Progress this week
January 26, 2020

- Proofread.
- Incorporated many on Andy's comments.
- Consolidated raw data figures into a single multipanel figure.

Previous word count: 3257
Previous page count: 13.5

Current word count:
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 January 26, 2020

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

Statistical methods for spatial prediction result in a high-dimensional inference problem that can present computational challenges, particularly for large datasets. For instance, 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, spatial prediction may require hundreds of thousands of latent variables.

Spatial point process models further complicate the situation; in addition to computational challenges due to a large number of latent variables and parameters, point process models also require evaluating 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 and, unlike a probability distribution, the integral over the entire domain does not necessarily equal one.

Many methods have been introduced for modeling spatial point patterns, including quadrature-based approximations *(cite Baddeley)*, pseudodata approaches *(cite Baddeley/Berman/Turner etc)*, and Markov chain Monte Carlo [12]. The number of parameters and latent variables in these models still poses a major computational challenge. This computational challenge led to the development of INLA [14].

(Add a paragraph about computational challenges. This is really the point of INLA (fast albeit approximate solutions). —Andy)

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

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 [14]. 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 exhibit 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.

We focus on spatial mapping using a hierarchical construction of the LGCP, and do so without discussing many classic spatial point process concepts such as Gibbs processes, Markov processes, point process densities, or the Papangelou conditional intensity function. For readers interested in the general spatial point process context in which the LGCP originates, we recommend other references [7, 8, 12].

(paragraph about partially observed point processes)

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

The remainder of Section 1 provides an overview of spatial LGCP models and INLA. Section 2 describes the methodology developed to fit LGCP models in INLA, and Section 3 presents example analyses of real-world data. Finally, Section 4 presents concluding remarks.

1.1. Log-Gaussian Cox Process

Notes from [12] *(delete once fully integrated into text)*:

- applications mentioned: seismology, ecology, forestry, geography, spatial epidemiology and materials science
- modern computing (e.g. McMC) allows for model-based inference without a stationarity assumption
- Minke whale distance sampling data used as an example of incomplete observation (they used a shot noise Cox process model)
- distribution of point process is the joint distrubition of n and the n events, or equivalently the distribution of the number of events in any arbitrary region
- in their presentation of the two Poisson process postulates, the second one is given the number of events in a region they are iid proportional to the intensity function
- Cox process is driven by a non-negative random process such that conditional on the driving process the Cox process is a Poisson process
- LGCP has no edge effects because the point process in the window is fully specified by the GP restricted to the window
- quite a bit about Gibbs processes, densities wrt stationary Poisson processes, and Papangelou conditional intensities (comments: The density was traditionally used to facilitate McMC simulation and ML fitting when the likelihood was not explicitly available. The conditional intensity of a Poisson process is the intensity

function because of independence among the points. The hierarchical setup of the LGCP lets us work directly with a Poisson process so the density and conditional intensity are not needed. LGCP models with constructed covariate can model Gibbs processes.)
- more things about infinite Gibbs and Markov processes that go over my head but we don't need because we deal with bounded and finite processes

Notation and terminology:

- process defined on $\mathcal{D} \subset \mathbb{R}^2$, domain of the intensity function, define $d = \dim(\mathcal{D})$
- observation window $\mathcal{S} \subset \mathcal{D}$
- *(Note: When thinking about sampling, we need to define three regions: the domain $\mathcal{D}$ over which the process mathematically operates, the region $\mathcal{R}$ over which inferences are desired, and the observed/sampled observation window $\mathcal{S}$. The general relationship is $\mathcal{S} \subset \mathcal{R} \subset \mathcal{D} \subset \mathbb{R}^d$, where all of the subset symbols taken to mean "subset or equal". The "fully surveyed" situation is $\mathcal{S} = \mathcal{R}$.)*
- $\mathbf{X}$ point process on $\mathcal{R}$, $\mathbf{x} = \{x_1, \ldots, x_n\}$ realized point pattern
- point $x \in \mathbf{x}$ called an event
- intensity function $\lambda(u)$
- x event, s numerical integration node, u arbitrary location in $\mathcal{D}$
- $z(u)$ a column vector of covariates/predictors at u
- "point" refers to a u unless clearly stated otherwise
- bold for sets and spatial processes, normal italics for spatial vectors
- y and variations will be used for objects derived from the point pattern, e.g. pseudodata

(insert paragraph giving overview of Cox process)

The LGCP is a Poisson process driven by a latent Gaussian process [13]. A Poisson process is characterized entirely by its intensity function $\lambda(u)$, 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 any region $\mathcal{B}$ follows a Poisson distribution with mean $\int_{\mathcal{B}} \lambda(u)\mathrm{d}u$.
(2) The numbers of events in disjoint regions are independent.

