

# The root-Gaussian Cox Process for Spatio-temporal Disease Mapping with Aggregated Data

Zeytu G. Asfaw<sup>1</sup>, Patrick E. Brown<sup>2,3</sup> and Jamie Stafford<sup>2</sup>

<sup>1</sup>School of Public Health, Addis Ababa University, Addis Ababa, Ethiopia

<sup>2</sup>Department of Statistical Sciences, University of Toronto, Toronto, Canada

<sup>3</sup>Center for Global Health Research, St. Michael's Hospital, Toronto, Canada

March 24, 2023

## Abstract

This paper focuses on the analysis of spatial data aggregated in space and time when the boundaries of geographic regions change over time. This can occur when reported cases of a health outcome are counted in regions over time and these regions change occasionally. We extend the spatial root-Gaussian Cox Process (RGCP), which uses the square-root link function rather than the usual log-link function, to the spatio-temporal case. A simulation study shows the algorithm can identify a spatio-temporal risk surface, and an analysis of malaria incidence in India is presented.

Keywords: EMS algorithm; spatial statistics; Gaussian random field; Matrix correlation and AR(1).

## 1 Introduction

When disease dynamics are affected in both space and time over longer periods hold more than one census of the population, computationally efficient algorithms and models providing interpretable inference are necessary. Besag et al., (1991) have formulated the dominant methodology for modeling random variation in disease risk for data collected over a short time interval.

These methods cannot handle aggregated data collected for extended periods with spatial boundaries changing during the study period. Case counts in a particular area are quite common because point locations for health outcome data are rare due to privacy and data quality issues. Besides, mapping disease risk often involves working with data that have been spatially aggregated to census regions or postal regions, either for administrative reasons or confidentiality. When studying rare diseases, data must be collected over a long period to accumulate a meaningful number of cases. These long periods can result in spatial boundaries of the census regions changing over time Nguyen et al. (2012).

Disease mapping aims to highlight the basic structure of scattered spatial health data, but it is relevant to apply such a methodology when the underlying risk is meant to be spatially, temporarily, and also, spatio-temporal structured. In the present study, we are concerned with the latter analysis, aggregated data is highly influenced by boundary changes over the study period in such a case investigating the temporal effect is optimally important. Besides, spatio-temporal analysis has additional benefits over purely spatial or temporal/time series because it allows the investigator to simultaneously study the persistence of patterns over-time and illuminate any unusual patterns. Of course, various continuously-indexed spatio-temporal process models have been constructed to characterize spatio-temporal dependence structures, but the computational complexity for model fitting and predictions grows in a cubic order with the size of the dataset, and application of such models is not feasible for large datasets and varying boundaries over data collection period (see, Mason et al.,1975; Lawson et al.,2000 and Sylvain et al. ,2015).

To this end, to extend purely spatial models to spatio-temporal models with varying boundaries, Fan et al. (2011) and Nguyen et al. (2012) have considered local-EM algorithms. But the drawback of these methodologies with local EM algorithms is that it is completely non-parametric and when their smoothing matrix is large, as is often the case, the implementation becomes computationally intensive. Brown and Stafford (2021) have proposed an alternative to local EM based on a root-Gaussian Cox Process (RGCP). It uses model-based inference and is computationally efficient. Moreover, it uses a square-root link function, rather than the usual log link, which leads to a generalized EMS algorithm. The current study extends the purely spatial model Brown and Stafford (2021) by incorporating temporarily varying risk surface and intensity functions.

The analysis of spatio-temporal data requires that both temporal correlation and spatial correlation be taken into account. While we consider both temporal and spatial variation simultaneously, it adds significant complexity to the data analysis process for two major reasons. First, continuous and discrete changes of spatial and non-spatial properties of spatio-temporal objects. Second, the influence of collocated neighboring spatio-temporal objects on one another. To reduce such complexity a simple, but commonly used class of spatio-temporal covariance model assumes a separable form that factors into a purely spatial and a purely temporal component. Besides, a geostatistical approaches for modelling spatio-temporal data rely on parametric covariance models and rather stringent assumptions such as stationarity, separability, and full symmetry. Thus, the present study has considered Matern spatial correlation with AR(1) temporal variation provides a recipe for computing the smoothing matrix at the S-step.

The paper is organized as follows. Sound motivations are written in the Introduction section. Section 2 describes the model formulation and heading to discretized the entire study regions to define disjoint regions and also, discretized the risk surface  $\lambda(\mathbf{s}, t)$ ; a root-Gaussian Cox Process is clearly defined and a Spatio-temporal EMS algorithms is derived. Besides, the implementation procedures are clearly written. The basic findings of the study are presented and discussed in Section 3. Finally, concluding remarks are provided in Section 4.

