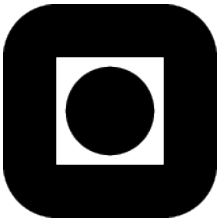
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Going off grid: Computationally efficient inference for log-Gaussian Cox processes

by

Daniel Simpson, Janine Illian, Finn Lindgren, Sigrunn H. Sørbye and Håvard Rue

PREPRINT
STATISTICS NO. 10/2011



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL
http://www.math.ntnu.no/preprint/statistics/2011/S10-2011.pdf
Daniel Simpson has homepage: http://www.math.ntnu.no/∼daniesi
E-mail: daniesi@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Going off grid: Computationally efficient inference for log-Gaussian Cox processes

Daniel Simpson*[1], Janine Illian[2], Finn Lindgren[1], Sigrunn H. Sørbye[3], and Håvard Rue[1]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[2]Centre for Research into Ecological and Environmental Modelling, University of St Andrews, Scotland
[3]Department of Mathematics and Statistics, University of Tromsø, Norway

November 1, 2011

Abstract

In this paper we introduce a new method for performing computational inference on log-Gaussian Cox processes (LGCP). Contrary to current practice, we do not approximate by a counting process on a partition of the domain, but rather attack the point process likelihood directly. In order to do this, we use the continuously specified Markovian random fields introduced by Lindgren et al. (2011). The inference is performed using the `R-INLA` package of Rue et al. (2009), which allows us to perform fast approximate inference on quite complicated models. The new method is tested on a real point pattern data set as well as two interesting extensions to the classical LGCP framework. The first extension considers the very real problem of variable sampling effort throughout the observation window and implements the method of Chakraborty et al. (Submitted). The second extension moves beyond what is possible with current techniques and constructs a log-Gaussian Cox process on the world's oceans.

1 Introduction

Datasets consisting, at least in part, of sets of locations at which some objects are present are ubiquitous in biology, ecology, economics, and a whole slew of other disciplines. In a number of situations, the most appropriate statistical models for this type of data are spatial point process models and such models have been extensively studied by statisticians and probabilists (Illian et al., 2008; Møller and Waagepetersen, 2003). They have not, however, been widely embraced by the scientists producing the data sets. The main reason for this is that *point process models are complicated and hard to fit.* As a result scientists often resort to using inappropriate methods. There is an interesting discussion of this issue in the context of "presence only" data sets in Chakraborty et al. (Submitted), which outlines a number of *ad hoc* approaches taken by the ecological community.

This difficulty is compounded by the fact that many real data sets typically do not have the simple structure of the classical point pattern data set considered in the statistical literature, i.e. that of a simple point pattern that has been observed everywhere within a simple, often rectangular plot. For instance, in real data sets the observation process is often not straight forward due to practical limitations or the observation window is highly complex. This includes data sets mapping the locations of bird species, for which very little data have been collected in the Himalayas due to obvious access issues. Therefore, on top of sampling issues that are relatively common such as

*Corresponding author. Email: `Daniel.Simpson@math.ntnu.no`

incompletely observed point patterns, positional errors, etc., this data set has a large hole in it where it is believed that these birds reside, but it is not practical to look for them. Very different, but similarly complex data deal with freak waves in the oceans. Even if we ignore the temporal aspect of the problem, along with the uncertainty in the observed locations, this data set remains complicated to deal with using standard methods. Here, the observation window is a region covering most of a sphere (the earth!) with a very complicated boundary. Motivated by data sets of this nature this paper aims to propose an easy to use, computationally efficient method for performing inference on spatial point process models that is sufficiently flexible to handle these and other data structures. In fact, these two examples inspired the case studies in Sections 8.2 and 8.3.

In this paper we will focus on log-Gaussian Cox processes (LGCPs), which are a class of flexible models that are particularly useful in the context of modelling aggregation relative to some underlying unobserved environmental field (Illian et al., 2011; Møller et al., 1998). However, standard methods for fitting Cox processes are computationally expensive and the MCMC methods that are commonly used are famously difficult to tune for this problem. Recently, Illian et al. (2011) developed a fast, flexible framework for fitting complicated LGCPs using integrated nested Laplace approximation (INLA), originally developed by Rue et al. (2009). This approach follows the current trend in the literature of constructing a Poisson approximation to the true LGCP likelihood and using this approximation to perform the inference. This approximation proceeds by placing a regular lattice over the observation window and counting the number of points in each cell. It can be shown that, if the regular lattice is fine enough and the latent Gaussian field is appropriately discretised, this approximation is quite good (Waagepetersen, 2004). It can, however, be very wasteful, especially when the intensity of the process is high or the observation window is large or oddly shaped (see Section 8.3 for an example of a point process on a very odd domain).

In this paper we essentially have two aims. The first aim is to re-examine the standard methodology for performing Bayesian inference on log-Gaussian Cox processes and propose an approach that is much more computationally efficient based on the stochastic partial differential equation models introduced by Lindgren et al. (2011). The key difference between this approach and those that exist in the literature is that the specification of the Gaussian random field is completely separated from the approximation of the LGCP likelihood. This leads to far greater flexibility, which allows us to tackle a number of exciting practical problems. The second aim is to demonstrate that this approach can be handled within the general INLA framework of Rue et al. (2009). This provides a unified modelling structure and an associated R-library and makes the methods that we develop accessible to and useful for scientists. In essence, we are aiming to make point process modelling *a blacker box*.

The paper proceeds as follows. In Section 2, we review the basic properties of log-Gaussian Cox processes and in Section 3 we review the standard approximation and discuss its limitations. The main thrust of our approach is outlined in general in Section 4, where we discuss spatially continuous approximations to Gaussian random fields. Section 5 is devoted to spatially Markovian random fields and their specification through stochastic partial differential equations. With these ingredients in place, we describe our approximate likelihood in Section 6 and its implementation in the `R-INLA` package in Section 7. Finally in Section 8, we present three examples: a standard example using tree data, a simulation study with an unsampled area, and a point process on the oceans.

2 Log-Gaussian Cox processes

Consider a bounded region $\Omega \subset \mathbb{R}^2$. A simple and frequently used point process model is the inhomogeneous Poisson process, in which the number of points within a region $D \subset \Omega$ is Poisson distributed with mean $\Lambda(D) = \int_D \lambda(s)\, ds$, where $\lambda(s)$ is the *intensity surface* of the point process. Given the intensity surface and a point pattern Y, the likelihood of an inhomogeneous Poisson process

is given by

$$\pi(Y|\lambda) = \exp\left(|\Omega| - \int_\Omega \lambda(s)\,ds\right) \prod_{s_i \in Y} \lambda(s_i). \tag{1}$$

This likelihood is analytically intractable as it requires the integral of the intensity function, which typically cannot be calculated explicitly. This integral can, however, be computed numerically using fairly traditional methods.

Treating the intensity surface as a realisation of a *random field* $\lambda(s)$ yields a particularly flexible class of point processes known as *Cox processes* or doubly stochastic Poisson processes (Møller and Waagepetersen, 2003). These are typically used to model aggregation in point patterns resulting from observed or unobserved environmental variation. In this paper we are particularly interested in the case of *log-Gaussian Cox processes*, where the intensity surface is modelled as

$$\log(\lambda(s)) = Z(s),$$

and $Z(s)$ is a Gaussian random field. Conditional on a realisation of $Z(s)$, a log-Gaussian Cox process is an inhomogeneous Poisson process. It follows that the likelihood for a LGCP is of the form (1), where the integral is further complicated by the stochastic nature of $\lambda(s)$ and methods for approximating this likelihood will be the focus of the next two sections. We note that the log-Gaussian Cox process fits naturally within the Bayesian hierarchical modelling framework. Furthermore, it is a latent Gaussian model, which allows us to embed it within the INLA framework. This embedding paves the way for extending the LGCP to include covariates, marks and non-standard observation processes while still allowing for computationally efficient inference (Illian et al., 2011). Therefore, approximating the likelihood in (1) is an essential 'baseline level' calculation for many practical problems. We will investigate this further in Section 8.