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

$$\log \lambda(u) = \mathbf{z}'(u)\boldsymbol{\beta}.$$

The LGCP adds another stochastic layer,

$$\log \lambda(u) = \boldsymbol{\Psi}(u) = \mathbf{z}(u)'\boldsymbol{\beta} + \mathbf{e}(u),$$

where $\mathbf{e}$ is a Gaussian process with mean 0.

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.

Möller and Waagepetersen (2007) present model checking tools for LGCP models [12].

(insert paragraph reviewing applications and uses of LGCP)

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 [14]. 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 [9]. It is readily implemented using the standalone INLA software or in R via the R-INLA package [10].

2. Methodology

INLA provides an efficient computational framework for fitting Bayesian models with latent Gaussian variables. On top of this framework, several tools are built to further simplify the fitting of spatial LGCP models. The stochastic partial differential equation (SPDE) approach provides dimension reduction for the spatial GP [11]. 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 [15]. With these computational improvements, researchers are now able to efficiently fit LGCP models to large spatial point pattern datasets and even incompletely-observed point patterns [16].

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 [4]. 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) [11].

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

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

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 [11]. They then construct approximations for positive integer values of ν by ν-fold convolution of $\mathbf{e}$ with itself. Finally, they use a finite element method to generalize 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 s_i at which to model $\mathbf{\Psi}(s_i) = \mathbf{z}(s_i)'\boldsymbol{\beta} + \mathbf{e}(s_i)$, 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 $\mathbf{e}(s_i)$ are modeled as a GMRF where the distribution of each $\mathbf{e}(s_i)$ depends only on the $\mathbf{e}(s_j)$ where s_i and s_j are connected by an edge. The GMRF representation is assumed to be a piecewise linear approximation of the continuous Gaussian field; values of $\mathbf{e}(s)$ for s not in the set of nodes are predicted by linear interpolation using the barycentric coordinates of 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 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 [5]. Simpson et. al. (2016) explain that this is also computationally wasteful [15].

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. For notational clarity, assume $\log[\lambda(u)] = \mathbf{\Psi}(u)$. The exact log-likelihood is

$$\ell(\lambda) = C - \int \lambda(u)\mathrm{d}u + \sum_{i=1}^{n} \log\left[\lambda(x_i)\right]$$

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 $\log[\lambda(u)]$, namely

$$\log\left[\lambda(u)\right] \approx \sum_{j=1}^{m} \psi_j \phi_j(u)$$

with $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_m)'$ a multivariate normal random vector and $\{\phi_1, \ldots, \phi_m\}$ a set

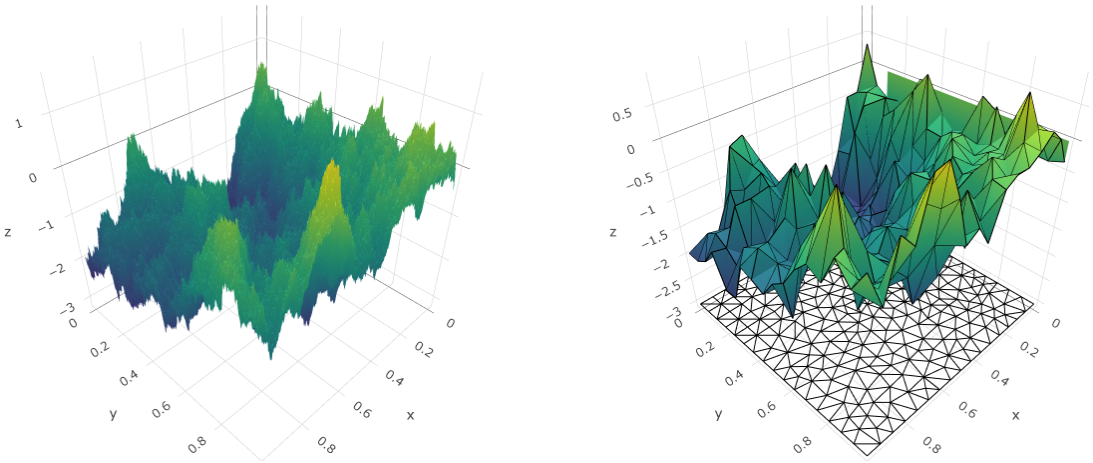


Figure 1. A realization of a spatial Gaussian process (left) and an approximation of that realization over a triangular mesh (right).