## 2 Model Formulation

In the geostatistical approach (see for example, Cressie 1993; Gelfand et al., 2010; Cressie and Wikle, 2011) data coming from monitoring networks are assumed to be realizations of a continuously indexed spatial process (random field) changing in time denoted by

$$Y(\mathbf{s}, t) = \{y(\mathbf{s}, t) : y(\mathbf{s}, t) \in D \subseteq \mathfrak{R}^2 \times \mathfrak{R}\}$$

which is indexed in space by  $\mathbf{s} \in \mathfrak{R}^2$  and in time by  $t \in \mathfrak{R}$

Cox processes provide useful and frequently applied models for aggregated spatial point patterns where the aggregation is due to a stochastic environmental heterogeneity, see e.g. Diggle (1983), Cressie (1993), Stoyan et al. (1995), and the references therein. A Cox process is 'doubly stochastic' as it arises as an inhomogeneous Poisson process (IPP) with a random intensity measure. The random intensity measure is often specified by a random

intensity function or as we prefer to call it an intensity process or surface. Here we have considered to two hierarchical level. The first level specifies a distribution for the location  $X_i$  and time  $T_i$  of each case  $i$  given an underlying risk surface  $\lambda(\mathbf{s}, t)$ . The second level specifies the process by which a portion of these event locations  $(X_i, T_i)$  is censored or aggregated to produce observed data  $\mathbf{Y}$  consisting of postal codes or names of census regions. Thus, in the present study, we have considered a natural model for the times and locations of cases is an inhomogeneous Poisson process with

$$[X_i, T_i] \sim \text{Poisson Process}[\lambda(\mathbf{s}, t)\mathcal{O}(\mathbf{s}, t)].$$

Here  $i = 1, 2, \dots, K$  indexes. The offset  $\mathcal{O}(\mathbf{s}, t)$  surface is the expected intensity of cases.

The ultimate objective of the present study is to derive an algorithm for making inference on a relative risk  $\lambda(\mathbf{s}, t)$  conditional on the point locations  $(X_i, T_i)$  when they are observed directly and the aggregated  $Y_i$  for those cases whose point locations have been censored. The likelihood function of an inhomogeneous Poisson process for  $\lambda(\mathbf{s}, t)$  (see Illian et al., 2008) is

$$\mathcal{L}(\lambda(\cdot, \cdot); \{X_i, T_i\}) = \left[ \prod_{i=1}^K \mathcal{O}(X_i, T_i) \lambda(X_i, T_i) \right] \exp \left[ - \int \int \mathcal{O}(u, v) \lambda(u, v) dudv \right]. \quad (1)$$

Now, we have a space-time process and the notation above could correspond to the  $X_i$  being residential locations of individuals within a common geographic region  $M$  where events happened at random times  $T_i$ .

## 2.1 A Discretized IPP

In the context considered here event locations  $X_i$  are area-censored, with observations  $x_i$  aggregated to a map of disjoint regions  $S_{ij}$  over time period  $t$  with  $M_j = \bigcup \{S_{ij}; j = 1, \dots, J_i\}$ . For such spatio-temporal data the  $S_{ij}$  might be geographic reporting regions - postal codes or enumeration areas are examples. The observed data become counts  $Y_{ij}$  of the number of events for region  $S_{ij}$  over a period of time  $T$  rather than event locations measured at a particular time  $T_i$ . More formally we set  $\mathbf{Y} = \{Y_{ij}; j = 1, 2, \dots, J_i\}$  where  $Y_{ij} = \{X_i \in S_{ij}, T_i \in C\}$  has a Poisson distribution with mean  $E(Y_{ij}) = \int_C \int_{S_{ij}} \lambda(u, v) \mathcal{O}(u, v) dudv$ , where  $C = [1, T]$ .

It is convenient to discretize  $\lambda(\mathbf{s}, t)$  and define it as a piecewise constant function. To discretize  $\lambda(\mathbf{s}, t)$  we divide the study area  $M = \bigcup_{j=1}^{J_i} M_j$  into a number of disjoint regions  $Q_\ell$  each with centroid  $q_\ell$  where  $Q = \{Q_\ell; \ell = 1, \dots, L\}$ . The  $Q$  with the smallest number of elements would be the regions obtained by overlaying all of the boundaries of the maps  $M_j$ . Another possible  $Q$  is the cells of a pixelated grid over  $M$  and observation time  $C_i, i = 1, \dots, T$ . We can now approximate  $\lambda(\mathbf{s}, t)$  by a piecewise constant function formed from the integrated values

$$\lambda_{i\ell} = \int_{C_i} \int_{Q_\ell} \frac{\lambda(u, v)}{\|C_i\| \|Q_\ell\|} dudv$$

so that  $\lambda(s, t) \approx \lambda(q_\ell, c_i); t \in C_i, s \in Q_\ell$ , and  $c_i$  is the mid point of interval  $C_i$ . We can now write the distribution of  $Y_{ij}$  as

$$Y_{ij} \sim \text{Pois} \left[ \sum_{\ell} \lambda_{i\ell} \mathcal{O}_{ij\ell} \right]$$

where

$$\mathcal{O}_{ij\ell} = \int_{C_i} \int_{S_{ij} \cap Q_\ell} \mathcal{O}(\mathbf{u}, v) dudv.$$

When the offset  $\mathcal{O}(\mathbf{s}, t)$  is constant within time periods with

$$\mathcal{O}(s, t) = \mathcal{O}_i(\mathbf{s}), t \in C_i$$

then

$$\mathcal{O}_{ij\ell} = \|C_i\| \int_{S_{ij} \cap Q_\ell} \mathcal{O}_i(u) du.$$

Since  $S_{ij} \cap Q_\ell$  is often empty (each census region is small and overlaps with only a small number of  $Q_\ell$ ),  $\mathcal{O}_{ij\ell}$  is often zero.