3 Computation on fine lattices is wasteful

A common method for performing inference with LGCPs is to take the observation window Ω and construct a fine regular lattice over it (Illian and Rue, 2010; Illian et al., 2011; Møller et al., 1998) and to then consider the number of points N_{ij} observed in each cell s_{ij} of the lattice. It is a simple consequence of the definition of a LGCP that N_{ij} may be considered as independent Poisson random variables, that is

$$N_{ij} \sim Po(\Lambda_{ij}),$$

where $\Lambda_{ij} = \int_{s_{ij}} \lambda(s)\,ds$ is the total intensity in each cell. Generally speaking, it is impossible to compute the total intensity for each cell and we therefore use the approximation $\Lambda_{ij} \approx |s_{ij}| \exp(z_{ij})$, where z_{ij} is a 'representative value' of $Z(s)$ within the cell s_{ij} and $|s_{ij}|$ is the area of cell s_{ij}. With this, the LGCP model has been transformed into the rather more benign GLM framework and can be treated appropriately. This method has been used to great success in a number of applications and it has been shown to converge to the true solution as the size of the cells decreases to zero (Waagepetersen, 2004).

Of course, the elephant in this particular room is the latent multivariate Gaussian vector $\mathbf{z}$ that contains the z_{ij}s—if $Z(s)$ is a general Gaussian random field, $\mathbf{z}$ will have a *dense* covariance matrix. This limits the above method to quite small lattices. We therefore consider more computationally efficient spatial models than standard Gaussian random fields. A common approach is to model $\mathbf{z}$ as a conditional autoregressive (CAR) model on the fine lattice and use this to perform fast computations (Rue and Held, 2005). The CAR approach has been used extensively in applications and has been implemented successfully within the INLA framework by Illian and Rue (2010); Illian et al. (2011). It should be noted that both of these methods rely heavily on the regularity of the lattice—somewhat surprisingly, it is quite difficult to construct a CAR model on an irregular lattice that is resolution consistent (Rue and Held, 2005), although there has been a recent breakthrough in this direction

by Lindgren et al. (2011). A third method, which is more in line with the general thrust of this paper, uses predictive process modelling to construct a low-rank continuous Gaussian field over Ω (Chakraborty et al., Submitted).

One of the primary contentions of this paper is that the methods listed above, although successful, are unsatisfactory. The problem is that the computational lattice has two fundamentally different roles. The first, and most natural, role is to approximate the latent Gaussian random field $Z(s)$. The second, and—in our opinion—unnatural, role of the computational lattice is to also approximate the position of the points. This is unfortunate as a lot of effort goes into collecting position data with a high degree of precision. Clearly, the finer the lattice is, the less the binning procedure *moves* the data. Therefore the quality of the likelihood approximation primarily depends on the size of the grid. As a result, we are required to compute on a much finer grid than is necessary for the approximation of the latent Gaussian field. Clearly, lattice based approaches are hence inherently computationally wasteful in the context of log-Gaussian Cox processes.

The inflexibility inherent in lattice-based methods has another interesting implication—we are unable to locally refine our approximation to the latent random field. In a lot of situations, this is not a problem, however, in the problem considered in Section 8.2 there is a large region that contains no data and will not, therefore, greatly impact the posterior inference. This begs the question: why are we wasting computational resources generating a high resolution approximation to the latent field over this area? Clearly a better and less computationally intensive solution would be to reduce the resolution in these areas without effecting the resolution in the areas that have been sampled. This is completely impossible with lattice-based methods. While this flexibility will not be needed in every problem, it is clearly convenient to have the option of changing the resolution of the approximation locally.

4 Continuous specification of the latent random field

The arguments in the previous section suggest that there are some strong disadvantages in using the common lattice-based framework to perform inference on log-Gaussian Cox processes. The barrier to efficient inference is the *double duty* that the lattice is engaged in. Therefore, we propose to use the computational mesh (note the change from lattice) *only* for representing the latent Gaussian random field. In order to do this we need a *continuous* specification of our random field model, which we will take to be a finite-dimensional basis function expansion

$$Z(s) = \sum_{i=1}^{n} z_i \phi_i(s), \tag{2}$$

where $\mathbf{z} = (z_1, z_2, \ldots, z_n)^T$ is a multivariate Gaussian random vector and the $\{\phi_i(s)\}_{i=1}^n$ are deterministic basis functions. There are three common approximations to Gaussian random fields that can be written in this form—process convolution models (Higdon, 1998; Xia and Gelfand, 2005), predictive process, or fixed-rank Kriging, models (Banerjee et al., 2008; Cressie and Johannesson, 2008), and the stochastic partial differential equation models of Lindgren et al. (2011).

The advantage of a finite dimensional, continuously specified Gaussian random field (2) is that it automatically allows us to compute the term in the likelihood (1) that depends on the data using the *exact* positions of the data points for any underlying grid using an (at most) $\mathcal{O}(n)$ operation. We note that if the basis functions have compact support, which is the case in the SPDE models, the field can be evaluated in $\mathcal{O}(1)$ operations!

The only thing that we then have left to do is to approximate the integral

$$\int_{\Omega} \exp\left(\sum_{i=1}^{n} z_i \phi_i(s)\right) ds, \tag{3}$$

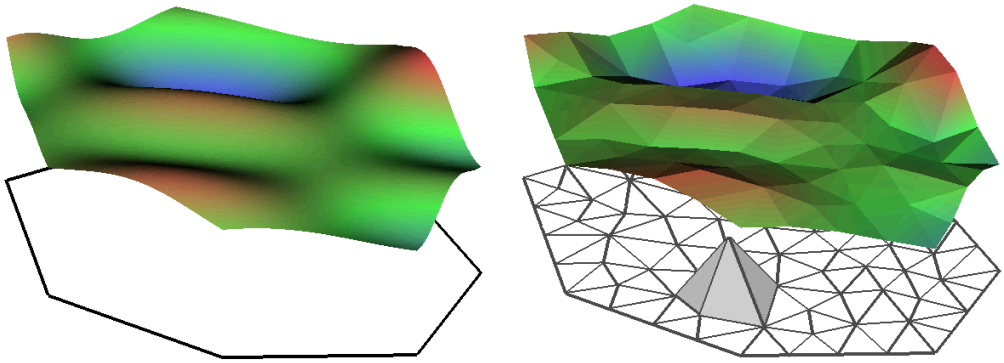


Figure 1: An example of a piecewise linear approximation to a surface. The grey pyramid is a representative basis function.

which can be attacked using standard numerical integration schemes. This may be done by admitting a tessellation on the observation window Ω and applying a Gaussian quadrature rule in each triangle. If the basis functions are chosen to be independent of the model parameters, we can compute and store the values of $\phi_i(s)$ at each integration point. Of course, this can require a lot of storage if the basis functions are not chosen to have compact support. The reason to prefer basis functions with compact support is that a numerical integration scheme requires the evaluation of the integrand in (3) at a large number of points in the observation window Ω. For each evaluation of the integrand, the number of terms in the sum is equal to the number of non-zero basis functions at the point at which it is being evaluated. Therefore, if the basis functions are non-zero everywhere, the evaluation of this integral will be very time consuming. We note that the basis functions used in kernel methods and predictive process-type models typically have global support, whereas the basis functions used in the SPDE models of Lindgren et al. (2011) always have very small supports.

