Biometrika (2016), **103**, 1, *pp.* 49–70
Printed in Great Britain

doi: 10.1093/biomet/asv064
Advance Access publication 5 February 2016

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

BY D. SIMPSON

Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, U.K.
d.simpson@bath.ac.uk

UW Space-Time Reading Group
Jim Faulkner
May 16, 2018

Log-Gaussian Cox process

- A simple point process on a bounded region $\Omega \subset \mathbb{R}^2$ is an inhomogeneous Poisson process.
- Let the intensity surface across the region be the function $\lambda(s)$.
- The number of points in region $D \subset \Omega$ follows a Poisson distribution with mean $\Lambda(D) = \int_D \lambda(s)ds$.
- The point pattern Y depends on the intensity surface.
- If $Z(s) = \log \lambda(s)$ is a Gaussian process, then the point process is known as a doubly-stochastic Poisson process or as a log-Gaussian Cox process.
- The likelihood is:

$$\pi(y \mid \lambda) = \exp\left\{|\Omega| - \int_\Omega \lambda(s)\mathrm{d}s\right\} \prod_{s_i \in y} \lambda(s_i)$$

Computation on a grid

- The most common inference method is to set up a regular lattice over the bounded region of interest, Ω.
- Let the number of points in cell s_{ij} be N_{ij}.
- Conditional on $Z(s)$, the N_{ij} can be considered independent Poisson random variables.
- Assume a constant value of $Z(s_{ij}) = z_{ij}$ within each grid cell, with $\Lambda_{ij} \approx |s_{ij}| \exp(z_{ij})$.
- Can assume z is multivariate normal with covariance function $C(i, j)$, but is computationally costly.
- Can use GMRF approximation and INLA instead.

Issues with grid approach

- Grid approximation converges to true solution as the size of the cells decreases to zero
- Increasing number of grid cells increases computational cost
- Lattice used to approximate the latent Gaussian field and used to approximate locations of points
- Binning the points is the main source of error
- Need more grid cells to accurately approximate the likelihood than needed to estimate the field
- Solution: construct a continuous approximation to the field in way that is still computationally efficient

Mátern SPDE

A Gaussian process $Z(s)$ with Mátern covariance function can be respresented as a stochastic differential equation

$$\tau \left(\kappa^2 - \Delta\right)^{\alpha/2} Z(s) = W(s)$$

where

- τ is a scaling parameter
- κ is a range parameter
- $\Delta = \sum_{i=1}^{d} \partial^2 / \partial s_i^2$ is the Laplacian operator
- $\alpha = \nu + d/2$, where ν is a smoothing parameter and d is the dimension (2 here)
- $W(s)$ is spatial white noise

Basis function representations

- A GP $Z(s)$ can represented over continuous time using a finite basis expansion:

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

  where $\boldsymbol{z} = \{z_1, \ldots z_n\}$ is a multivariate Gaussian vector and $\{\phi_i(s)\}_{i=1}^{n}$ is a set of linearly independent deterministic basis functions.
- This approach has been used in various ways, including the Karhunen-Loeve decomposition, process convolutions, fixed-rank kriging, and SPDE approximations.

GMRF solutions to SPDE

Using piecewise-linear basis functions as test functions, the set of weak solutions to the SPDE for $\alpha = 2$ and $j = 1, \ldots n$ is

$$\int_\Omega \phi_j(s)\tau \left(\kappa^2 - \Delta\right) Z(s)ds \stackrel{\mathrm{d}}{=} \int_\Omega \phi_j(s)W(s)ds$$

After substituting the basis expansion for $Z(s)$ we get

$$\tau \sum_{j=1}^{n} z_j \int_\Omega \left(\kappa^2 - \Delta\right) \phi_j(s)\phi_j(s)ds \stackrel{\mathrm{d}}{=} \int_\Omega \phi_j(s)W(s)ds$$

GMRF solutions to SPDE

These solutions can be represented in matrix notation as (assuming Neumann boundary conditions):

$$\tau(\kappa^2 \boldsymbol{C} + \boldsymbol{G})\boldsymbol{z} \stackrel{\mathrm{d}}{=} \boldsymbol{w}$$

where $\boldsymbol{w} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{C}})$ and $\tilde{C}_{i,j} = \int_\Omega \phi_i(s)\phi_j(s)ds$. It follows that

$$\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{Q}^{-1})$$

where $\boldsymbol{Q} = \tau^2(\kappa^2 \boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1}(\kappa^2 \boldsymbol{C} + \boldsymbol{G})$ and the elements of $\boldsymbol{G}$ and the sparse approximation of $\tilde{\boldsymbol{C}}$ are

$$C_{i,i} = \int_\Omega \phi_i(s)ds \text{ and } G_{i,j} = \int_\Omega \nabla\phi_i(s)\nabla\phi_j(s)ds$$