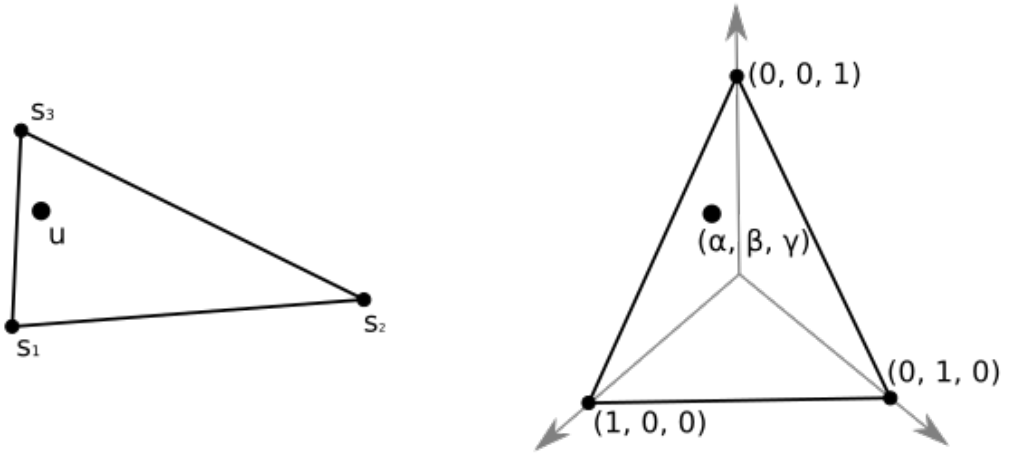


Figure 2. An illustration of the linear transformation from a mesh triangle to the simplex. (α, β, γ) are the barycentric coordinates of u.

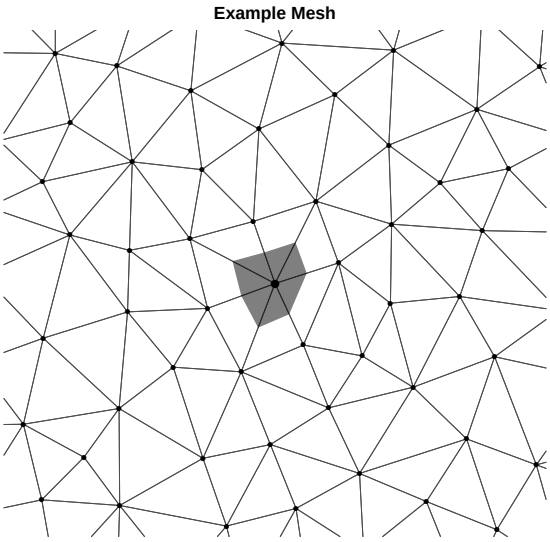


Figure 3. An illustration of the nodes (dots) and the weighting scheme for one node (large dot). This node represents the shaded region, so its weight is proportional to the shaded area. The shaded region is constructed by connecting the midpoints of the node's edges.

of linearly independant basis functions. This defines a numerical integration scheme at nodes s_i with weights α_i and yields the approximation,

$$\begin{aligned}\ell(\lambda) &\approx C - \sum_{i=1}^{m} \alpha_i \exp\left[\sum_{j=1}^{m} \psi_j \phi_j(s_i)\right] + \sum_{i=1}^{n}\sum_{j=1}^{m} \psi_j \phi_j(x_i) \\ &= C - \boldsymbol{\alpha}' \exp\left[\mathbf{A}\boldsymbol{\psi}\right] + \mathbf{y}'\mathbf{A}_2\boldsymbol{\psi}.\end{aligned}$$

This is equivalent to the likelihood of independent Poisson random variables with means $\alpha_i\eta_i$ *(translate the confusing algebra from the appendix)* observations $y_i = 0$ at the nodes s_i and $y_i = 1$ at the events x_i. 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 and construct a pseudodata vector $\mathbf{y}$.

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 [5]. 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 [6].

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 [15, 16].

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 $\lambda(u)$ be the intensity of the parent process, and let an event at u (if it exists) be observed with probability $p(u)$. The observed point process is a thinned point process with intensity

$$\lambda_p(u) = p(u)\lambda(u),$$

and if the parent process is a Poisson process then the observed process is also a Poisson process [12].

We seek to make posterior inferences about the parent intensity, $\lambda(u)$. If $p(u)$ is known, its value at each node is used to adjust the SPDE integration weights. Most usefully, if it is known that $p(s) = 0$ at certain nodes s because they were outside the sruveyed domain, the weights become zero so those nodes do not contribute to the integral. If $p(u)$ 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

$$\log\left[\lambda_p(u)\right] = \log\left[p(u)\right] + \sum_{j=1}^{m} \psi_j \phi_j(u).$$

Thus, any log-linear model for $p(u)$ 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 account for the unknown detection function [16].