5 Stochastic PDEs and Markov random fields

While process convolution and fixed-rank Kriging models can be written in the form (2), the corresponding basis functions do not typically have compact support and, therefore, evaluation of the likelihood will still be expensive. We will, therefore, use the SPDE models of Lindgren et al. (2011) to construct $Z(s)$. This choice also has the pleasing property of being strongly related to the modelling strategy of Illian et al. (2011), which is based on conditional autoregressive models.

The basic idea of Lindgren et al. (2011) is that, given a surface, an appropriate lower resolution approximation to the surface can be constructed by sampling the surface in a set of well designed points and constructing a piecewise linear interpolant (Figure 1). We will, therefore, take the basis functions in (2) to be a set of piecewise linear functions defined over a triangular mesh, which gives us more geometric flexibility than a traditional grid-based method. An important question is 'Which classical random fields will yield computationally efficient piecewise linear representations?' It turns out that the answer is a subset of the Matérn random fields, which are zero-mean Gaussian stationary, isotropic random fields with covariance function

$$c(h) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa h)^\nu K_\nu(\kappa h), \qquad h \geq 0,$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind, $\nu > 0$ is the smoothing parameter, $\kappa > 0$ is the range parameter, and σ^2 is the variance. The subset of Matérn random fields with efficient piecewise linear representations occur when $\nu + d/2$ is an integer, where d is the dimension of the space.

When $\nu + d/2$ is an integer, a computationally efficient piecewise linear representation can be constructed by using a different representation of the Matérn field $x(s)$, namely as the stationary solution to the stochastic partial differential equation (SPDE)

$$(\kappa^2 - \Delta)^{\alpha/2} x(s) = W(s), \tag{4}$$

where $\alpha = \nu - d/2$ is an integer, $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(s)$ is spatial white noise. This representation had first been constructed by Whittle (1954, 1963) while proving (among other things) that the classical second order conditional autoregression (CAR(2)) model limits to a Matérn field with $\nu = 1$.

Piecewise linear approximations to *deterministic* partial differential equations are commonly constructed in physics, engineering and applied mathematics using the finite element method and it is of no surprise that this method was also successfully used by Lindgren et al. (2011) to construct efficient representation of the appropriate Matérn fields. When $\alpha = 2$, the final outcome of their procedure replaces the slightly frightening SPDE (4) with the very pleasant equation for the weights in the basis expansion (2)

$$(\kappa^2 \mathbf{C} + \mathbf{G} - \mathbf{B})\mathbf{z} \sim N(\mathbf{0}, \mathbf{C}), \tag{5}$$

where $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{G}$ are sparse matrices with easily calculable entries (it's an exercise in primary school-level geometry!)

$$\begin{aligned} C_{ii} &= \int_{\Omega} \phi_i(s)\, ds, \\ G_{ij} &= \int_{\Omega} \nabla\phi_i(s) \cdot \nabla\phi_j(s)\, ds, \\ B_{ij} &= \int_{\partial\Omega} \phi_i(s) \partial_n \phi_j(s)\, ds, \end{aligned}$$

$\partial\Omega$ is the boundary of Ω, $\partial_n \phi_j(s)$ is the normal derivative of $\phi_j(s)$. A detailed presentation of these models can be found in (Lindgren et al., 2011), where it is also shown that these models lead exactly to the classical CAR models when computed over a regular lattice.

The presence of the matrix $\mathbf{B}$ in (5) is particularly interesting. This matrix encodes information about the process on the boundary of the observation window Ω. The effect of physical boundaries in spatial models has received very little attention in the literature (a notable example in the context of Bayesian smoothing is Wood et al. (2008)). For the remainder of this paper, we will set $\mathbf{B} = \mathbf{0}$, which corresponds to no-flux boundary conditions. We will discuss the interpretation of this condition in Section 8.3.

All of the discussion in this section is predicated on the idea that we can find good piecewise linear representations of a class of random fields. It is therefore imperative to construct a mesh over the observation window in a way that will not introduce excessive approximation bias. A major difference between point process models and other standard spatial models is that in a point process model *there is data everywhere*. As such, we are obliged to make our triangulation cover the space in a fairly regular way. On the other hand, we are unlikely to be able to infer the fine scale latent structure in areas where we don't have points or there has been little sampling effort. This suggests that a sensible meshing strategy would be to construct a regular triangulation of the observation window and refine it in areas where there are a large number of points.

6 Approximating the likelihood

With the continuous SPDE model in place, we are now in a position to attack the intractable likelihood (1). In this section, we will outline a procedure for approximating the likelihood that

extends the standard approximation to the non-lattice, unbinned data case. The log-likelihood

$$\log(\pi(y|Z)) = |\Omega| - \int_\Omega \exp(Z(s))\, ds + \sum_{i=1}^{N} Z(s_i)$$

consists of two terms: the stochastic integral, and the evaluation of the field at the data points. While the continuously specified SPDE models allow us to compute the sum term exactly, we will need to approximate the integral by a sum. Consider a deterministic integration rule of the general form

$$\int_\Omega f(s)\, ds \approx \sum_{i=1}^{p} \tilde{\alpha}_i f(\tilde{s}_i),$$

for fixed, deterministic nodes $\{\tilde{s}_i\}_{i=1}^p$ and weights $\{\tilde{\alpha}_i\}_{i=1}^p$. Using this integration rule, we can construct the approximation

$$\begin{aligned}\log(\pi(y|\boldsymbol{z})) &\approx C - \sum_{i=1}^{p} \tilde{\alpha}_i \exp\left(\sum_{j=1}^{n} z_j \phi_j(\tilde{s}_i)\right) + \sum_{i=1}^{N}\sum_{j=1}^{n} z_j \phi_j(s_i) \\ &= C - \tilde{\boldsymbol{\alpha}}^T \exp(\boldsymbol{A}_1 \boldsymbol{z}) + \mathbf{1}^T \boldsymbol{A}_2 \boldsymbol{z}, \end{aligned} \tag{6}$$

where C is a constant that doesn't depend on anything important, $[\boldsymbol{A}_1]_{ij} = \phi_j(\tilde{s}_i)$ is the matrix that extracts the value the latent Gaussian model (2) at the integration nodes $\{\tilde{s}_i\}$, and $[\boldsymbol{A}_2]_{ij} = \phi_j(s_i)$ is the matrix that evaluates the latent Gaussian field at the observed points $\{s_i\}$.

The advantage of approximating the log-likelihood by (6) is that it is of the standard Poisson form. In particular, given $\boldsymbol{z}$ and $\boldsymbol{\theta}$, the approximate likelihood consists of $N+p$ independent Poisson random variables. To see this, we write $\boldsymbol{\eta} = (\boldsymbol{z}^T \boldsymbol{A}_1^T, \boldsymbol{z}^T \boldsymbol{A}_2^T)^T$ and $\boldsymbol{\alpha} = (\tilde{\boldsymbol{\alpha}}^T, \mathbf{0}_{N\times 1}^T)^T$. Then, if we construct some fake 'observations' $\boldsymbol{y} = (\mathbf{0}_{p\times 1}^T, \mathbf{1}_{N\times 1}^T)^T$, the approximate likelihood factors as

$$\pi(\boldsymbol{y}|\boldsymbol{z}) \approx C \prod_{i=1}^{N+p} (\alpha_i \eta_i)_i^{y_i} e^{-\alpha_i \eta_i}, \tag{7}$$

which is just a product of conditionally independent Poisson random variables with mean $\alpha_i \eta_i$ and observed value y_i.