GMRF solutions to SPDE

The benefits of this formulation are:

- $\boldsymbol{Q}$ is sparse so can use sparse matrix operations for GMRF
- $Z(s)$ is specified over continuous space
- Can use INLA for inference

Likelihood

- Likelihood for inhomogeneous point process

$$\pi(Y \mid \lambda) = \exp\left\{|\Omega| - \int_{\Omega} \lambda(s)\mathrm{d}s\right\} \prod_{s_i \in Y} \lambda(s_i)$$

- Associated log-likelihood

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

Likelihood approximation

Using a numerical integration rule $\int_\Omega f(s)ds \approx \sum_{i=1}^{p} \tilde{\alpha}_i f(\tilde{s}_i)$ for fixed nodes $\{\tilde{s}_i\}_{i=1}^{p}$ and weights $\{\tilde{\alpha}_i\}_{i=1}^{p}$, we can construct an approximate log-likelihood:

$$\log \pi(y|z) \approx C - \sum_{i=1}^{p} \tilde{\alpha}_i \exp\left\{\sum_{j=1}^{n} z_j \phi_j(\tilde{s}_i)\right\} + \sum_{i=1}^{N}\sum_{j=1}^{n} z_j \phi_j(s_i)$$

where C is a constant and $Z(s)$ is replaced by the basis expansion.

Likelihood approximation

We can write the log-likelihood approximation in matrix notation as

$$\log \pi(y|z) \approx C - \tilde{\boldsymbol{\alpha}}^T \exp\{\boldsymbol{A}_1 \boldsymbol{z}\} + \mathbf{1}^T \boldsymbol{A}_2 \boldsymbol{z}$$

where

- $[A_1]_{ij} = \phi_j(\tilde{s}_i)$ is matrix of basis functions evaluated at integration nodes
- $[A_2]_{ij} = \phi_j(s_i)$ is matrix of basis functions evaluated at observation locations

Likelihood approximation

Let

- $\log \boldsymbol{\eta} = (\boldsymbol{z}^T \boldsymbol{A}_1^T, \boldsymbol{z}^T \boldsymbol{A}_2^T)^T$
- $\boldsymbol{\alpha} = (\tilde{\boldsymbol{\alpha}}, \mathbf{0}_{N\times 1}^T)^T$
- $\boldsymbol{y} = (\mathbf{0}_{p\times 1}^T, \mathbf{1}_{N\times 1}^T)^T$

The likelihood can then be re-written in Poisson form:

$$\pi(y \mid z) \approx C \prod_{i=1}^{N+p} \eta_i^{y_i} \exp(-\alpha_i \eta_i)$$