3. Applications

3.1. Beilschmiedia Pendula Lauraceae Dataset

For an example analysis of real-world data, we consider the locations of $n = 3605$ *Beilschmiedia pendula Lauraceae* trees in 1000 m by 500 m plot in a tropical rainforest [12]. The data are available as the `bei` dataset in the `spatstat` R package [2]. The point pattern is accompanied by elevation and gradient covariates measured on a square lattice with 5 m spacing. The point pattern exhibits inhomogeneity which appears to be associated with elevation and the the elevation (Figure 4).

3.1.1. Model Specification

We use the log-Gaussian Cox process model with intensity

$$\log\left[\lambda(u)\right] = \beta_0 + \beta_1 z_1(u) + \beta_2 z_2(u) + \mathbf{e}(u)$$

where $z_1(u)$ is the elevation at u, $z_2(u)$ is the magnitude of the elevation gradient at u, and $\mathbf{e}(u)$ is a zero-mean Gaussian process with Matérn covariance and $\alpha = 2$.

State priors — using INLA's defaults which (I think) are uniform but I need to confirm than or adjust them to something reasonable.

3.1.2. Fitting in R-INLA

We fit the model in R using the `INLA` package and its implementation of the SPDE approach. Some manual preprocessing was needed to employ the Poisson factorization. Our R code can be found in the supplementary materials. This section provides a conceptual explanation of the model-fitting process.

The first task is to construct a mesh for the piecewise linear approximations. A 25 meter resolution appears adequate to accurately represent the covariate surfaces. We used tools in the `INLA` package to create a Delaunay triangulation with a maximum edge length of 25 m. The resulting mesh has $m = 2145$ nodes $s_1, \ldots, s_{2145}$ and 4096 triangles (Figure 5). We used linear interpolation to approximate the covariates $z_1(s_i)$ and $z_2(s_i)$ at each node. After creating the mesh, `INLA` can initialize an object to represent the covariance structure on the mesh.

The next step is to define the pseudodata and calculate the numerical integration weights. The pseudodata vector is

$$\mathbf{y} = (0, \ldots, 0, 1, \ldots, 1)'$$

consisting of 2145 zeros corresponding to the mesh nodes and 3604 ones representing the events. The weight vector is

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{2145}, 0, \ldots, 0)'.$$

The alphas equal one-third of the area of the triangles of which each node is a vertex, calculated by the `INLA` package. The zeros correspond to the events, which do not contribute to the numerical integration.

Talk about the $\mathbf{A}$ *matrix.*

The final bit of setup is to combine the pseudodata, covariates at the nodes, and a node index variable into a data list for the `inla()` function. This function is then called with arguments specifying a formula for the linear predictor, a Poisson family model (which defaults to a log link), the data list, the $\mathbf{A}$ matrix, and the $\boldsymbol{\alpha}$ vector as the exposure parameter.

3.1.3. Results

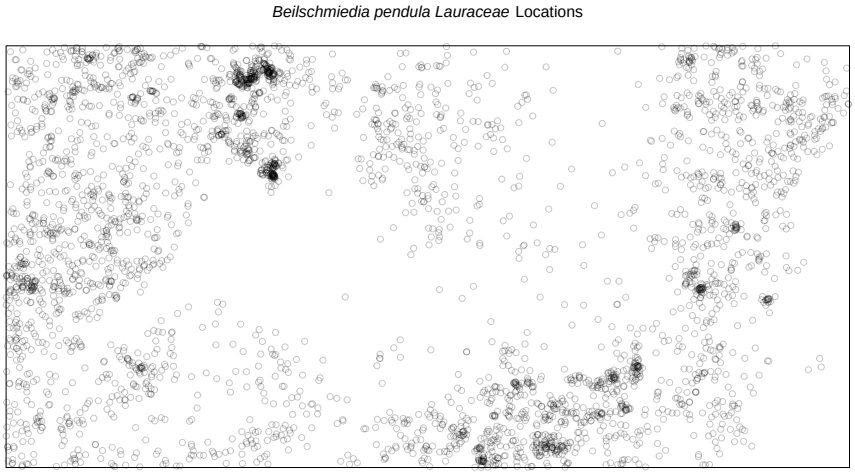
insert intervals for betas and covariance parameters

decide which of figures 8, 9, 10, 11 are needed, then combine into a single multi-panel figure.