Numerical integration schemes that lead to likelihood approximations of the form of the form (7) were also considered by Baddeley and Turner (2000), under the unusually esoteric name of "Berman-Turner devices", for approximating pseudolikelihoods of Gibbs-type point processes. However, to the best of our knowledge, these ideas have not been extended to LGCPs, most probably due to the paucity of computationally efficient continuously specified Gaussian random field models. We also note that, in contrast to Baddeley and Turner (2000), we have strong opinions on the underlying numerical integration scheme. Rather than placing "one [...] point, either systematically or randomly", we strongly advise the reader to avail themselves of the last few decades (or arguably centuries) of numerical analysis and place the integration points at good (or even optimal) points! The simplest option is to attach to each node in the mesh a region V_i for which the value of the basis function $\phi_i(s)$ is greater than the value of any other basis function. This construction, shown in Figure 2, corresponds to the important notion of the *dual mesh*. The corresponding integration rule sets $\tilde{s}_i$ to be the node location and $\tilde{\alpha}_i = |V_i|$ to be the volume of the dual cell. It is easy to show that this approximation (known as the midpoint rule) is second order accurate. Of course, we can use the structure of the computational mesh in other ways when constructing the integrator. In particular, we could construct an integration scheme as the sum of (optimal) Gaussian integration rules on the individual triangles in the mesh. The weights and integration points for general triangles are well known and can be found in most books on numerical analysis or finite element methods (Ern and Guermond, 2004).

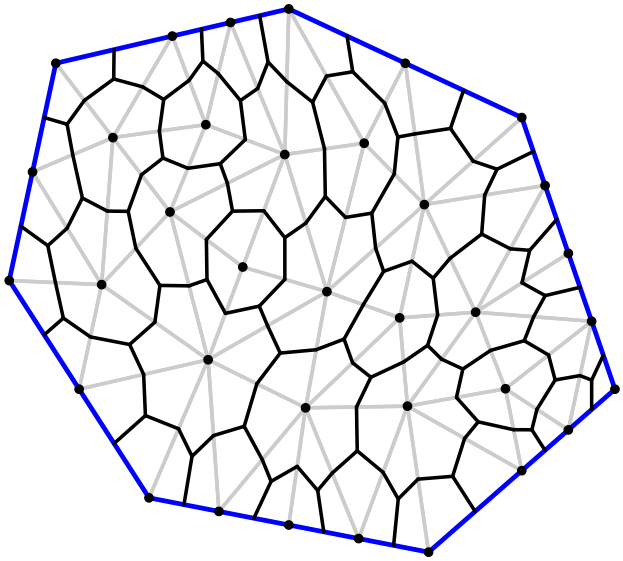


Figure 2: The dual mesh (black) is constructed by joining the centroids of the primal triangular mesh (grey). The volumes of these dual cells define the weights of an integration scheme based at the nodes of the primal mesh.

7 Implementation in R-INLA

The log-Gaussian Cox process model discussed above can be implemented efficiently within the R-INLA package, which is a user-friendly interface for the methods introduced by Rue et al. (2009) and Lindgren et al. (2011). This package allows us to quickly and accurately perform full Bayesian inference on LGCPs in *seconds* rather than the hours that an MCMC routine would require. Illian et al. (2011) argue that the R-INLA package significantly reduces the barrier faced by applied scientists when attempting to fit spatial point process models. To this end, we note that in spite of the varying complexity in the models considered, the code for solving them is essentially identical.

The first task is to construct the spatially continuous SPDE model. In order to do this, it is necessary to construct a mesh over the observation window. This proceeds by constructing a boundary object and using it to construct a mesh. This mesh can then be used to construct an SPDE model object, which contains the relevant information about the continuous model. The following snippet of code constructs a Matérn model with $\alpha = 2$ on the unit square.

```
> loc.bnd <- matrix(c(0,0, 1,0, 1,1, 0,1), 4, 2, byrow=TRUE)
> segm.bnd <- inla.mesh.segment(loc.bnd)
> mesh <- inla.mesh.create(boundary=segm.bnd, refine=list(max.edge=0.1))
> spde <- inla.spde.create(mesh, model="matern")
```

The `refine` parameter in the `inla.mesh.create` function tells the meshing software that the largest triangle edge in the mesh should be no larger than 0.1.

With the continuous Gaussian random field model in hand, we now turn to the problem of evaluating it at the appropriate points. It follows from (6) that we need two matrices: one to evaluate the field at the integration points and one to evaluate it at the observed points. As we are using a simple vertex-based integration scheme, the first evaluation matrix ($\boldsymbol{A}_1$) is the identity matrix. The second matrix can be constructed through a call to `inla.mesh.project`, which is a general piece of code for constructing evaluation matrices.

```
> LocationMatrix = inla.mesh.project(mesh, loc)$A
> IntegrationMatrix = sparseMatrix(i=1:nV,j=1:nV,x=rep(1,nV)) #nV is the size of the mesh
```

```
> ObservationMatrix=rBind(IntegrationMatrix,LocationMatrix)
```

Finally, we need to construct the likelihood for the data. Following (7), the approximate likelihood is Poisson and the "data" for the model is

```
> fake_data = c(rep(0,nV),rep(1,nData)) #nData is the no. of data points
```

In `R-INLA`, the Poisson likelihood takes the form

$$P(y = n) = \frac{\lambda^n}{n!} \exp(-\lambda),$$

where $\lambda(x) = E \exp(x)$. From (7), the integration weights are just the volumes of the control volumes surrounding each vertex (see Figure 2). Rather than re-compute these weights, we can use the convenient fact that they are simply the diagonal elements of the matrix $\boldsymbol{C}$ in (5).

```
> IntegrationWeights = diag(spde$internal$c0)
> E_point_process =c(IntegrationWeights,rep(0,nData))
```

With all of the ingredients for the model in place, we can now fit the model. The `INLA` call is as follows.

```
> formula = y ~ 1 + f(idx, model=spde)
> result = inla(formula,
  data=list(y = fake_data, idx = c(1:nV)),
  family="poisson",
  control.predictor=list(A=ObservationMatrix),
  E=E_point_process)
```

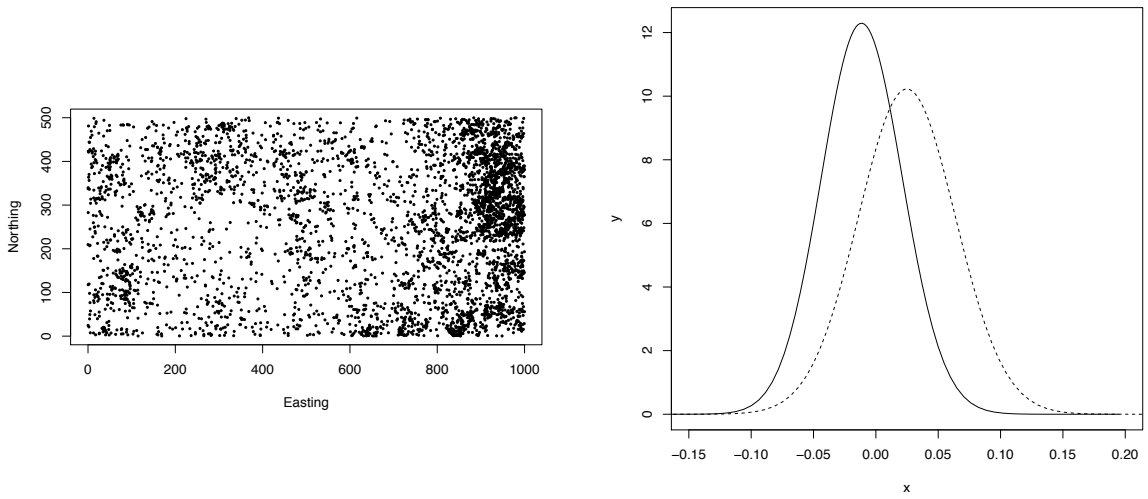
The only "non-standard" part of the `INLA` call is the `control.predictor` statement, which is required when there is a matrix (ObservationMatrix) linking the data to the latent field (Simpson et al., 2011a).