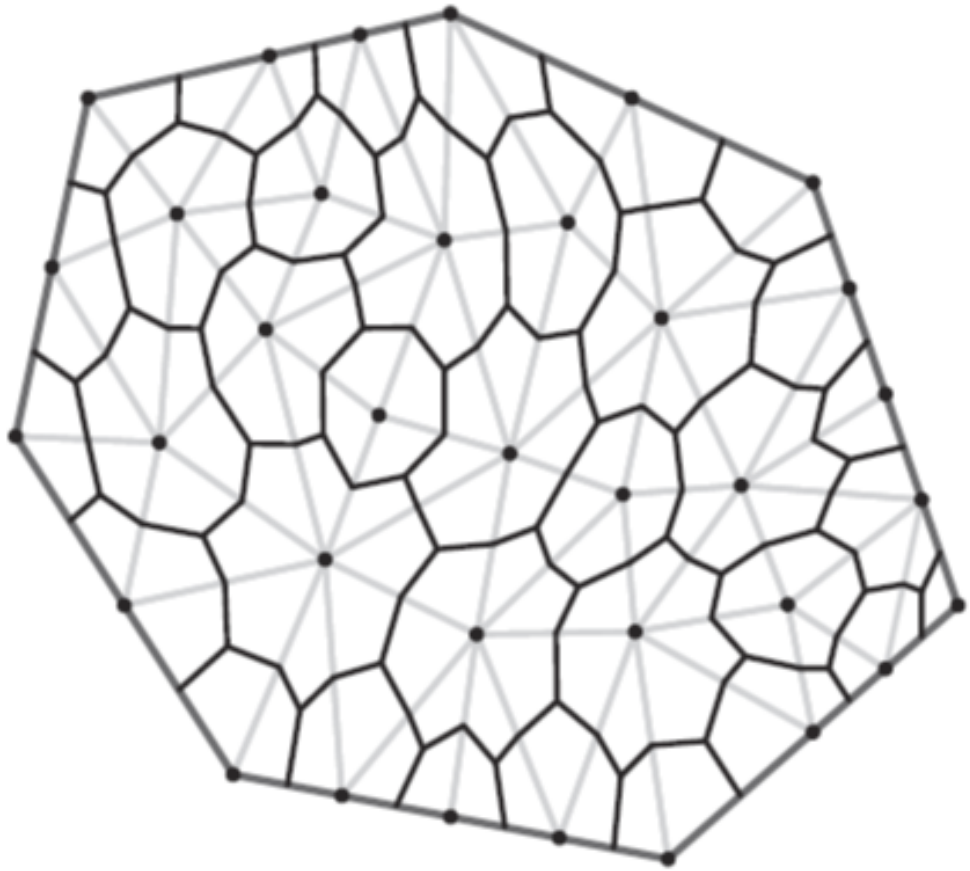

Fig. 1. The dual mesh (black) is constructed by joining the centroids of the primal triangular mesh (grey). The volumes of these dual cells define the weights of an integration scheme based at the nodes of the primal mesh.

Data example 7.2

- Locations of 4,294 trees on 50 ha plot in rainforest.
- Intensity related to phosphorus concentration
- Model 1: SPDE LGCP
  - $Z(s) = \mu + \beta P(s) + x(s)$
  - $\kappa = 0.0014$ and $\log(\tau) \sim N(0, 1000)$
  - How many basis functions?
- Model 2: Lattice LGCP
  - $z = \mu \mathbf{1} + \beta P + x$
  - $x \sim N(0, \tau^{-1} \boldsymbol{Q}^{-1})$ as intrinsic RW2, and $\tau \sim Ga(1, 10^{-5})$
  - How many grid cells?
- Both models fit with INLA

