Progress this week
December 1, 2019

- Added a figure 3 illustrating SPDE weighting.
- Added an outline of simulation ideas in §3.1.
- Added an outline of the example analysis in §3.2.
- Began writing code for §3.2.

Previous word count: 2068 **Current word count:**
Previous page count: 5.5 **Current page count:**

REVIEW ARTICLE

The Integrated Nested Laplace Approximation applied to Spatial Point Process Models

Kenneth Flagg and Andrew Hoegh

Montana State University, Bozeman, MT

ARTICLE HISTORY
Compiled December 6, 2019

ABSTRACT
This template is for authors who are preparing a manuscript for a Taylor & Francis journal using the LaTeX document preparation system and the `interact` class file, which is available via selected journals' home pages on the Taylor & Francis website.

KEYWORDS
INLA, spatial prediction, log-Gaussian Cox process, spatial point process

1. Introduction

Spatial prediction is a high-dimensional inference problem. When the goal of statistical modeling is to produce a graphical map of a random variable over space, the model ultimately must be able to predict that random variable at every pixel of the image. A map image will typically be at least several hundred by several hundred pixels, so in total there can easily be hundreds of thousands of pixels requiring predictions. Thus, even when a model has only half a dozen parameters, it may include hundreds of thousands of latent variables.

Spatial point process models further complicate the situation with difficult likelihoods. *(Cite some computational papers — Baddely?)* Both maximum likelihood and Bayesian model fitting require integrating the intensity function over space, but the integral is generally not available in closed form. Many methods have been introduced including quadrature-based approximations *(cite Baddeley)*, pseudodata approaches *(cite Baddeley/Berman/Turner etc)*, and Markov chain Monte Carlo [8].

Development of the integrated nested Laplace approximation (INLA) has made accurate approximate model fitting considerably more feasible for a particular class of log-Gaussian Cox process (LGCP) models. INLA was developed to fit Bayesian hierarchical models with many latent Gaussian variables [10]. A key part of INLA's computational simplicity is that it calculates the posterior distribution of each latent Gaussian variable one at a time; that is, it provides only the posterior marginal distributions rather than the full joint distribution.

When using a LGCP for spatial mapping, two aspects make INLA a suitable approach. First, the LGCP is driven by a spatial Gaussian process (GP), so the latent variables are Gaussian. Second, even though the latent variables are expected to ex-

CONTACT Kenneth Flagg. Email: kenneth.flagg@montana.edu

hibit spatial dependence, their full joint distribution is not needed. In most situations it suffices to map their predicted values, variance, and upper and lower interval bounds pointwise across space.

This article provides a review of recent advances in the fitting of LGCP models via INLA, including dimension reduction by triangulation, a likelihood factorization that avoids gridding, and incorporation of sampling or false negatives.

1.1. Log-Gaussian Cox Process

The LGCP is a Poisson process driven by a latent Gaussian process [9]. A Poisson process is characterized entirely by its intensity function $\lambda(\mathbf{s})$, which gives the mean number of events per unit area, and the process satisfies the following two properties *(find a good citation).*

(1) The number of events in a region $\mathcal{S}$ follows a Poisson distribution with mean $\int_{\mathcal{S}} \lambda(\mathbf{s}) d\mathbf{s}$.
(2) The numbers of events in disjoint regions are independent.

Commonly, covariates are incorporated via a log-linear model for the intensity,

$$\log \lambda(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta}.$$

The LGCP adds another stochastic layer,

$$\log \lambda(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + Z(\mathbf{s}),$$

where Z is a Gaussian process.

Where the spatial Poisson process has a (stochastically) fixed intensity to be estimated from data, the spatial LGCP induces a hierarchical model where the intensity function is itself a random process to be predicted.

1.2. Integrated Nested Laplace Approximation

INLA was developed with the goal of providing fast, accurate, deterministic approximations for posterior marginal distributions of the parameters in latent Gaussian models [10]. The setting is Bayesian generalized additive models but is kept very general; this includes linear models, models with nonlinear (spline, random walk, etc.) functions of predictors, and models with temporal and/or spatial dependence. Such models commonly have a large number of Gaussian latent variables or parameters with Gaussian priors and relatively few non-Gaussian parameters. The INLA approach makes repeated use of Laplace expansion, numerical integration, and numerical search. The method has established usefulness for LGCP [5]. It is readily implemented using the standalone INLA software or in R via the R-INLA package [6].