The number of events in disjoint regions are independent and the log-likelihood for  $\boldsymbol{\lambda}$  is

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum_{ij} \left\{ Y_{ij} \log \left[ \sum_{\ell} \lambda_{i\ell} \mathcal{O}_{ij\ell} \right] - \sum_{\ell} \lambda_{i\ell} \mathcal{O}_{ij\ell} \right\}. \quad (2)$$

## 2.2 A Root-Gaussian Cox Process

For an non-homogeneous Poisson process it is common to model the intensity function as a random effect rather than a parameter. For example, we may model the intensity using a latent Gaussian process where  $\lambda(s, t) = g[\theta(\mathbf{s}, t)]$ ,  $g$  is a link function, and

$$\theta(s, t) \sim \mathcal{GP}\{\mu(s, t), \rho[(\mathbf{s}, t), (\mathbf{s}', t')]\}$$

where  $\rho$  is a spatio-temporal covariance function.

Now, we have considered the unusual step of defining the link function as  $g[\theta(\mathbf{s}, t)] = [\theta(\mathbf{s}, t)]^2$  rather than  $g[\theta(\mathbf{s}, t)] = \exp[\theta(\mathbf{s}, t)]$ , of course a popular choice for such models is the log-Gaussian process but here we propose a root-Gaussian cox process because it conveniently leads to comparators of the Spatio-temporal EMS algorithms while the log-Gaussian process does not. We discretize the Gaussian process by

$$\theta(s, t) = \theta_{i\ell}; \mathbf{s} \in Q_\ell \text{ and } t \in C_i$$

so that  $\lambda_{i\ell} = \theta_{i\ell}^2$ . Setting  $\boldsymbol{\theta} = \{\theta_{11}, \theta_{22}, \dots, \theta_{LT}\}$  and assume that  $\boldsymbol{\theta}$  is distributed as a Gaussian random field (GRF):  $\mathcal{N}(\boldsymbol{\alpha}, \Sigma_{\phi, \psi, \sigma^2})$  over a period of time. We further assume  $\boldsymbol{\alpha} = \mathbf{1}$  acknowledging the possibility that the intensity surface is constant with any systematic variability captured by the offset function. A spatial temporal covariance matrix is very dense since all time points and spatial locations are dependent. Moreover, if we consider spatial location size  $n_1$  and time points  $n_2$ , the resulting matrix gets very large. For example, if  $n_1 = 100$  and  $n_2 = 40$  and now we have to work with a 4000 x 4000 covariance matrix. To make computations with less demanding and more efficient, we formulate the covariance matrix in separable form. The covariance is the product of a valid 2d spatial and a valid 1d temporal covariance/ correlation function with

$$Cov[\theta(\mathbf{s}, t), \theta(\mathbf{s}', t')] = \sigma^2 \rho_1(\|\mathbf{s} - \mathbf{s}'\|; \psi) \rho_2(|t - t'|; \phi).$$

The covariance matrix can be written as the Kronecker product of the two separate covariance matrices

$$\Sigma = \sigma^2 V_1(\psi) \otimes V_2(\phi)$$

where  $V_1(\psi)$  has size  $L \times L$  and having elements  $[V_1(\psi)]_{jj'} = \rho_1(\|\mathbf{s}_j - \mathbf{s}_{j'}\|; \psi)$  and similarly,  $V_2(\phi)$  has size  $T \times T$  and having elements  $[V_2(\phi)]_{ii'} = \rho_2(|t_i - t_{i'}|; \phi)$ . The property of Kronecker product,

$$[\Sigma(\sigma^2, \psi, \phi)]^{-1} = \sigma^2 [V_1(\psi)]^{-1} \otimes [V_2(\phi)]^{-1}$$

is useful, where  $[V_1(\psi)]^{-1} = P_1$ ,  $[V_2(\phi)]^{-1} = P_2$ . Moreover, we have considered  $P_1$  as the precision matrix of a Gaussian Markov Random Fields and

$P_2$  as that of an autoregressive AR(1). AR processes of order one are a particularly parsimonious model choice, in which the next value of the process depends only on the previous one and its precision matrix is tridiagonal. It is denoted by  $\mathbf{P}_2(\phi)$  and expressed by

$$P_2(\phi) = \frac{1}{\phi^2} \begin{bmatrix} 1 & -\phi & \dots & 0 \\ -\phi & 1 + \phi^2 & \phi & \dots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \dots & \phi & 1 \end{bmatrix}.$$

$P_1(\psi)$  is also sparse, with non-zero entries only for pairs of cells which are neighbors (see, Brown and Stanford, 2021). The penalized log-likelihood of  $\theta$  is

$$\begin{aligned} \mathcal{L}_p(\boldsymbol{\theta}) &= \sum_{ij} \left\{ \mathbf{Y}_{ij} \log \left[ \sum_{il} \mathcal{O}_{ijl} \theta_{il}^2 \right] - \sum_{il} \mathcal{O}_{ijl} \theta_{il}^2 \right\} \\ &\quad - \frac{1}{2\sigma^2} (\boldsymbol{\theta} - \mathbf{1})^T [P_1(\psi) \otimes P_2(\phi)] (\boldsymbol{\theta} - \mathbf{1}). \end{aligned} \quad (3)$$

Note the first term of  $\mathcal{L}_p$  is of the same form as the likelihoods (2) but where it has been reparametrized in terms of  $\boldsymbol{\theta}$  while the second has the form of a generalized quadratic penalty for that likelihood.

### 2.3 A Spatio-Temporal EMS Algorithm

We can now consider various approaches for estimating  $\boldsymbol{\lambda}$ . In what follows we derive expressions for the first two derivatives of  $\mathcal{L}_p(\boldsymbol{\theta})$ . The first produces an iteration (9).