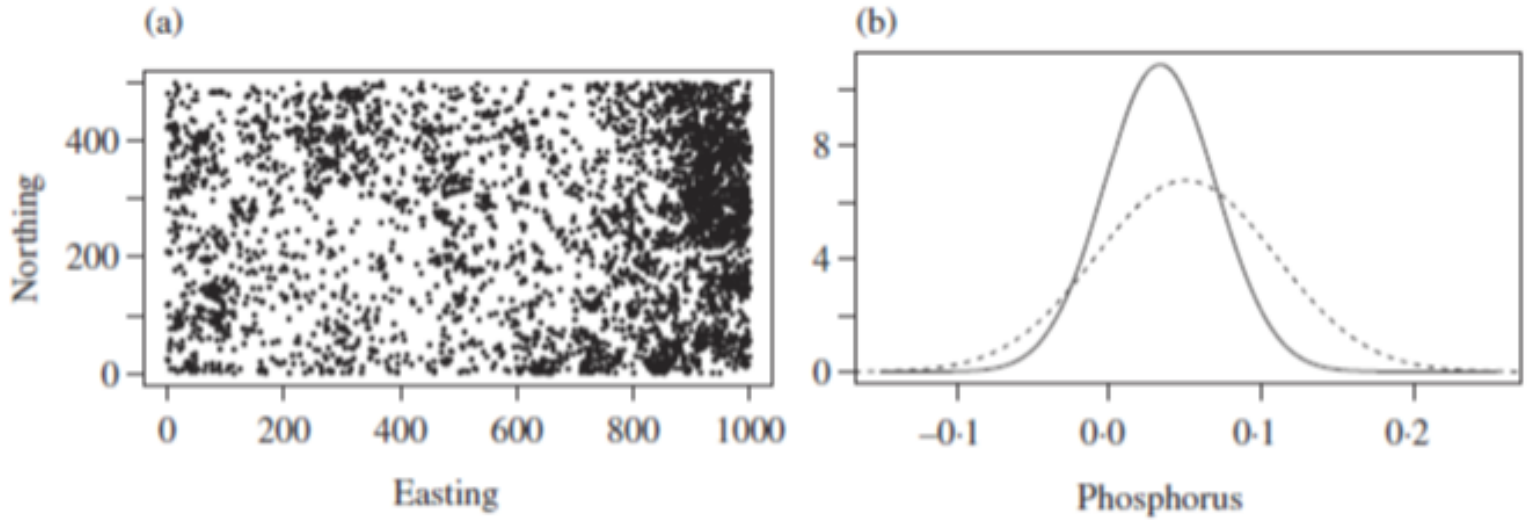

Fig. 2. The effect of soil potassium levels on the location of *Protium tenuifolium*: (a) locations of *Protium tenuifolium*; (b) the posterior covariate effect of phosphorus obtained using the standard lattice method (dashed) and the stochastic partial differential equation approach (solid).

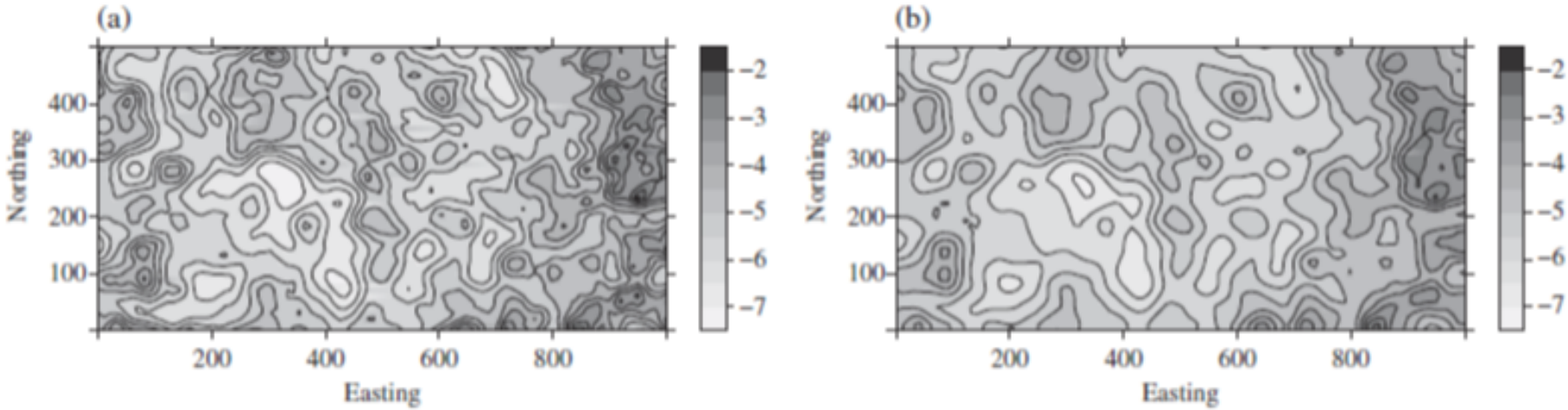

Fig. 3. Estimated spatial effects for *Protium tenuifolium*: (a) using a standard lattice point process model; (b) using the stochastic partial differential equation approach.

Data example 7.3

- Subarea with reduced sampling effort
- $\lambda(s) = S(s) \exp\{Z(s)\}$, where $S(s)$ is function for sampling effort
- Make $S(s) = 0$ in rectangle, $S(s) = 1$ elsewhere
- This results in no contribution to likelihood - make mesh coarse there
- Compared to complete simulated data
- Also compared two meshes – uniform and with coarse in rectangle.
  - little difference in resulting posterior marginals
  - coarse mesh resulted in 35% reduction in computation time