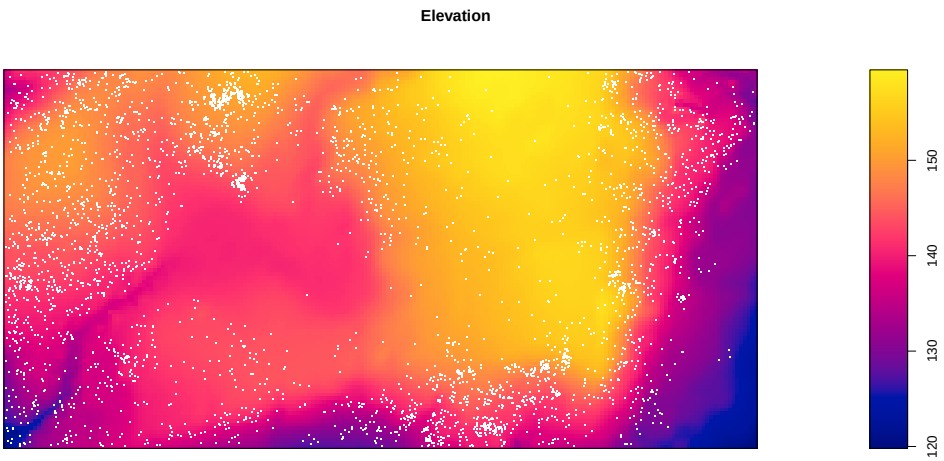
The posterior mean of the fixed effect component of the linear predictor,

$$E(\beta_0|\mathbf{x}) + E(\beta_1|\mathbf{x})z_1(u) + E(\beta_2|\mathbf{x})z_2(u),$$

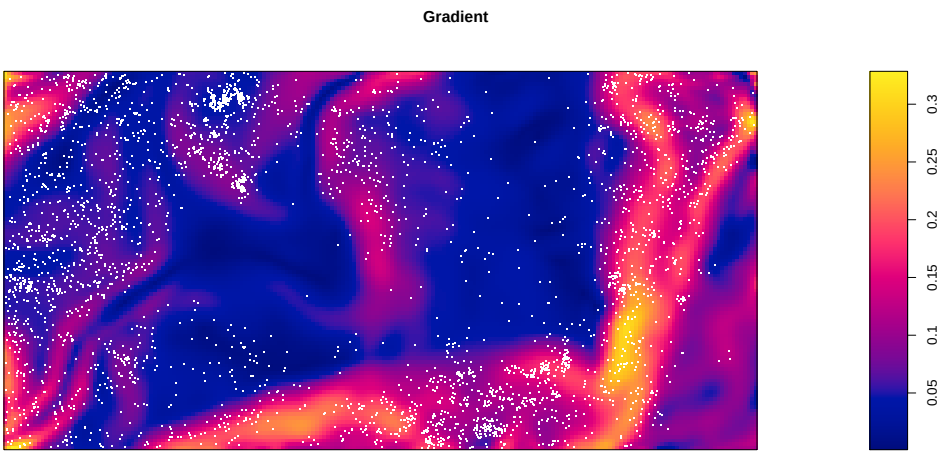
appears as an average of the elevation and gradient surfaces (Figure 8). The fixed effects do not explain all of the heterogeneity of the point pattern; some spatial structure remains to be described by the posterior predicted GP (Figure 9).



(a) Locations of trees (events).



(b) Elevation in meters, with events overlaid.



(c) Magnitude of the elevation gradient, with events overlaid.

Figure 4. Plots of the event and covariate data.

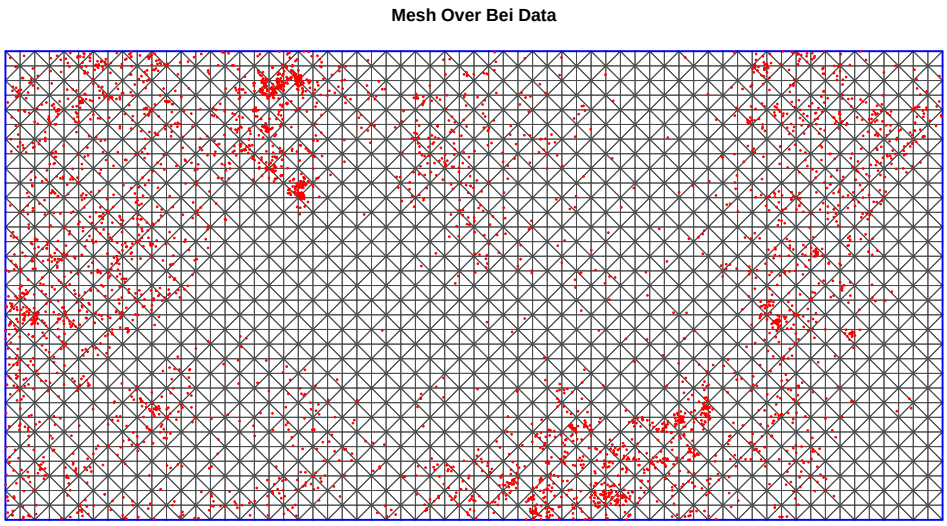


Figure 5. Triangular mesh with events overlaid.