A working piece of code containing the snippets in this section can be found in `simulated.R` in the supplementary material, which simulates a LGCP on the unit square and performs the appropriate inference in INLA. The supplementary material also contains the code for the simulated examples below (Sections 8.2 and 8.3). In order to run these examples it is necessary to update `R-INLA` to the latest version with the command

```
> inla.upgrade(testing=TRUE)
```

8 Examples

In this section we will consider the application of log-Gaussian Cox processes in three increasingly complicated situations and demonstrate the applicability of our methods. The first case study is a rather pedestrian example of a LGCP with covariates fitted to a real data set observed everywhere in a rectangular area. This is exactly the situation considered in Rue et al. (2009) and Illian and Rue (2010), who use the standard LGCP approximation on a lattice. The second example is a simulation study in the vein of Chakraborty et al. (Submitted), where the point pattern is incompletely observed due to varying sampling effort across the region of interest. The third case study is where, in our opinion, the power of our method is best showcased. Inspired by the problem of mapping the risk associated with freak waves on oceans, we have constructed a point process defined only on the world's oceans. This is a point process over a very irregular, multiply connected bounded region on a sphere and, to the best of our knowledge, there is no other method that can solve this problem.



(a) Data

(b) A regular mesh

Figure 3: (Left) The location of *Protium tenuifolium*. (Right) The posterior for the effect of Phosphorus. The dashed line shows the posterior covariate weight for the lattice method, while the solid line corresponds to the approach described above.

8.1 A vanilla example: Rainforest data

In this case study we deal with a standard application of spatial point process models: species association with soil properties in tropical rainforests. The complete data set consists of the location of all trees with diameter at breast height of 1cm or greater of a total of 319 species within a 50 ha plot in a rainforest plot on Barro Colorado Island in Panama that has never been logged. We model the large spatial pattern formed by 4294 trees of species *Protium tenuifolium*, shown in Figure 3(a), relative to the covariate phosphorus (Condit, 1998; Hubbell et al., 1999, 2005). The plot on Barro Colorado Island is only one plot within a network of a large network of 50 ha plots that have been established as part of an international effort to understand species survival and coexistence in biodiverse ecosystems (Burslem et al., 2001).

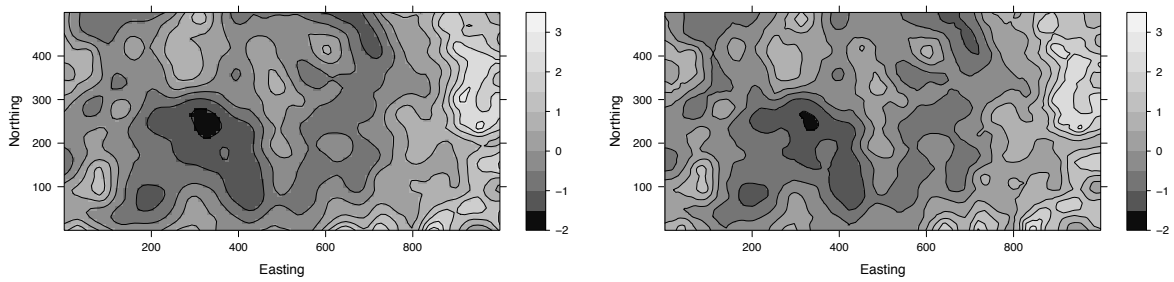
Data sets with a similar structure have recently been analysed in the literature both with descriptive (Law et al., 2009) and model-based approaches (Waagepetersen and Guan, 2009; Waagepetersen, 2007; Wiegand et al., 2007). Rue et al. (2009) as well as Illian and Rue (2010) use INLA to fit a log-Gaussian Cox process to similar data, while Illian et al. (2011) fit a joint model to both the pattern and covariates. In this paper we fit a simple model, where the latent field is given by

$$Z(s) = \mu + \beta P(s) + x(s),$$

where μ is a constant mean, $P(s)$ is a spatially varying covariate describing the level of phosphorus in the soil and $x(s)$ is an SPDE model with $\alpha = 2$. For the purpose of comparison, we fit a lattice model with linear predictor

$$\boldsymbol{z} = \mu\mathbf{1} + \beta\boldsymbol{P} + \boldsymbol{x},$$

where $\mathbf{1}$ is a vector of ones, $\boldsymbol{P}$ is a phosphorous covariate, $\boldsymbol{x}$ is a CAR(2) model. Both of the models required around 12 seconds to run in `R-INLA`. The posterior means for the spatial random effects are shown in Figure 4 and they are approximately the same. The posteriors for the effect of soil phosphorous on the location of trees are shown in Figure 3(b).



(a) Lattice method

(b) SPDE approach

Figure 4: This figure shows the two spatial effects obtained when using different point process models to model the point pattern formed by trees. The figure on the left was obtained using the standard lattice method, while the figure on the right was constructed from the method introduced above.

8.2 A more complex example: Incorporating variable sampling effort

One of the major challenges when applying spatial point process models to real data sets is that the point pattern is very rarely captured exactly (Chakraborty and Gelfand, 2010; Chakraborty et al., Submitted). As such, it is important that sampling effort is included in the observation process. In this example, we will consider the case where there is a hole in the data set: an area in which there have been no measurements, but we expect that presences are possible. This type of situation occurs, for instance, when considering the spatial distribution of an animal species over an area that contains an area that is impossible to survey for topographical or political reasons (Elith et al., 2006). In a related situation, data sampling effort varies spatially and is higher in areas where the scientists expect a good chance of presence (c.f. the preferential sampling model of Diggle et al., 2010).

Following Chakraborty et al. (Submitted), we include *known* sampling effort in our model by writing the intensity as

$$\lambda(s) = S(s)\exp(Z(s)),$$

where $S(s)$ is a known function describing the sampling effort at location s. In this example, we will assume that the point pattern has been observed perfectly except in a rectangle (see Figure 5(a)), where the pattern is not observed. We therefore define $S(s)$ to be zero inside this rectangle and one everywhere else. It is straightforward to see from (1) that, with this choice of $S(s)$, the unsampled area does not contribute to the integral in the likelihood. We can therefore choose the mesh to be quite coarse in this area, as long as it does not adversely affect the SPDE approximation to the random field. Figure 5(b) shows a mesh that has been coarsened in a rectangular region corresponding to a hole in the sampling effort.