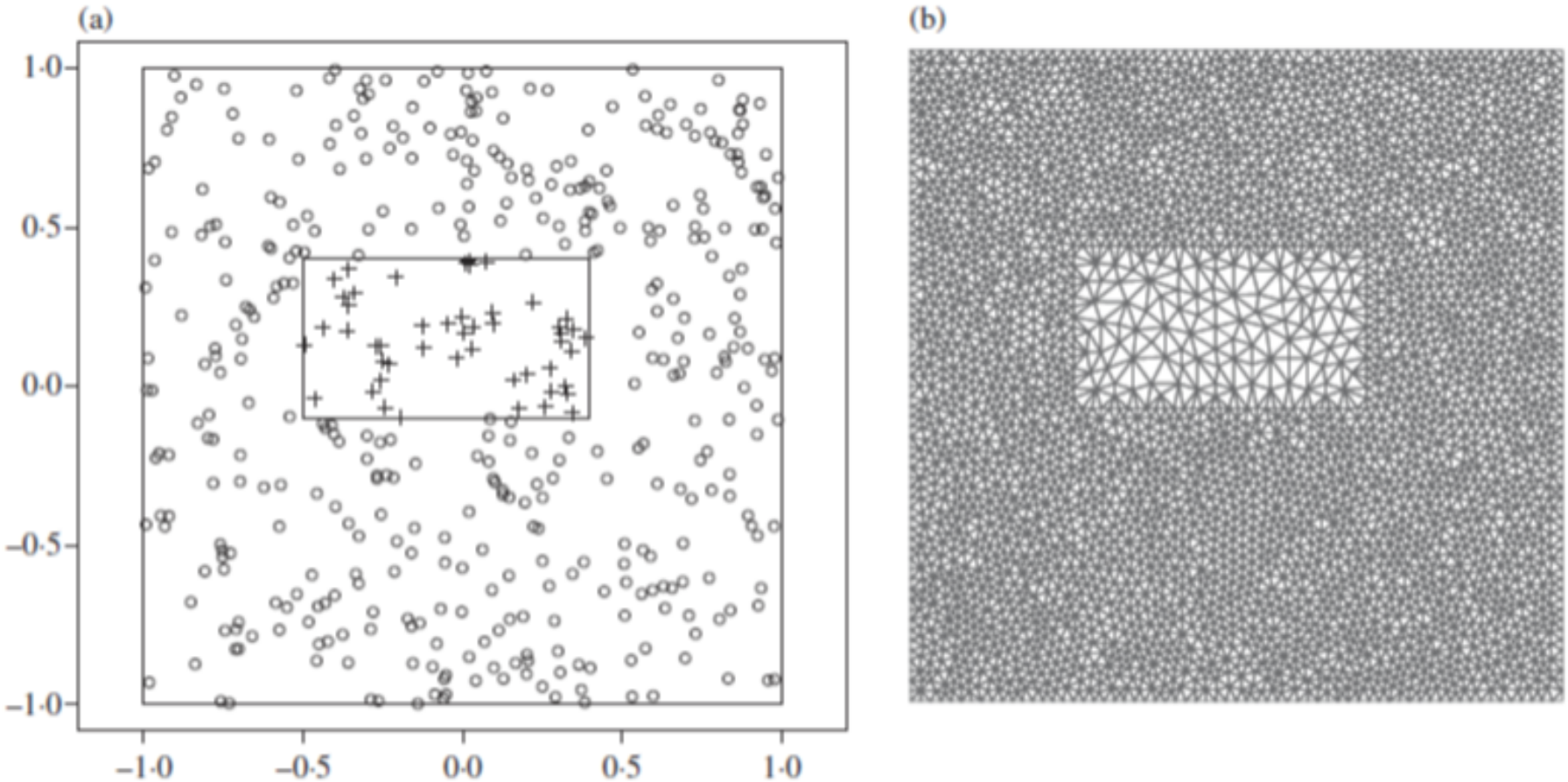

Fig. 4. Simulated data with a hole in the sampling effort: (a) an inner rectangle delineating the area in which there was no sampling, with plus signs representing the points that were missed due to incomplete sampling; (b) a mesh that takes into account the lack of sampling effort in the rectangular region.