Hereafter we re-index  $k$  as  $k = (i - 1)L + \ell$  and differentiating once with respect to  $\theta_k$  gives

$$\begin{aligned} \frac{\partial \mathcal{L}_p(\boldsymbol{\theta})}{\partial \theta_k} &= 2 \sum_{ij} \left\{ Y_{ij} \frac{\mathcal{O}_{ijl} \theta_k}{\sum_m \mathcal{O}_{ijm} \theta_{(i-1)L+m}^2} - 2 \mathcal{O}_{ijl} \theta_k \right\} - \mathbf{P}_k^T (\boldsymbol{\theta} - \mathbf{1}) \\ &= 2 \sum_{ij} \left\{ \frac{\mathcal{O}_{il}}{\theta_k} \left[ Y_{ij} \frac{\mathcal{O}_{ijl} \theta_k^2}{\mathcal{O}_{il} \sum_m \mathcal{O}_{ijm} \theta_{(i-1)L+m}^2} \right] - 2 \mathcal{O}_{il} \theta_k \right\} - \mathbf{P}_k^T (\boldsymbol{\theta} - \mathbf{1}) \\ &= 2 \boldsymbol{\mu}(\boldsymbol{\lambda}^r)_k - 2 \mathcal{O}_{il} \theta_k - \mathbf{P}_k^T (\boldsymbol{\theta} - \mathbf{1}). \end{aligned} \quad (4)$$

Here

$$\mathbf{P} = \frac{1}{\sigma^2} P_1(\psi) \otimes P_2(\phi)$$

and

$$\begin{aligned}\mathcal{O}_k &= \sum_{ij} \mathcal{O}_{ij\ell}, \\ \boldsymbol{\mu}(\boldsymbol{\lambda}^r)_k &= \theta_k \sum_{ij} \frac{Y_{ij} \mathcal{O}_{ij\ell}}{\sum_m \mathcal{O}_{ijm} \lambda^{(i-1)L+m}},\end{aligned}\tag{5}$$

and  $i$  and  $\ell$  are fixed.

We may now differentiate with respect to  $\boldsymbol{\theta}$  in vector-matrix notation, with  $\Theta = \text{diag}(\boldsymbol{\theta})$  and

$$\begin{aligned}\frac{\partial \mathcal{L}_p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= 2\mathcal{O}\Theta^{-1}\boldsymbol{\mu}(\boldsymbol{\lambda}) - 2\mathcal{O}\boldsymbol{\theta} - \mathbf{P}(\boldsymbol{\theta} - \mathbf{1}) \\ &= 2\mathcal{O}\Theta^{-1}\boldsymbol{\mu}(\boldsymbol{\lambda}) - 2\mathcal{O}\boldsymbol{\alpha} - 2\mathcal{O}(\boldsymbol{\theta} - \mathbf{1}) - \mathbf{P}(\boldsymbol{\theta} - \mathbf{1}) \\ &= 2\mathcal{O}\Theta^{-1}(\boldsymbol{\mu}(\boldsymbol{\lambda}) - \boldsymbol{\eta}(\boldsymbol{\lambda})) - 2(\mathcal{O} + \frac{1}{2}\mathbf{P})(\boldsymbol{\theta} - \mathbf{1}) \\ &= 2\mathcal{O}\Theta^{-1}\tilde{\boldsymbol{\mu}}(\boldsymbol{\lambda}) - 2(\mathcal{O} + \frac{1}{2}\mathbf{P})\Theta^{-1}(\boldsymbol{\lambda} - \boldsymbol{\eta}(\boldsymbol{\lambda})).\end{aligned}\tag{6}$$

Setting  $\frac{\partial \mathcal{L}_p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$  and solving for  $\boldsymbol{\lambda}$  yields

$$\mathcal{O}\Theta^{-1}\tilde{\boldsymbol{\mu}}(\boldsymbol{\lambda}) = (\mathcal{O} + \frac{1}{2}\mathbf{P})\Theta^{-1}(\boldsymbol{\lambda} - \boldsymbol{\eta}(\boldsymbol{\lambda}))\tag{7}$$

$$(\boldsymbol{\lambda} - \boldsymbol{\eta}(\boldsymbol{\lambda})) = \Theta(\mathcal{O} + \frac{1}{2}\mathbf{P})^{-1}\mathcal{O}\Theta^{-1}\tilde{\boldsymbol{\mu}}(\boldsymbol{\lambda})\tag{8}$$

$$\boldsymbol{\lambda} = \mathcal{S}(\boldsymbol{\lambda})\tilde{\boldsymbol{\mu}}(\boldsymbol{\lambda}) + \boldsymbol{\eta}(\boldsymbol{\lambda}).\tag{9}$$