Practically speaking, the changes necessary to add sampling effort to basic point process code used in the previous section are minimal. The only change is in the `E` parameter in the INLA code, the line now reads

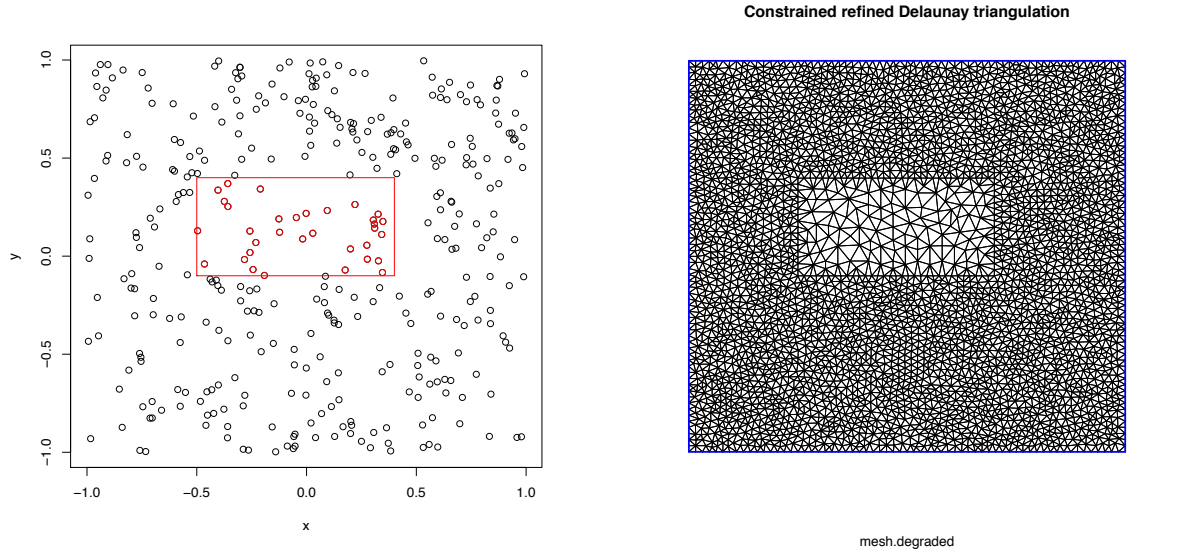
```
E_point_process =c(IntegrationWeights*sampled,rep(1,nData))
```

where `sample` is a 0/1 variable that indicates if the mesh node is in the sampled area. This method can be extended in a straightforward manner to cover more complicated designs, although Chakraborty et al. (Submitted) suggest it is necessary to assume that the design is known.

In order to test our method on this type of problem, we have simulated a log-Gaussian Cox process on $[-1, 1] \times [-1, 1]$ and removed the points from the rectangle $[-0.5, 0.4] \times [-0.1, 0.4]$ to simulate the variable sampling mentioned above. The simulated data set is shown in Figure 5 and the difference in the posterior mean generated from the full data and the censored data is shown in Figure 6. There is very little difference between the two posterior means outside of the censored area, whereas there are, unsurprisingly, missing features from within the censored area. From our point of

view, the most interesting figure in this example is Figure 7. We ran the analysis using two different meshes with the same maximum edge length. The first mesh (dotted lines) was a regular lattice that covered the entire domain and contained 4225 points. The second mesh (dashed lines) is an irregular mesh consisting of 3850 points that was de-refined in the censored area shown in Figure 5(b). Figure 7 compares the posterior marginals for the parameters for these two meshes and it can be seen that they are, for all intents and purposes, identical. It is, however, important to have some points inside the censored area to ensure that the random field behaves properly—it would be incorrect to simply remove the censored area from the mesh.

Computing on the mesh that was correctly adapted to the problem unsurprisingly resulted in a significant decrease in computational time. With the regular grid, the full inference took 114.55 seconds on a 2009 Macbook Pro, whereas the computation on the irregular mesh required only 81.02 seconds—a saving of 29%! This justifies our statement that automatically using regular lattices is computationally wasteful.



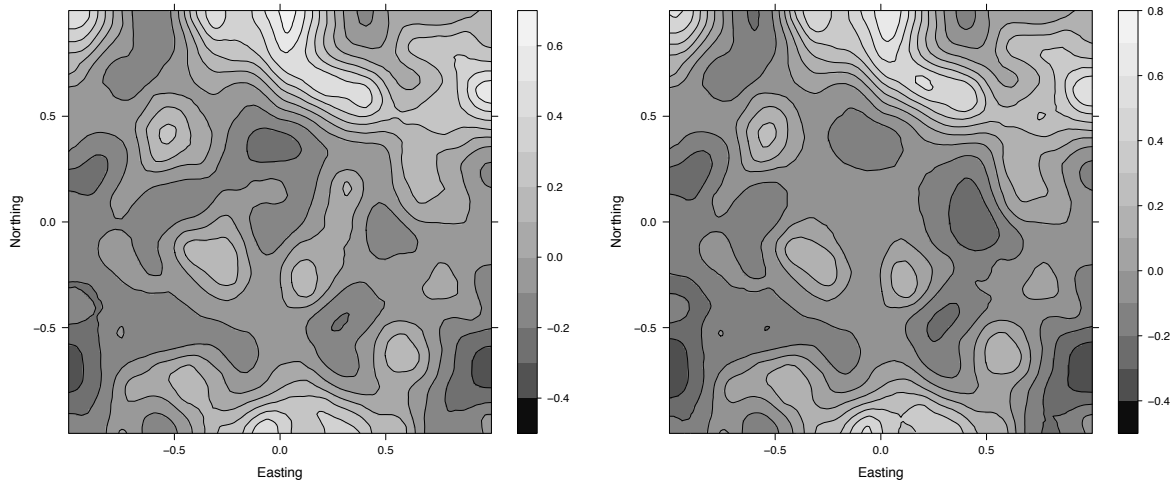
(a) Simulated data

(b) An mesh that takes into account the sampling effort

Figure 5: (Left) Simulated data with a hole in the sampling effort. The red rectangle borders the area in which there was no sampling, and the red circles show the points that were missed due to incomplete sampling. (Right) A mesh that takes into account the lack of sampling effort in the rectangular region.

8.3 An even more complex example: A point process over the ocean

The previous two case studies have been observed over quite simple observation windows: namely rectangles. However, point processes often occur over complicated domains and the topology, topography and geometry of the domain will typically be significant when modelling the covariance structure (c.f. the discussion of Wood et al. (2008) in the context of spatial smoothers). For this case study, we have simulated a LGCP on the oceans. Point processes over the oceans would be useful, in particular, when modelling the risk of freak waves. To the best of our knowledge, no existing methods for point processes can be used directly to attack this problem. In contrast, the methodology developed in this paper can be applied *with essentially no change*! In fact, the most difficult task was finding a suitable map for the boundaries of the continents. As we are only working with simulated data, we built very basic continent outlines that were sufficient to show that it is possible to perform inference on a log-Gaussian Cox process over the oceans.



(a) Complete data

(b) Incomplete data

Figure 6: This figure shows the posterior mean of the spatial effect when using the full simulated point pattern (left) and the partially observed point pattern (right). We note that the large scale features of both fields are very similar in areas at which the point pattern was sampled.

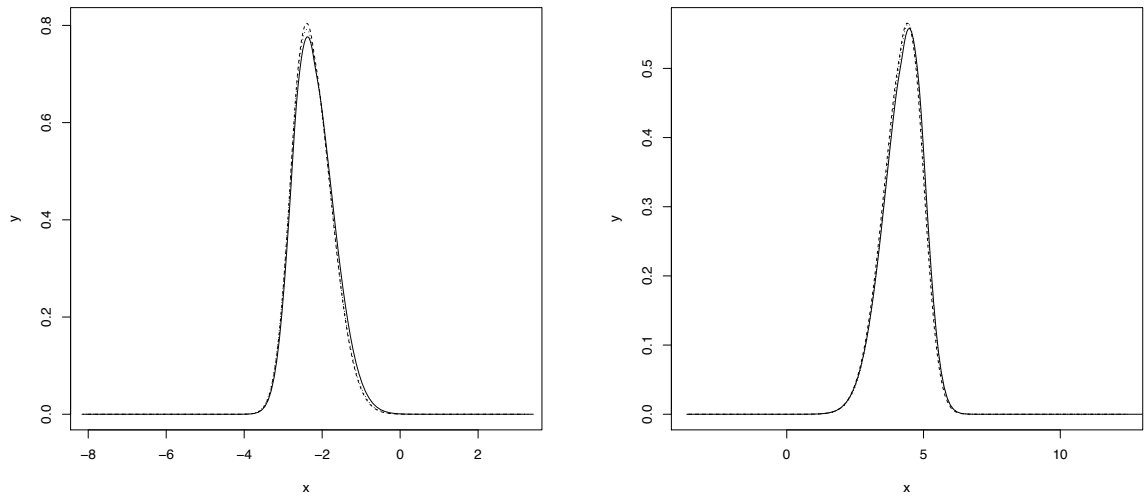
The oceans form a non-convex, multiply connected bounded region on the sphere and it is, therefore, necessary to construct a Gaussian random field model over this region. The main complication, beyond those considered by Lindgren et al. (2011) is that we need a model for the covariance at the boundary. This is a difficult issue that has essentially been ignored in the statistics literature. As we are working with simulated data, we have the luxury to choose a relatively simple, yet still realistic boundary model. In particular, as we would expect that wave heights vary more near the coast than in the deep ocean and, as the designation of a "freak wave" is relative to the expected wave height, we will require that the random field has more uncertainty near the boundary. This can be achieved very naturally using Neumann boundary conditions. The easiest way to see this is to consider the one dimensional case where, using Theorem 1 in Appendix A.4 of Lindgren et al. (2011), we can see that the variance of the one dimensional SPDE model with $\alpha = 3/2$ (which corresponds to an $\alpha = 2$ model in 2D) and Neumann boundary conditions on $[0, 1]$ is