3.1.4. Model Checking

For a more general discussion of residuals, see [3]. The classic Pearson residual,

$$\frac{\text{Observed count} - \text{Expected count}}{\sqrt{\text{Expected count}}},$$

is a special case.

Should make the whole four-panel plot if citing Baddeley et. al.

Pearson residuals (Figure 12): small magnitude in the region with no observed events and low posterior intensity (not possible to see much variation), the model overestimates the intensity in the canyon (too small to be accurately represented on the mesh), random with no obvious spatial patterns elsewhere, a few large-magnitude positive residuals (maybe there is some clustering?), tried a few different grids (not shown) and saw the same patterns

4. Conclusion and Discussion

not really possible to get a posterior distribution of the intensity function because it is a nonlinear transformation of several parameters and we do not have their joint posterior

the mesh for is finer than needed to describe the GP — think about coarsening and refining based on the detail needed to represent the covariates (is there literature on this?)

shows that mapping based on Bayesian LGCP models is now practical with useful accuracy in short amount of time

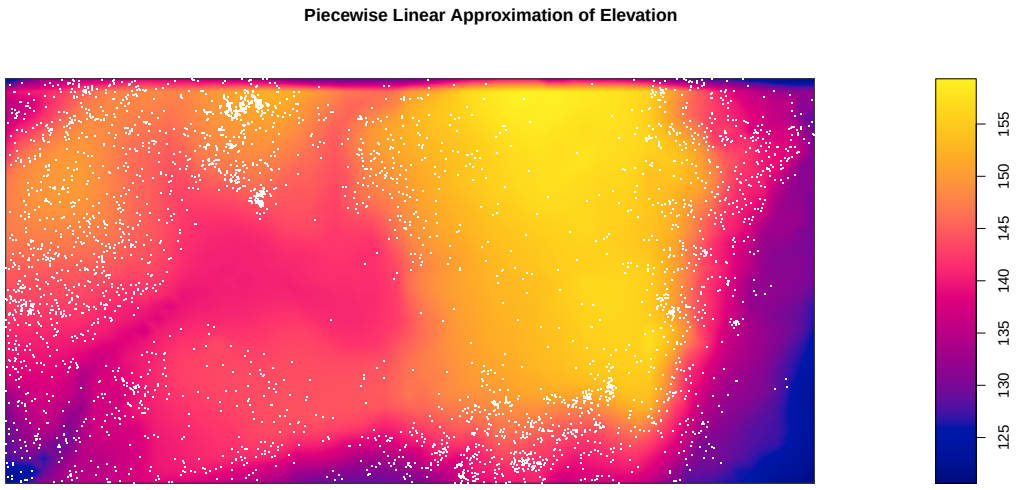


Figure 6. Piecewise linear approximation of the elevation surface.

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] A. Baddeley and R. Turner, *Spatstat: An R package for analyzing spatial point patterns*, Journal of Statistical Software 12 (2005), pp. 1–42.

[3] A. Baddeley, R. Turner, J. Møller, and M. Hazelton, *Residual analysis for spatial point processes*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (2005), pp. 617–666.

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

[5] 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.

[6] 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.

[7] N. Cressie, *Statistics for Spatial Data*, Wiley, 1993.

[8] P. Diggle, *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed., CRC Press, 2013.

[9] 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.

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

[11] 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.

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

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

[14] H. Rue, S. Martino, and N. Chopin, *Approximate bayesian inference for latent gaussian*

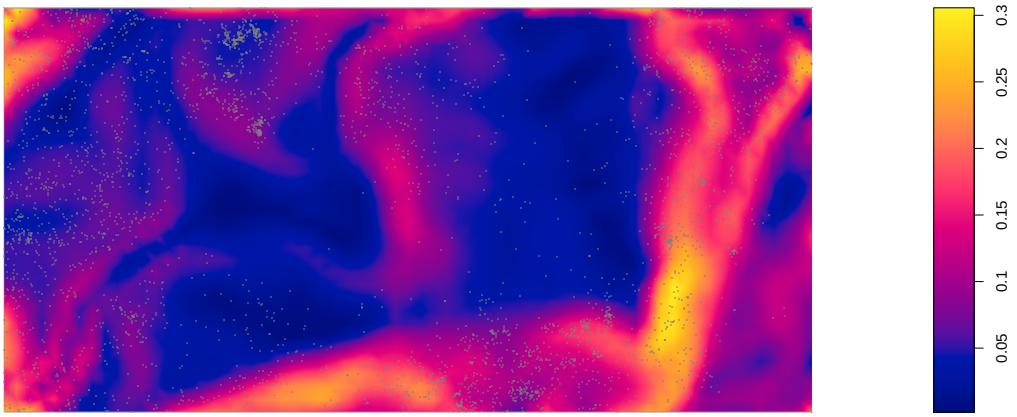


Figure 7. Piecewise linear approximation of the gradient surface.

models by using integrated nested laplace approximations, Journal of the royal statistical society: Series b (statistical methodology) 71 (2009), pp. 319–392.