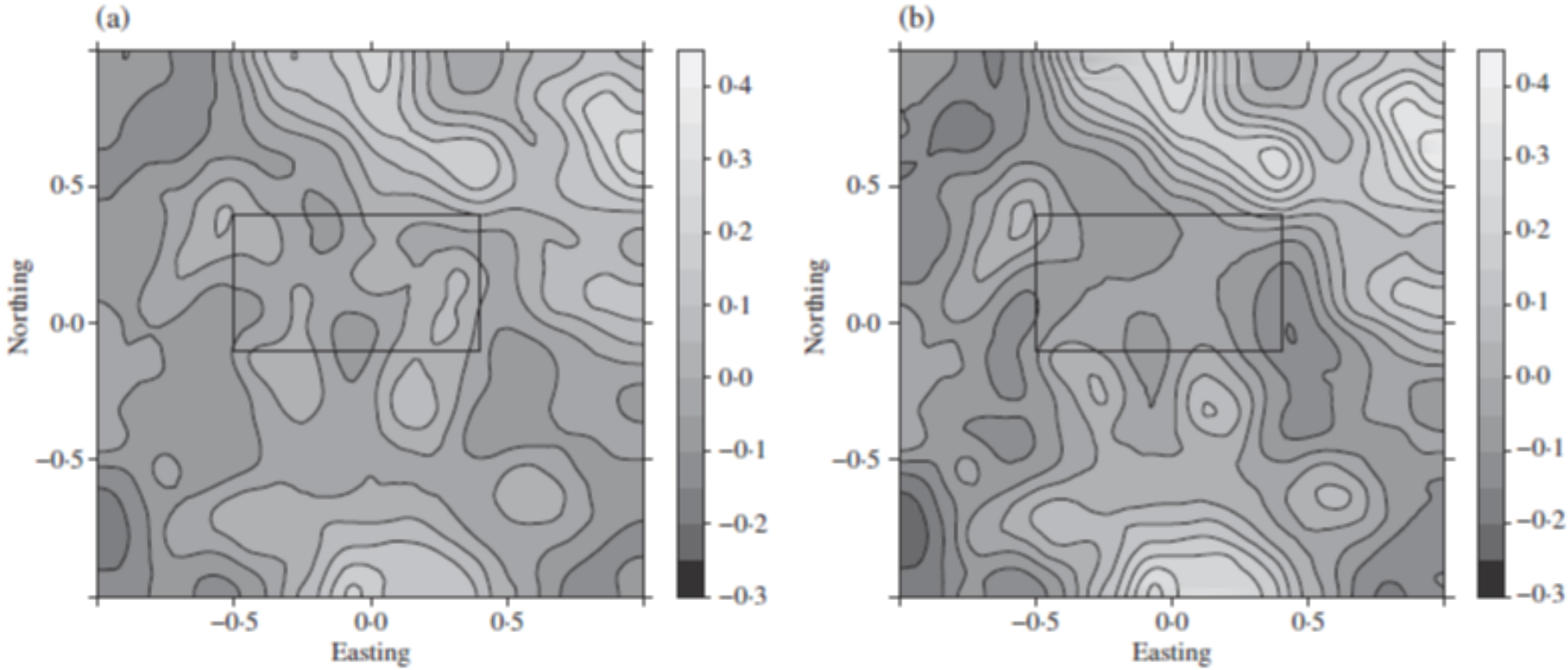

Fig. 5. The posterior mean of the spatial effect for variable sampling effort (see § 7·3): (a) using the complete simulated point pattern; (b) using the incomplete, partially observed point pattern. The large-scale features of the two fields are similar in areas in which the point pattern was sampled.

Data example 7.4

- Point process over entire ocean
- Motivated by model of risk of freak waves
- Expect freak wave height more variable near coast
- Used Neumann boundary conditions – doubles variance near boundaries
- Simulated field with 913 points