2. Methodology

INLA provides an efficient computational framework for fitting Bayesian models with latent Gaussian variables. On top of this framework are built several tools to further simplify the fitting of spatial LGCP models. The stochastic partial differential equation (SPDE) approach provides dimension reduction for the spatial GP [7]. The SPDE approach employs a numerical integration scheme which can also be used to approximate the LGCP likelihood and negate the need to grid the events into Poisson counts [11]. With these computational improvements, researchers are now able to efficiently fit LGCP models to incompletely-observed point patterns [12].

2.1. The SPDE Approach

Because the LGCP includes a Gaussian process (GP), efficient computation for Gaussian processes is critical when working with LGCP models.

The GP imposes a dense covariance matrix on the latent variables [2]. For GPs with a Matérn covariance function, a Gaussian Markov random field (GMRF) approximation can simplify computation, requiring only a sparse covariance structure.

The GMRF approximation is motivated by the fact that Gaussian fields with Matérn covariances are solutions to the below stochastic partial differential equation (SPDE) [7].

$$(\kappa^2 - \Delta)^{\alpha/2} Z(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \qquad \mathbf{s} \in \mathbb{R}^d, \quad \kappa > 0, \quad \alpha = \nu + d/2, \quad \nu > 0$$

Here, $\mathcal{W}$ is a Gaussian white noise process with variance 1, and Δ is the Laplacian operator. The stationary solution Z is a Gaussian field with a Matérn covariance function with scaling parameter κ (approximately inversly proportional to the range) and smoothness parameter ν. The variance is a function of κ, ν, and d.

Lindgren, Rue, and Lindström investigate the limit as $\nu \to 0$ for $d = 2$, finding that the solution is a GMRF on a unit lattice [7]. They then construct approximations for positive integer values of ν by ν-fold convolution of Z with itself. Finally, they use a finite element method to genrealize the approximation to arbitrary triangulations of the support. This approximation has considerable computational benefits because the GMRF has a sparse covariance structure; the only nodes with nonzero covariances are those directly connected by edges in the triangulation.

In practice, the SPDE approach goes as follows. Choose nodes $\mathbf{s}_i$ at which to model $Z(\mathbf{s}_i)$, then build a triangular mesh using these nodes. Typically the nodes will include locations where data or covariates are available, then the rest will be filled in with a Delaunay triangulation under some edge length constraints. The $Z(\mathbf{s}_i)$ are be modeled as a GMRF where the distribution of each $Z(\mathbf{s}_i)$ depends only on the $Z(\mathbf{s}_j)$ where $\mathbf{s}_i$ and $\mathbf{s}_j$ are connected by an edge. The GMRF representation is assumed to be a piecwise linear approximation of the continuous Gaussian field; values of $Z(\mathbf{s})$ for $\mathbf{s}$ not in the set of nodes are predicted using linear interpolation using the barycentric coordinates of $\mathbf{s}$.

The SPDE approach is implemented in the `R-INLA` package along with tools for constructing meshes. A wrapper for easily fitting LGCP models is provided in the `inlabru` package [1].

2.2. Going Off the Grid

Point pattern data are sometimes known as presence-only data, underscoring the fact that information about where events did *not* occur is both important and often overlooked. There have been many proposed methods to account for regions that were observed to contain no events (in contrast to unobserved regions where it is unknown