[15] 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.

[16] 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.

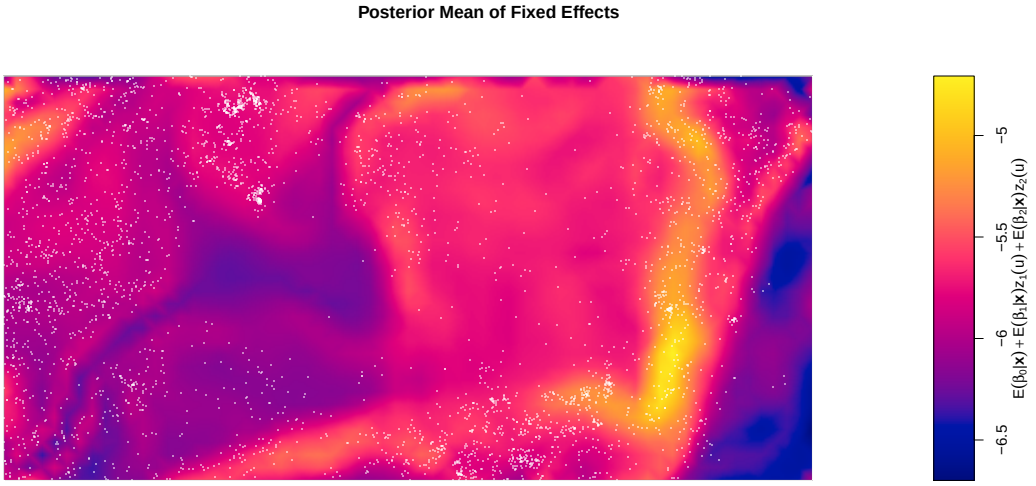


Figure 8. Posterior mean surface for the fixed effects.

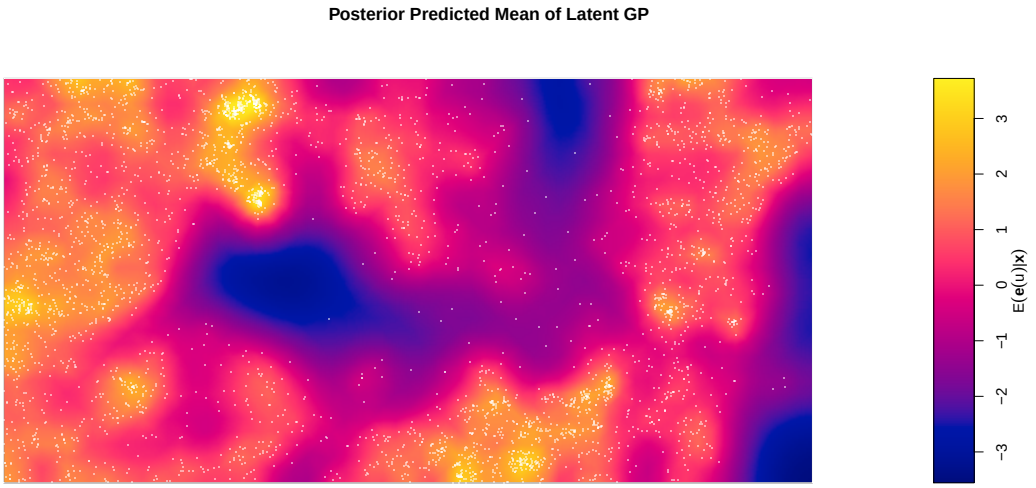


Figure 9. Posterior mean of prediction surface.