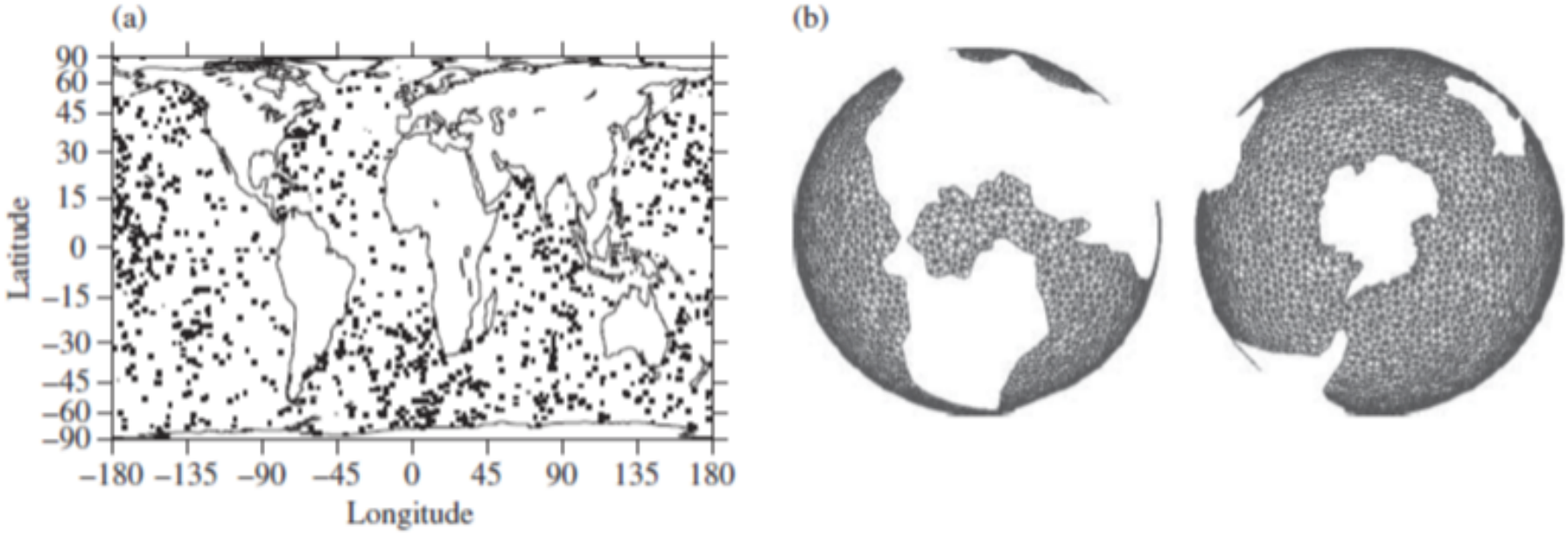

Fig. 6. (a) A simulated log-Gaussian Cox process over the oceans. (b) A mesh that covers the oceans.

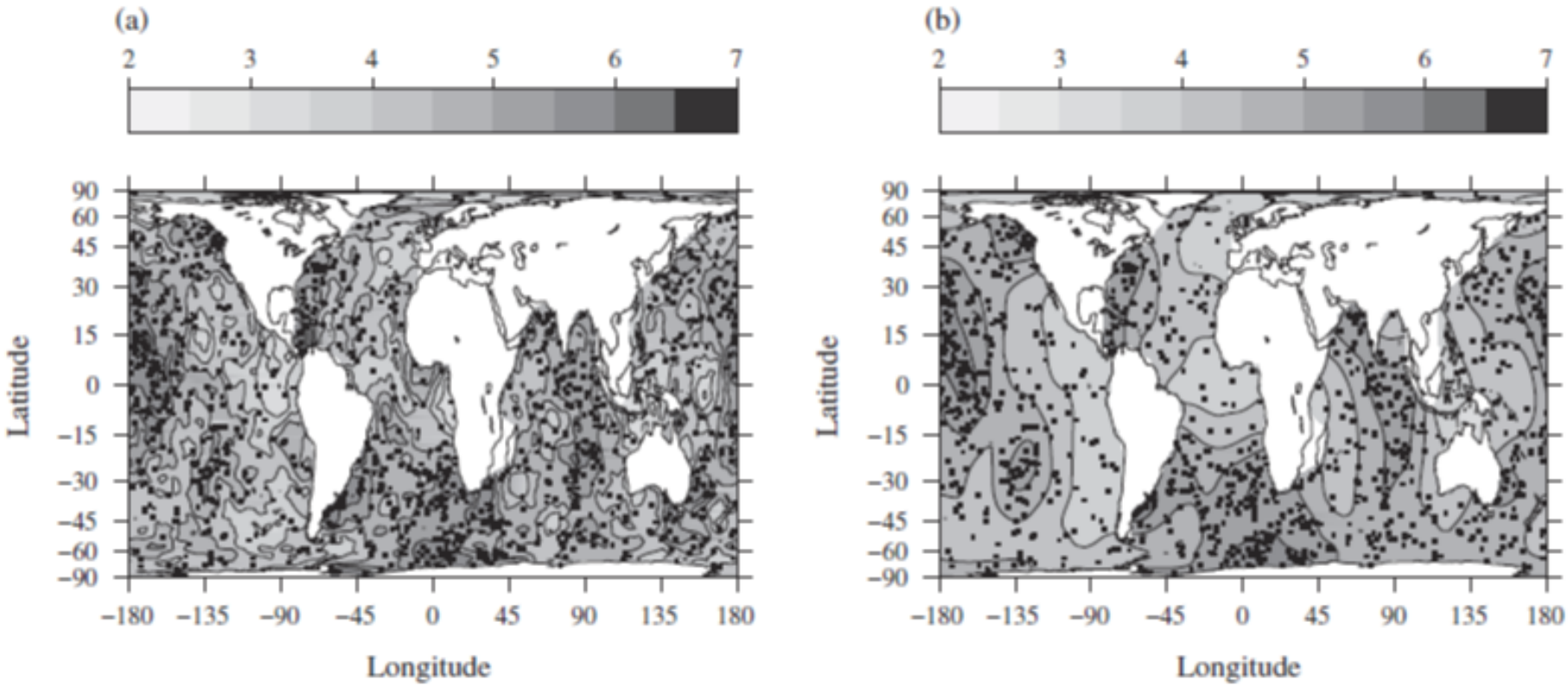

Fig. 7. Inference for a point process over the oceans: (a) true surface from the latent Gaussian random field used to generate the sample in Fig. 6; (b) posterior mean of the latent spatial effect. Note that the large-scale behaviour is the same in both panels.