$$\text{cov}(x(s), x(s)) = 2 \sum_{k=0}^{\infty}{}' r_m(2k) + \sum_{k=-\infty}^{\infty} r_m(2|s-k|),$$

where $r_m(s)$ is the isotropic covariance function corresponding to the model on $\mathbb{R}$ and the dash on the first sum indicates that the $k = 0$ term is halved. From this expression, it is clear that, for sufficiently large κ, the variance is approximately constant (and equal to the first sum) in the centre of $[0, 1]$ and it doubles at the endpoints. A similar result holds over rectangular regions in 2D,

The simulated point process is shown in Figure 8, which was constructed by simulating a Gaussian random field associated with the mesh in Figure 8(b). The resulting point pattern has 1142 points. With essentially no changes to the code, the inference was performed on this model and the posterior mean is shown in Figure 9(b). The posterior mean shows the same large-scale features as the sample that was used to generate the LGCP (Figure 9(a)), with the expected loss of information due to the uninformative nature of point pattern data.

Some interesting effects induced by the boundary conditions can be seen in Figure 10. The pointwise standard deviation of the posterior latent Gaussian field is shown in Figure 10(a). The standard deviation is reasonably constant away from the coasts, whereas it is much higher near the



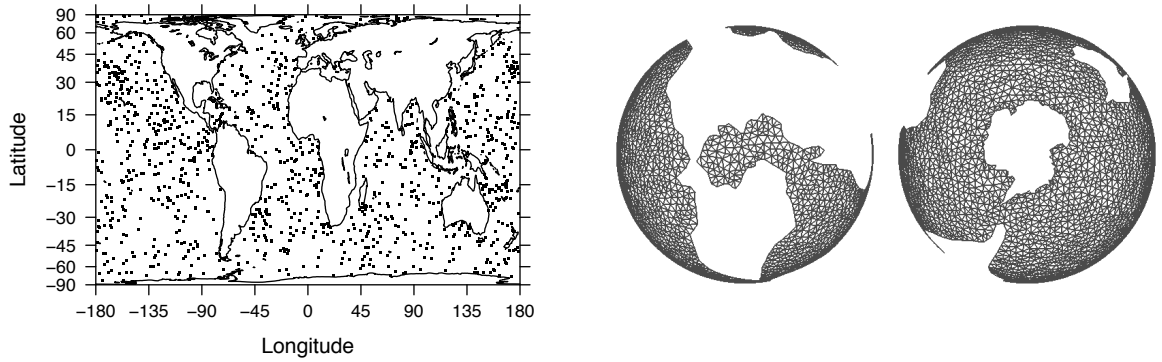
(a) Posterior for $\log(\tau)$

(b) Posterior for $\log(\kappa^2)$

Figure 7: This figure shows the effect of coarsening the mesh on the posterior estimates of the parameters. The dashed line corresponds to the anisotropic mesh in Figure 5(b), while the dotted line corresponds to using a regular grid with the same maximum edge length as the fine portion of the anisotropic mesh. For comparison purposes, we have plotted the posterior generated from the correctly observed point pattern (solid line).

boundaries. There are also some interesting effects in the Gulf of Carpentaria (Australia) and the North Sea (between the UK and Scandinavia). This is an effect of the prior model, which increases the variance near the boundaries and in areas with high curvature.

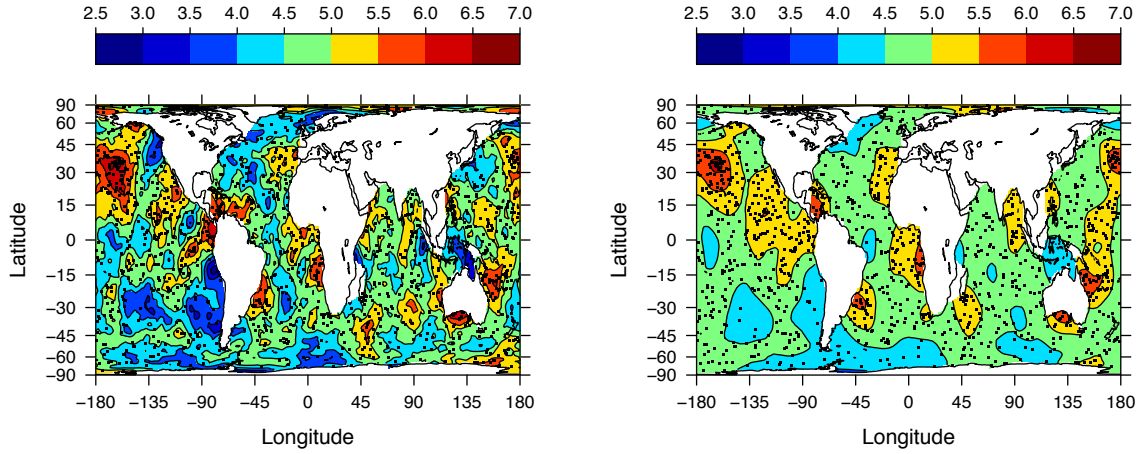
In the context of freak wave modelling, Figure 10(b), which shows probability that the log-risk will be greater than 5.5, is probably the most important result. This type of map can easily be computed using the function `inla.pmarginal`. Once again we see pronounced effects near the coastlines.



(a) Simulated data

(b) A mesh over the oceans

Figure 8: (Left) A simulated log-Gaussian Cox processes over the oceans. (Right) A mesh that covers the oceans.



(a) True surface

(b) Reconstructed surface

Figure 9: This shows the true sample from the latent Gaussian random field used to generate the sample in Figure 8 (Left) and the posterior mean of the latent spatial effect (Right). We note that the large scale behaviour is the same for both figures.