Differentiating  $\mathcal{L}_p$  with respect to  $\boldsymbol{\theta}_{i\ell}$  and solving the resulting equations yields a generalized EMS iteration for  $\boldsymbol{\lambda}$

$$\boldsymbol{\lambda}^{r+1} = \mathcal{S}(\boldsymbol{\lambda}^r)\tilde{\boldsymbol{\mu}}(\boldsymbol{\lambda}^r) + \boldsymbol{\eta}(\boldsymbol{\lambda}^r)\tag{10}$$

where  $\tilde{\boldsymbol{\mu}}(\boldsymbol{\lambda}^r) = \boldsymbol{\mu}(\boldsymbol{\lambda}^r) - \boldsymbol{\eta}(\boldsymbol{\lambda}^r)$ ,  $\boldsymbol{\eta}(\boldsymbol{\lambda}^r) = \boldsymbol{\theta}^r$

$$\begin{aligned}\mathcal{S}(\boldsymbol{\lambda}^r) &= \Theta_r \left\{ \mathcal{O} + \frac{1}{2}\Sigma_{\phi, \psi, \sigma^2}^{-1} \right\}^{-1} \mathcal{O}\Theta_r^{-1} \\ &= \Theta_r \left\{ \mathcal{O} + \frac{1}{2\sigma^2} P_1(\psi) \otimes P_2(\phi) \right\}^{-1} \mathcal{O}\Theta_r^{-1}\end{aligned}\tag{11}$$

with  $\Theta_r = \text{diag}(\theta_k^r)$ ,  $\mathcal{O} = \text{diag}(\mathcal{O}_k)$ . Note that at each iteration of the algorithm we set  $\boldsymbol{\theta}^r = \sqrt[r]{\boldsymbol{\lambda}^r}$ .

## 2.4 Implementation

The algorithm implemented in R and the set of procedures summarized hereunder.

**Step 1:** The spatial surface is discretized on a rectangular area containing the study region with lattice of square grid cells  $G_{i\ell}$  with  $\theta(s, t) = \hat{\theta}_{i\ell}$ ;  $(s, t) \in G_{i\ell}$  and  $\ell = 1, 2, \dots, L$ ;  $i = 1, 2, \dots, T$ .

**Step 2:** The discretized model is parametrized as  $\lambda_{i\ell} = \theta_{i\ell}^2$  with

$$[\theta_{11}, \theta_{22}, \dots, \theta_{LT}]^T \sim \mathbf{N}(\mathbf{1}, \sigma^2 \Sigma_{\psi, \phi}^{-1}),$$

where  $\Sigma_{\psi, \phi}^{-1}$  is the Kronecker product of a Matern correlation matrix and a temporal AR(1) precision matrix.

**Step 3:** The Matern correlation matrix is approximated using a Markov random field on the square grid of the  $G_\ell$ , as described in Lindgren et al.(2011). This matrix is computed with the `maternGmrfPrec` function in the `geostatsp` package (Brown, 2015), and includes the edge correction from section 5.1.4 of Rue and Held (2005). Similarly, the AR(1) temporal precision matrix is constructed as a sparse-band matrix specifying its nonzero superdiagonals.

**Step 4:** The EMS recursion is implemented as

$$\begin{aligned} \boldsymbol{\lambda}^{(r+1)} &= 2\sigma^2 \Theta^{(r)} \tilde{\mathbf{S}}^{-1} \Theta^{(r)} \boldsymbol{\mu}(\boldsymbol{\lambda}^{(r)}) + \Theta^{(r)} [\mathbf{1} - 2\sigma^2 \tilde{\mathbf{S}}^{-1} \mathcal{O} \mathbf{1}] \\ \boldsymbol{\mu}(\boldsymbol{\lambda}^r)_k &= \theta_k \sum_j \frac{Y_{ij} \mathcal{O}_{ij\ell}}{\sum_m \mathcal{O}_{ijm} \lambda_{(i-1)L+m}} \\ \tilde{\mathbf{S}} &= 2\sigma^2 \mathcal{O} + \Sigma_{\psi, \phi}^{-1}. \end{aligned} \tag{12}$$

Since the  $\tilde{\mathbf{S}}$  matrix does not change with the iteration  $r$ ; a sparse Cholesky decomposition of  $\tilde{\mathbf{S}}$  can be performed once and used to solve  $\tilde{\mathbf{S}} \mathbf{b} = \Theta^{(r)} \boldsymbol{\mu}(\boldsymbol{\lambda}^{(r)})$  for  $\mathbf{b}$  at each iteration.

**Step 5:** Convergence of the EMS iteration is achieved when the first derivatives in (6) are all below a threshold.

## 3 Results

### 3.1 Simulation study and validation with Kentucky data

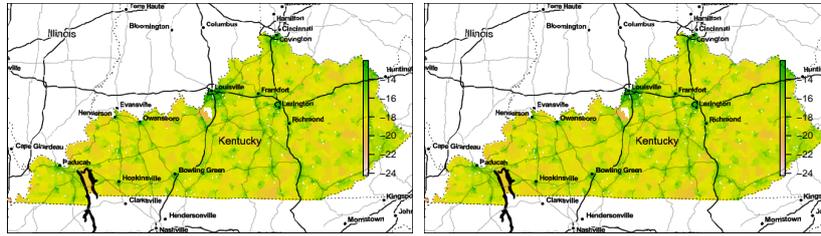
Simulation studies are used to obtain empirical results about the performance of statistical methods, particularly for the evaluation of new methods and algorithms. Similarly, to evaluate the performance of the newly formulated root-Gaussian Cox process for Spatiotemporal models, we have considered the geographic boundaries of the state of Kentucky in the United States of America where data is aggregated by administrative regions PUMA and county. We have simulated events in four different time points, Fig 1 (c, d, e, and f) and also, heading to aggregate the events by PUMA, Fig 1(l and i) and County, Fig 1(h and j).