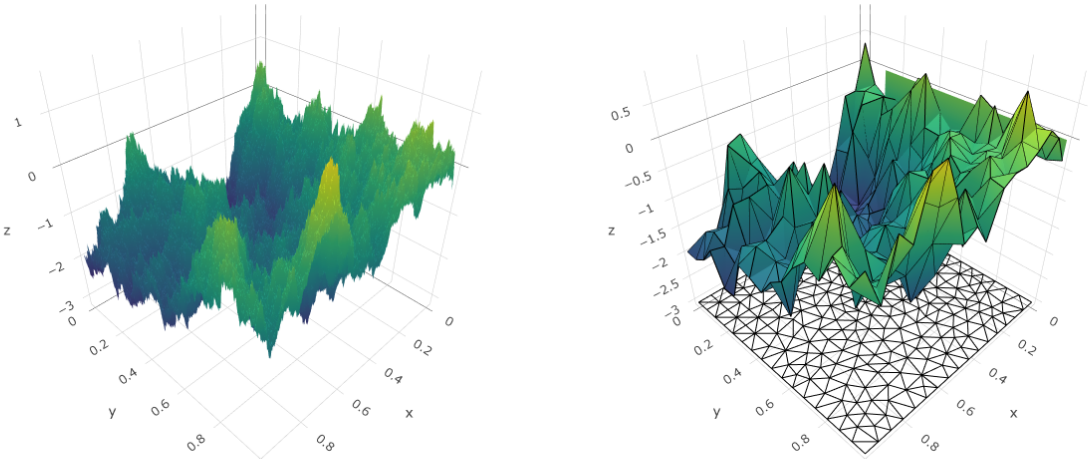

Figure 1. A realization of a spatial Gaussian process (left) and an approximation of that realization over a triangular mesh (right).

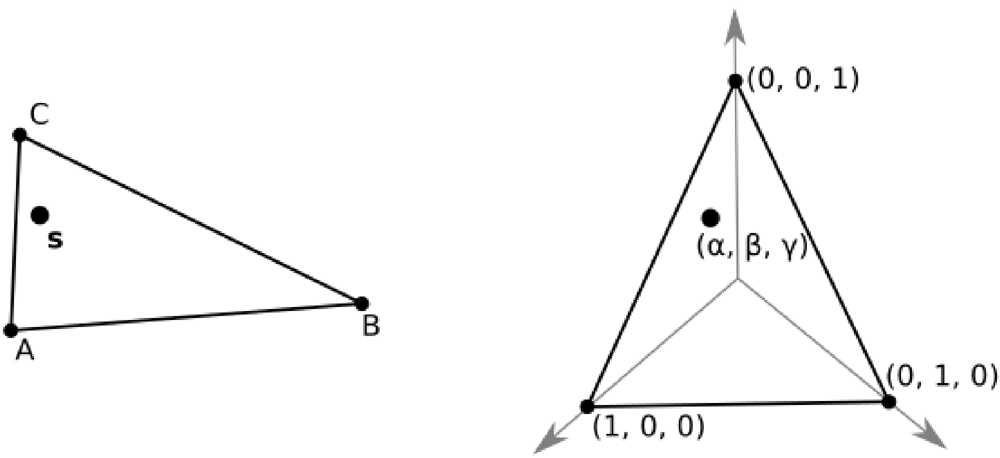

Figure 2. An illustration of the transformation from a mesh triangle to the simplex. (α, β, γ) are the barycentric coordinates of $\mathbf{s}$

if any events are present). Many of these methods involve imputation of dummy points or discretization. In the world of maximum likelihood, perhaps the most well-developed of these use approximations based on logistic regression on presence/absence information in small disjoint regions *(cite a bunch of Baddely etc papers)*.

Another alternative, which is probably the most popular approach to Bayesian fitting of LGCP models but is also common in frequentist analyses, is to grid the domain and model the induced Poisson counts. It has long been understood that results are sensitive to the discretization scheme [3]. Simpson et. al. (2016) explain that this is also computationally wasteful [11].

Simpson et. al. took LGCP inference "off grid" by intoducing a computationally-efficient approximation to the Poisson process likelihood that requires the intensity function only to be evaluated at the locations of observed events and at the nodes of a mesh. Thus, the SPDE approach can be employed to model the intensity surface and the same nodes reused in evaluation of the Poisson process likelihood. The result is a substantial improvement in both computing time and accuracy of the approximation compared to gridding.