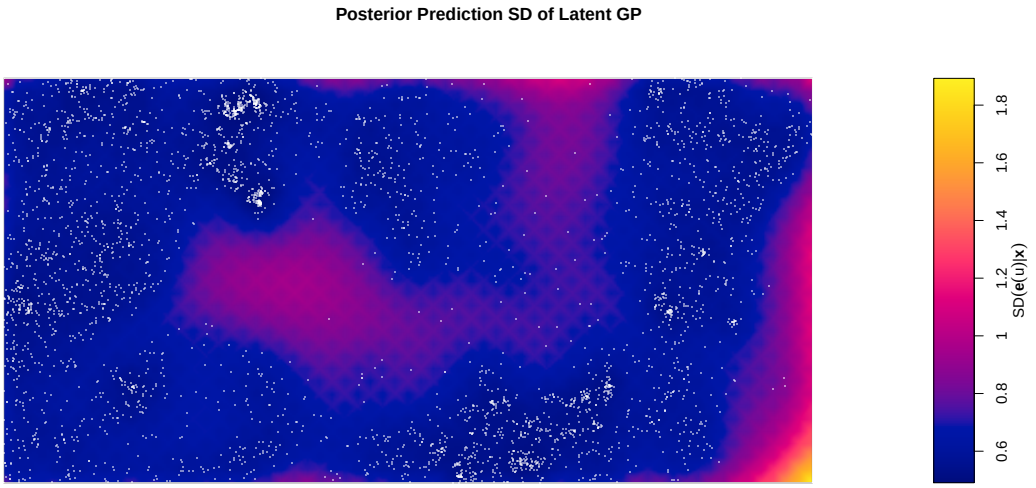


Figure 10. Pointwise posterior standard deviation of predictions.

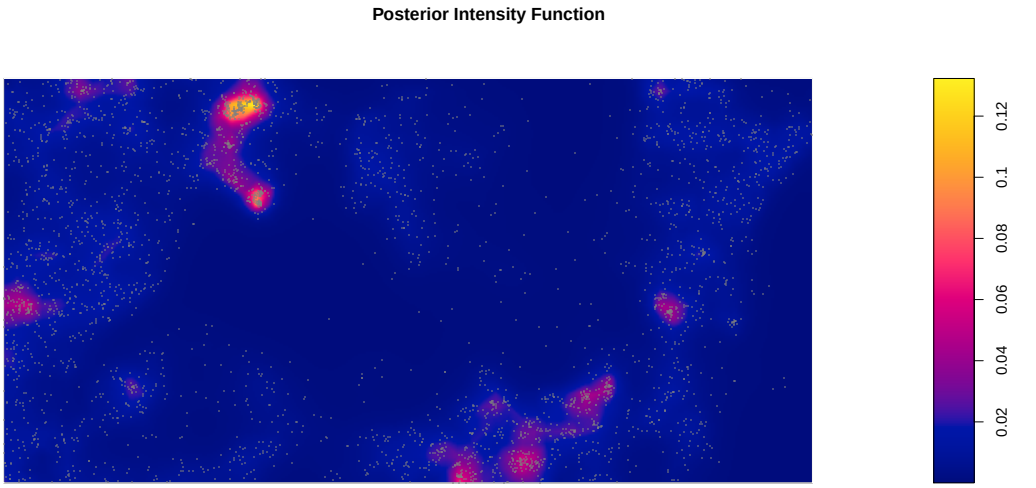


Figure 11. Posterior intensity surface in events per square meter, calculated using the piecewise linear approximate covariate surfaces and the posterior means of the intercept, coefficients, and GP.

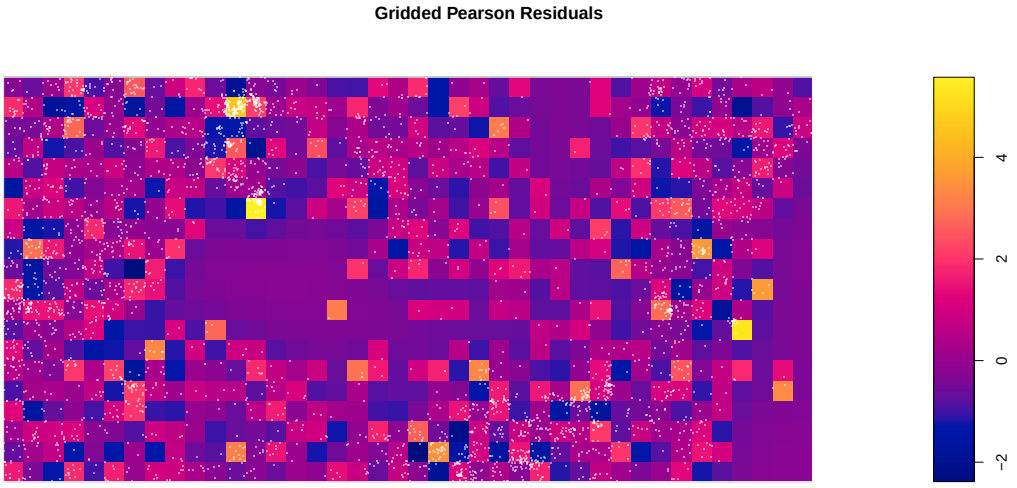


Figure 12. Pearson residuals calculated on a grid.