The procedures are here, we have simulated the incidence of a fictitious disease in the state of Kentucky, aggregating events by county and also by Public Use Microdata Area (PUMA). Population-based offsets are available from the US census at the fairly fine spatial resolution of 3157 census tracts, shown in Figures 1a and 1b. Note that the offsets here are expected intensity per unit of surface area, not the expected area-level counts used for spatially discrete models. A spatiotemporal root-Gaussian Cox process was simulated using a Matern correlation function with a shape parameter of 1, and standard deviation of 0.25, and a range of 100km. For a time we have used exponential covariance with a variance of 1 and a scale of 10. We have considered 1100 grid cells and 4-time points.

The estimated and true intensity risk surface  $\lambda(s, t)$  are compared at each time point. For example, the first simulated risk surface Fig 2a compared with Fig 2b at the same time 1. Similarly, the estimates Fig 2c, 2e, and 2g were compared to their counterpart time point Fig 2d, 2f, and 2h, respectively. With the comparison we have estimated through simulation and true value/real datasets, we have shown close to the same outputs and thus, the newly formulated RGCP is recommended for the analysis of aggregated data whenever spatial boundary changes over the study period time t.

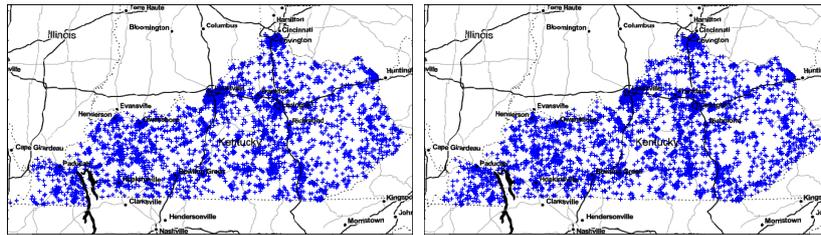
### 3.2 Validation with India Malaria Data

Here again, we have considered Indian malaria data in the same procedures as the states of Kentucky in the USA. This could be considered the second validation of the newly formulated RGCP. We have considered Indian malaria



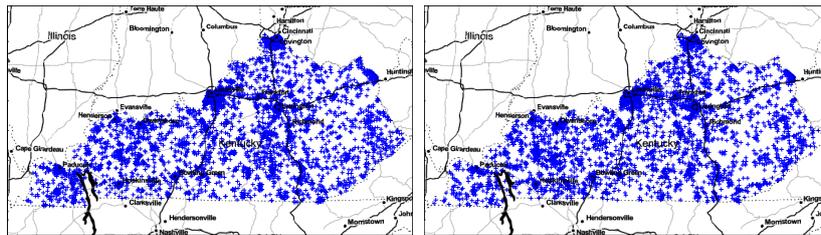
(a) offset 1

(b) offset 2



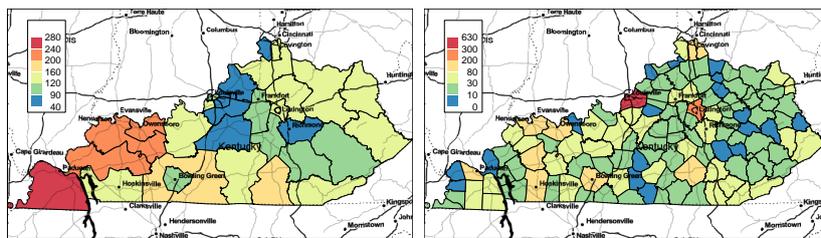
(c) event, time 1

(d) event, time 2



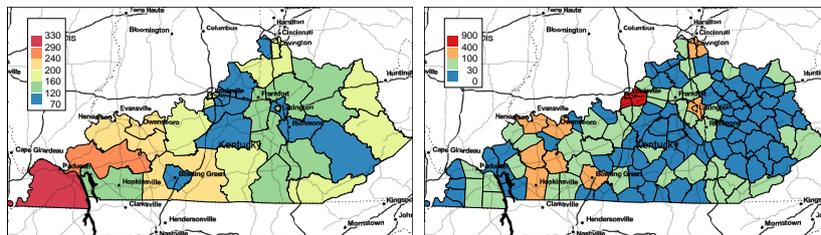
(e) event, time 3

(f) event, time 4



(g) PUMA, time 1

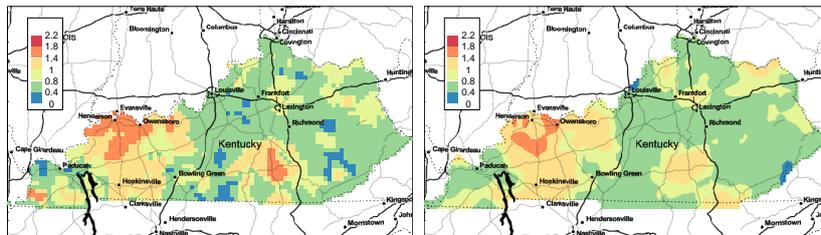
(h) County, time 2



(i) PUMA, time 3

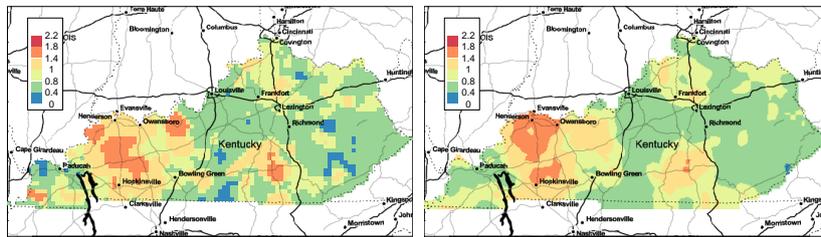
(j) County, time 4

Figure 1: Simulated root-Gaussian Cox process aggregated by county and Public Use Micro-data Area (PUMA). Background map ©Stamen Design