The approximation arises from a factorization of the Poisson processes likelihood. The exact log-likelihood is

$$\ell(\lambda) = C - \int \lambda(\mathbf{s})d\mathbf{s} + \sum \log\left[\lambda(\mathbf{s_i})\right]$$

where C is a normalizing constant. The log-intensity is projected into the space spanned by a finite set of basis function representation is used for $\log[\lambda(\mathbf{s})]$, namely

$$\log\left[\lambda(\mathbf{s})\right] \approx \sum_j z_j \phi_j(\mathbf{s})$$

with $\mathbf{z} = (z_1, \ldots, z_m)'$ a multivariate normal random vector and $\{\phi_1, \ldots, \phi_m\}$ a set of linearly independant basis functions. This defines a numerical integration scheme at nodes $\tilde{\mathbf{s}}_i$ with weights $\tilde{\alpha}_j$ and yields the approximation,

$$\begin{aligned}\ell(\lambda) &\approx C - \sum_i \tilde{\alpha}_i \exp\left[\sum_j z_j \phi_j(\tilde{\mathbf{s}}_i)\right] + \sum_i \sum_j z_j \phi_j(\mathbf{s}_i) \\ &= C - \tilde{\boldsymbol{\alpha}}' \exp\left[\mathbf{A}_1 \mathbf{z}\right] + \mathbf{1}' \mathbf{A}_2 \mathbf{z}.\end{aligned}$$

This is equivalent to the likelihood of independent Poisson random variables with means *(translate the confusing algebra from the appendix)*. Thus, the SPDE approach, which can be easily fit in R-INLA, can be combined with a Poisson GLM to rapidly fit a LGCP model. There is no need to compute grid counts, only to define dummy points at the mesh nodes.

2.3. Variable Sampling Effort

There has long been a gap between point process modeling theory and practice, where point process models are only fit to data from completely-observed domains under the assumption that every event was detected perfectly. In practice, this is not the case as it may be impossible or impractical to census the entire region of interest. For example, line-transect surveys routinely generate point pattern data but are analyzed after aggregation rather than using a point process model. Another issue is that of false negatives, in other words events which exist but are not detected during the survey. False negatives are an accepted part of species abundance surveys, where the species of interest may be camouflaged or hidden in thick plant cover.

The idea of the incompletely-observed domain has been around for some time. Brix and Möller (2001) fit an LGCP model to weed data observed in rectangular frames and used a Metropolis-adjusted Langevin algorthim to predict the intensity outside of the observed frames [3]. Chakraborty et. al (2011) discuss nonhomogeneous Poisson process modeling as a richer alternative to ecological presence/absence models and describe their data as "degraded" in the sense that sampling bias prevented the entire region from being fully observed; they fit their model by aggregating the point pattern to counts in grid cells and using MCMC [4].

With INLA, the SPDE approach, and the "off grid" approximation facilitating the routine fitting of LGCP models, there is now renewed interest in accounting for variable sampling effort in spatial point process models [11, 12].

Sampling effort is accounted for using the theory of thinned point processes. Thinning refers to the events of one point process being kept or discarded probabilistically. Let $\lambda(\mathbf{s})$ be the intensity of the parent process, and let an event at $\mathbf{s}$ be observed with probability $p(\mathbf{s})$. The observed point process is a thinned point process with intensity

$$\lambda_p(\mathbf{s}) = p(\mathbf{s})\lambda(\mathbf{s}).$$

We seek to make posterior inferences about the parent intensity, $\lambda(\mathbf{s})$. If $p(\mathbf{s})$ is known, its value at each node is used to adjusts the SPDE integration weights. Most usefully, if it is known that $p(\mathbf{s}) = 0$ at certain nodes $\mathbf{s}$ because they were outside the sruveyed domain, the weights become zero so those nodes do not contribute to the integral. If $p(\mathbf{s})$ is unknown, it can be modeled. Taking the logarithm of the above and substituting in the basis function representation, we have the log-linear model

$$\log\left[\lambda_p(\mathbf{s})\right] = \log\left[p(\mathbf{s})\right] + \sum_j z_j \phi_j(\mathbf{s}).$$

Thus, any log-linear model for $p(\mathbf{s})$ that can be fit by INLA and the SPDE approach can be incorporated into the LGCP model. For example, Yuan et. al (2017) fit such a model to data from a line-transect survey using a spline model to acocunt for the unknown detection function [12]