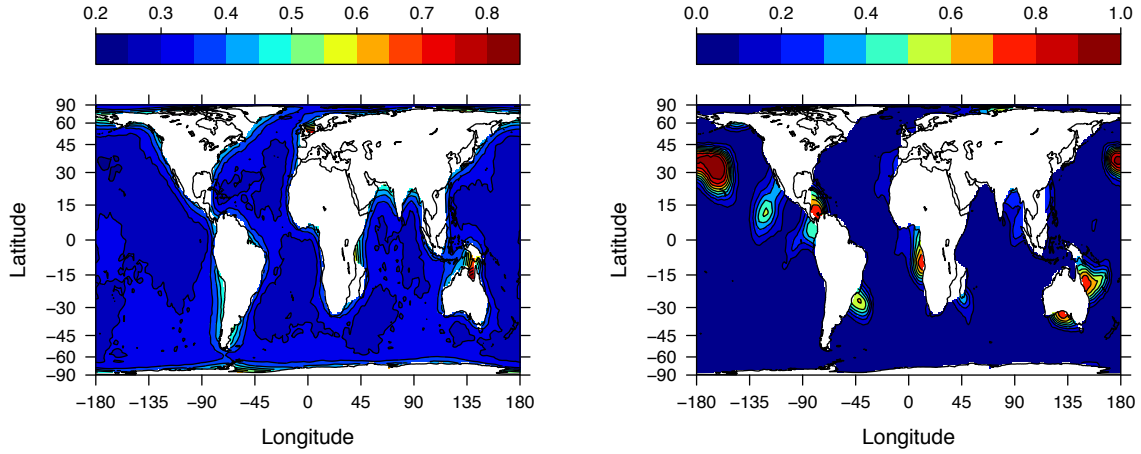
9 Discussion and future work

In this paper we have developed a new, computationally efficient approximation to log-Gaussian Cox processes that bypasses the requirement that they have to be defined over a regular lattice. Furthermore, by exploiting the computational and modelling advantages of the stochastic partial differential equation models of Lindgren et al. (2011), we are able to attack a variety of interesting new problems. We note that the approximation introduced above is also valid when using kernel methods (Higdon, 1998), predictive processes (Banerjee et al., 2008) or fixed-rank Kriging (Cressie and Johannesson, 2008). The problem with using these methods in this context is that their basis functions are typically non-local and, therefore, the point evaluation matrices $\boldsymbol{A}_i$ in (7) are dense (see Simpson et al., 2011b, for a further discussion of the choice of basis functions in spatial statistics).

In Section 8.3, we considered a point process over a bounded region of the sphere. To the best of our knowledge, there are no other applicable inference methods for this problem. As such, there is also no work on modelling boundary effects for point process models, and precious little work done even in the general spatial statistics literature on this problem. Therefore, an interesting and challenging problem is the construction of good boundary models. We have argued heuristically that Neumann, or no-flux, boundary conditions increase the variance at the boundaries. Similarly, it can be easily seen that Dirichlet boundary conditions, which corresponding to fixing the value of the field on the boundaries, decrease the variance. It would be interesting to study the effect of other boundary conditions in the statistical context.

There is work to be done on the theoretical properties of the approximation presented in this paper. We would expect convergence results similar to those of Waagepetersen (2004) for the lattice case. However, we would also hope for sharper asymptotic results, as the approximation properties of the finite element basis provide us with good bounds on the convergence of the underlying random field model (Simpson et al., 2011b). Similarly, the interplay between the random field approximation, the numerical integration scheme and the posterior accuracy should be investigated. The ultimate aim of this theory would be to construct *spatially adaptive* methods for inferring log-Gaussian Cox processes. The computational savings from the simple *a priori* adaptation in Section 8.2 suggests that this would be a worthwhile idea to pursue.

Finally, it is worth noting that the approximation in Section 6 applies even when the latent



(a) Posterior standard deviation

(b) Posterior risk map

Figure 10: The pointwise posterior standard deviation for the log risk surface (Left). The risk map $P(\log(\lambda(s)) > 5.5)$ computed from the posterior marginals (Right).

random field $Z(s)$ is not Gaussian. The only requirement is that it has the basis function expansion (2) and that the statistical properties of $\boldsymbol{z}$ is known. In particular, this approximation applies to SPDE models with non-Gaussian noise. This has been investigated for type-G Lévy processes, and especially for Laplace random fields, by Bolin (2011). Similarly, replacing Gaussian white noise with Poisson noise would result in shot-noise Cox process models of the Matérn type. It would be very interesting to investigate these models!

Acknowledgements

The authors gratefully acknowledge the financial support of Research Councils UK for Illian.

The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651,DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042,DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

References

A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point processes. *New Zealand Journal of Statistics*, 42:283–322, 2000.

S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang. Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society, Series B*, 70(4):825–848, 2008.

D. Bolin. Spatial matérn fields driven by non-gaussian noise. Technical Report 2011:4, Lund University, 2011.

D. F. R. P. Burslem, N. C. Garwood, and S. C. Thomas. Tropical forest diversity – the plot thickens. *Science*, 291:606–607, 2001.

A. Chakraborty and A. Gelfand. Analyzing spatial point patterns subject to measurement error. *Bayesian Analysis*, 5(1):847–872, 2010.

A. Chakraborty, A. E. Gelfand, A. M. Wilson, A. M. Latimer, and J. A. S. Jr. Point pattern modelling for degraded presence-only data over large regions. *Journal of the Royal Statistical Society, Series C*, Submitted.

R. Condit. *Tropical Forest Census Plots.* Springer-Verlag and R. G. Landes Company, Berlin, Germany, and Georgetown, Texas., 1998.

N. A. C. Cressie and G. Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B*, 70(1):209–226, 2008.

P. Diggle, R. Menezes, and T. Su. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2):191–232, 2010.

J. Elith, C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Marion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Sobero'on, S. Williams, M. S. Wisz, and N. E. Zimmermann. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29:129–151, 2006.

A. Ern and J. Guermond. *Theory and practice of finite elements.* Springer, New York, 2004.

D. Higdon. A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, 5(2):173–190, 1998.

S. P. Hubbell, R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, and S. L. de Lao. Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science*, 283:283: 554–557, 1999.

S. P. Hubbell, R. Condit, and R. B. Foster. Barro Colorado Forest Census Plot Data, 2005. URL `http://ctfs.si/edu/datasets/bci`.

J. B. Illian and H. Rue. A toolbox for fitting complex spatial point process models using integrated Laplace transformation (INLA). Technical report 6, Department of mathematical sciences, Norwegian University of Science and Technology, 2010.

J. B. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns.* Wiley, Chichester, 2008.

J. B. Illian, S. H. Sørbye, and H. Rue. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Submitted*, 2011.

R. Law, J. B. Illian, D. F. R. P. Burslem, G. Gratzer, C. V. S. Gunatilleke, and I. A. U. N. Gunatilleke. Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology*, 97:616–628, 2009.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 73(4):423–498, September 2011.

J. Møller and R. Waagepetersen. *Statistical inference and simulation for spatial point processes*, volume 100 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London, 2003.

J. Møller, A. R. Syversveen, and R. P. Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482, 1998.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392, 2009.

D. Simpson, F. Lindgren, and H. Rue. Fast approximate inference with inla: the past, the present and the future. In *ISI 2011*, 2011a.

D. Simpson, F. Lindgren, and H. Rue. Think continuous: Markovian models in spatial statistics. Technical Report 2011/09, Department of Mathematical Sciences, Norwegian University of Science and Technology, 2011b.

R. Waagepetersen. Convergence of posteriors for discretized log Gaussian Cox processes. *Statistics & Probability Letters*, 66(3):229–235, 2004.

R. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society, Series B*, 71:685–702, 2009.

R. P. Waagepetersen. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63(1):252–258, 2007.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41(3/4):434–449, 1954.

P. Whittle. Stochastic processes in several dimensions. *Bull. Inst. Internat. Statist.*, 40:974–994, 1963.

T. Wiegand, S. Gunatilleke, N. Gunatilleke, and T. Okuda. Analysing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology*, 88:3088–3012, 2007.

S. N. Wood, M. V. Bravington, and S. L. Hedley. Soap film smoothing. *Journal of the Royal Statistical Society, Series B*, 70(5):931–955, 2008.

G. Xia and A. Gelfand. Stationary process approximation for the analysis of large spatial datasets. ISDS Discussion Paper 2005-24, Duke University, Durham, NC, 2005.