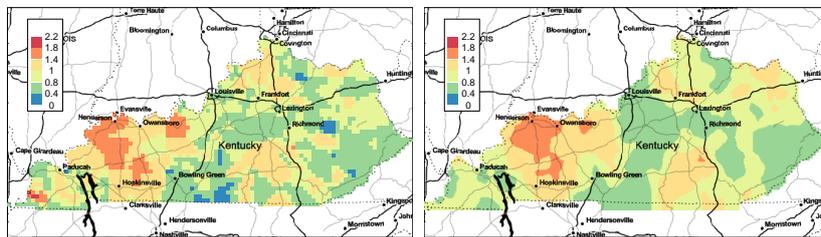
(a) estimate, time 1

(b) true, time 1



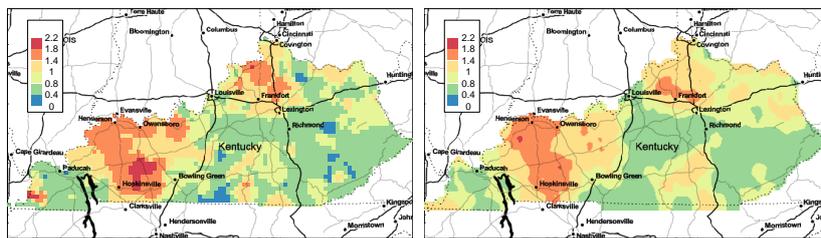
(c) estimate, time 2

(d) true, time 2



(e) estimate, time 3

(f) true, time 3



(g) estimate, time 4

(h) true, time 4

Figure 2: Estimated and true intensity surfaces at each time point for simulated data. Background map ©Stamen Design

cases due to real data availability, and it is a highly malaria-prevalent nation. Besides, the country carries 2% of the global malaria case burden and 2% of global malaria deaths WHO (2021).

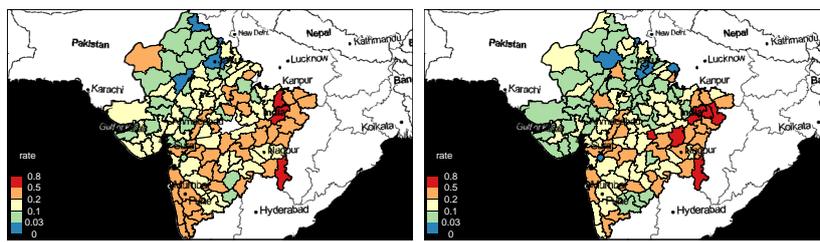
Figure 3 shows the proportion of malaria cases reported in four states in the west of India which are of the strain *Plasmodium falciparum*, the strain of malaria most dangerous to humans. The data come from the National Vector Borne Disease Control Programme (see Cohen et al., 2010), and are reported at the district level. Figures 3a, 3b, 3c, and 3d show the raw proportions of samples that were positive for *Plasmodium falciparum*, with the counts ranging from 0 to 16,000 with a median of 200. District boundaries are shown superimposed in Figure 3i. The red lines show the boundaries of the 123 districts in 1999, the green lines show new boundaries created when the number of districts expanded to 135 in 2000, and the blue lines show additional boundaries from 141 districts in 2001 and 2002.

An RGCP was fit to these data using an offset surface for each year calculated as number of samples tested in each district multiplied by the raw proportion for the entire study region in that year, and converted to an intensity by dividing each districts expected count by its surface area. Similarly, we have made comparison the real data analysis Fig (3a - 3d) with the estimated risk surface Fig (3e - 3h) and founding the same conclusion with section 3.1 above.

## 4 Concluding Remarks

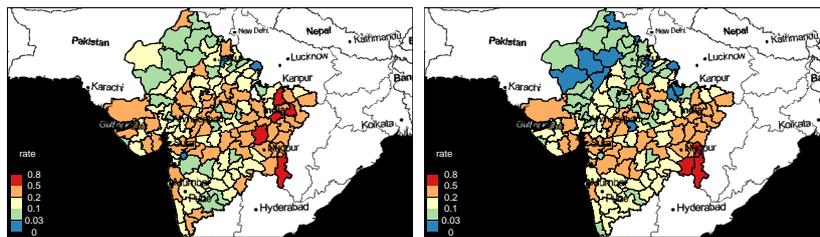
This paper has demonstrated how an extension of the local-EM algorithm from Brown and Stafford (2021) can be used for modelling spatio-temporal data. Compared to the spatial model in Brown and Stafford (2021), the spatio-temporal model in this paper involves a much larger smoothing matrix, fitting the model with four time points quadruples the number of rows and columns increasing the number of total entries 16-fold. By exploiting sparsity of the smoothing matrices and the fact that the smoothing matrix does not change over EM iterations, an efficient and tractable algorithm is developed for a difficult spatio-temporal problem.