3. Applications

3.1. Simulation Study

ideas:
simulate LGCPs
fit via Simpson method
vary mesh coarseness
explore bias/variability in parameter posteriors
explore spatial prediction error in latent GP

3.2. Data Application

Examples with data, maybe `bei` *dataset or Victorville. Should read like a tutorial because there aren't many good resources yet.*

3.2.1. Data and Model

3.2.2. Fitting in R-INLA

explain the mesh
explain the weight calculations
walk through the spatial predictions

3.2.3. Results

3.2.4. Model Checking

4. Conclusion and Discussion

References

[1] F.E. Bachl, F. Lindgren, D.L. Borchers, and J.B. Illian, *inlabru: an R package for bayesian spatial modelling from ecological survey data*, Methods in Ecology and Evolution 10 (2019), pp. 760–766.

[2] M. Blangiardo and M. Cameletti, *Spatial and Spatio-temporal Bayesian Models with R-INLA*, Wiley, 2015.

[3] A. Brix and J. Møller, *Space-time multi type log gaussian cox processes with a view to modelling weeds*, Scandinavian Journal of Statistics 28 (2001), pp. 471–488.

[4] A. Chakraborty, A.E. Gelfand, A.M. Wilson, A.M. Latimer, and J.A. Silander, *Point pattern modelling for degraded presence-only data over large regions*, Journal of the Royal Statistical Society: Series C (Applied Statistics) 60 (2011), pp. 757–776.

[5] J.B. Illian, S.H. Sørbye, and H. Rue, *A toolbox for fitting complex spatial point process models using integrated nested laplace approximation (inla)*, The Annals of Applied Statistics (2012), pp. 1499–1530.

[6] F. Lindgren and H. Rue, *Bayesian spatial modelling with R-INLA*, Journal of Statistical Software 63 (2015), pp. 1–25.

[7] F. Lindgren, H. Rue, and J. Lindström, *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 (2011), pp. 423–498.

[8] J. Møller and R. Waagepetersen, *Modern spatial point process modelling and inference*, Scandinavian Journal of Statistics 34 (2007), pp. 643–711.

[9] J. Møller, A.R. Syversveen, and R.P. Waagepetersen, *Log gaussian cox processes*, Scandinavian journal of statistics 25 (1998), pp. 451–482.

[10] H. Rue, S. Martino, and N. Chopin, *Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations*, Journal of the royal statistical society: Series b (statistical methodology) 71 (2009), pp. 319–392.

[11] D. Simpson, J.B. Illian, F. Lindgren, S.H. Sørbye, and H. Rue, *Going off grid: Computationally efficient inference for log-gaussian cox processes*, Biometrika 103 (2016), pp. 49–70.

[12] Y. Yuan, F.E. Bachl, F. Lindgren, D.L. Borchers, J.B. Illian, S.T. Buckland, H. Rue, T. Gerrodette, *et al.*, *Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales*, The Annals of Applied Statistics 11 (2017), pp. 2270–2297.

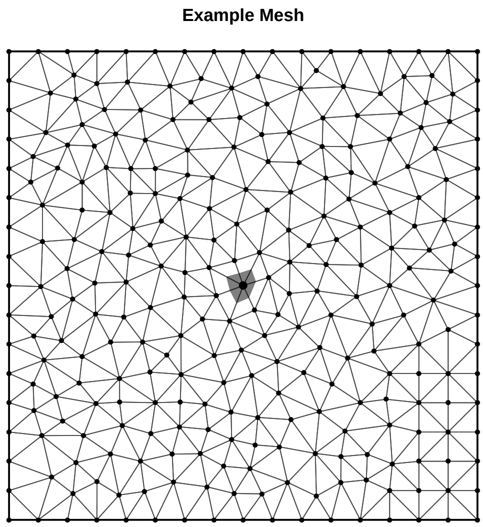

Figure 3. An illustration of the nodes (dots) and the weighting scheme for one node (large dot). This node represents the shaded region, so its weight is proportional to the shaded area. The shaded region is constructed by connecting the midpoints of the node's edges.