The simulation study presented shows the algorithm produces estimated risk surfaces which are very similar to the truth, and the application to the India malaria data shows the algorithm can produce useful results for a real dataset. Further research is required to estimate the model parameters and



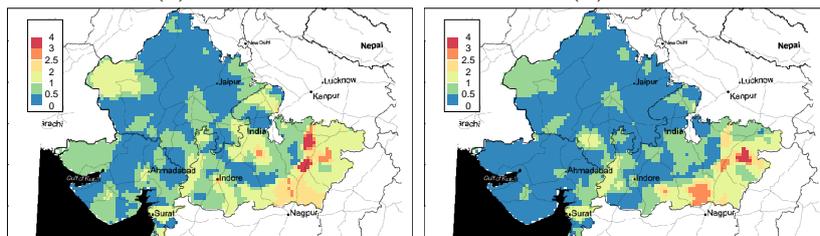
(a) 1999

(b) 2000



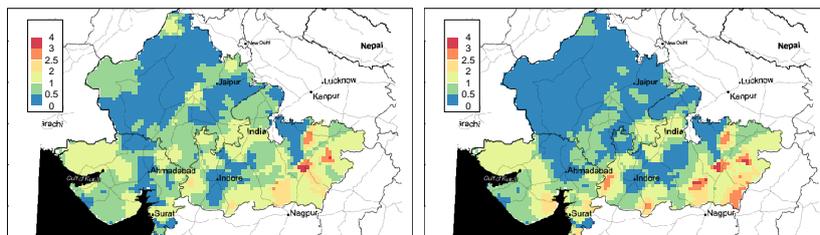
(c) 2001

(d) 2002



(e) Estimated risk, 1999

(f) Estimated risk, 2000



(g) Estimated risk, 2001

(h) Estimated risk, 2002



(i) Boundaries

Figure 3: *Plasmodium falciparum* as a proportion of all recorded Malaria cases, by district in the India states of Rajasthan, Madhya Pradesh, and Maharashtra, for selected years (a-d). Superimposed boundaries and predictions from an RGCP (e-h). Background map ©Stamen Design

prediction intervals for the estimated surface, the methods used by Brown and Stafford (2021) for spatial data should translate directly to the spatio-temporal case.

## Declarations

## Availability of data and material

Code and data for the simulation study are available from the authors upon request.

## Competing interests

The authors declare that they have no competing interests.

## Funding

Not applicable

## Acknowledgments

The authors are supported by Norwegian Programme for Capacity Development in Higher Education and Research for Development through HU-PhD-Mat-Stat-Sci project along with Center for Global Health Research, University of Toronto, Toronto, Canada. Background maps in Figures 1, 2 and 3 have map tiles by Stamen Design under the CC By 3.0 license using data by OpenStreet Map available under the Open Database License.

## References

1. Besag, J., York., and Mollie, A. (1991). Bayesian images restoration, with two applications in spatial statistics, 1991. *Annals of the Institute of Statistical Mathematics* 23, 1 - 20.
2. Brown, P.E.(2015). Model-based geostatistics the easy way. *Journal of Statistical Software*, 63.

3. Brown, E. P., and Stafford, S. (2021). *The root-Gaussian Cox Process and a generalized EMS algorithms*. Spatial Statistics, to appear.
4. Cohen, A.A., Dhingra, N., Jotkar, R. M., Rodriguez, P.S., Sharma, V.P., and Jha, P. (2010). The summary index of malaria surveillance (sims): a stable index of malaria within india. Population health metrics, 8(1):1.
5. Cressie, N. (1993), Statistics for Spatial Data, revised ed., Wiley, New York.
6. Cressie N, Wikle C (2011) Statistics For Spatio- Temporal Data. Wiley
7. Fan, C.P.S., Stafford, J., and Brown, P.E.(2011). Local-EM and the EMS Algorithms. *Journal of Computational and Graphical Statistics* 20, 750 - 766.
8. Diggle, P.J. (1983). Statistical Analysis of Spatial Point Patterns. Academic Press, London.
9. Gelfand A, Diggle P, Fuentes M, Guttorp P (eds) (2010) Handbook of Spatial Statistics. Chapman and Hall
10. Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., (2008). Spatial Analysis and Modelling of Spatial Point Patterns. Wiley and Sons.
11. Lindgren, F., Rue, H., and Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423-498.
12. Lawson A., Biggeri A. B., Boehning D., Lesaffre E., Viel J.-F., Clark A., Schlattmann P. and F. D. (2000). Disease mapping models: an empirical evaluation. *Statistics in medicine*, 19:2217-2241.

13. Mason T., MacKay F., Hoover R., Blot W. and Fraumeni J. (1975). Atlas of Cancer Mortality for US counties, 1950-1969. Technical report, Department of Health, Education, and Welfare Publication, Washington.
14. Nguyen, P., Brown, P. E., and Stafford, J. (2012). Mapping Cancer Risk in Southwestern Ontario with changing census boundaries. *Biometrics*, 68.
15. Rue, H. and Held, L. (2005). Gaussian Markov Random Fields: Theory and Applications. CRC Press, London. <https://doi.org/10.1201/9780203492024>
16. Stoyan, D., Kendall, W.S. and Mecke, J. (1995). Stochastic Geometry and Its Applications. 2nd ed. Wiley, Chichester.
17. Sylvain Colya, Myriam Charras-Garridoa, David Abriala, Anne-Franoise Yao-Lafourcadeb (2015). Spatiotemporal disease mapping applied to infectious diseases. *Procedia Environmental Sciences* 26 ( 2015 ) 32 37
18. World Health Organisation (2021). World Malaria Report 2021