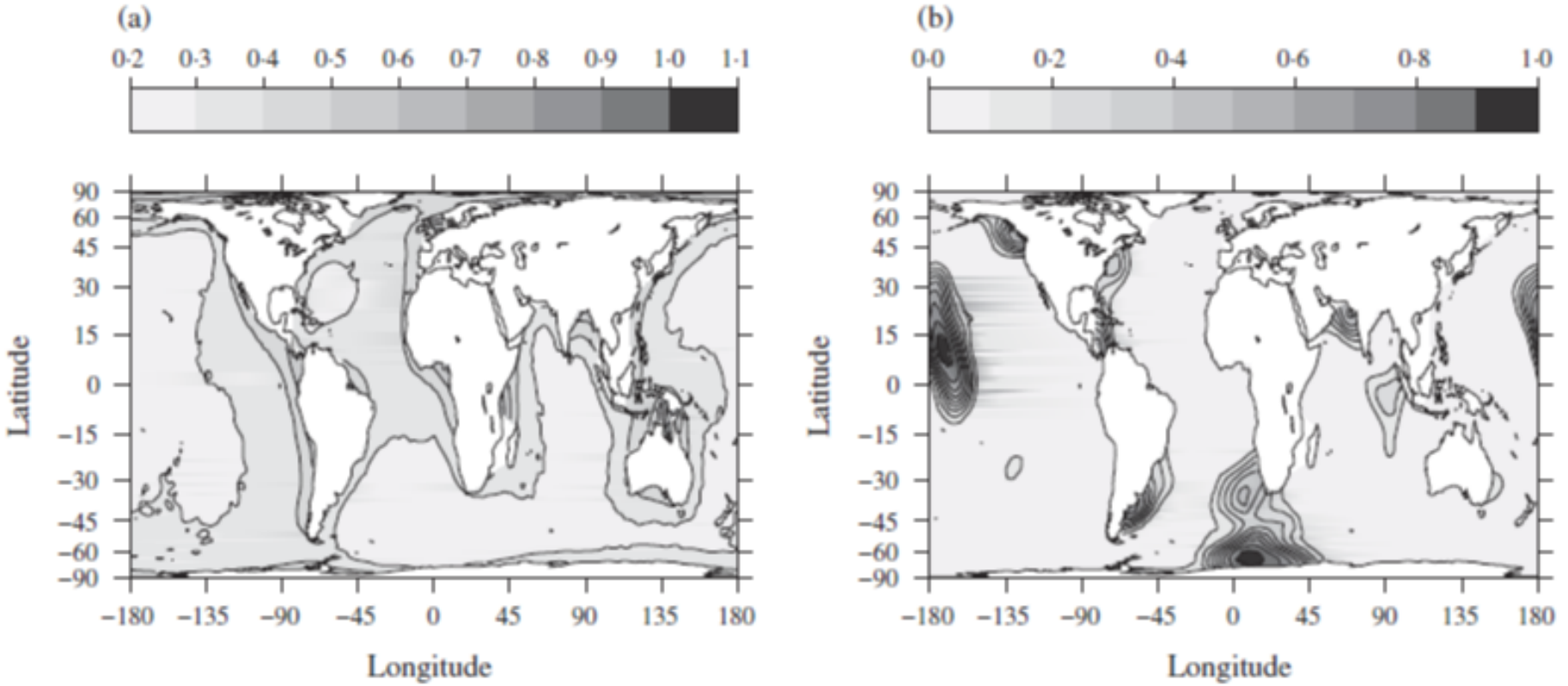

Fig. 8. Inference for a point process over the oceans: (a) the pointwise posterior standard deviation for the log risk surface; (b) the posterior risk map $\mathrm{pr}\{\log \lambda(s) > 5{\cdot}5 \mid y\}$.

Some tutorial links

- http://www.r-inla.org/examples/tutorials
- https://haakonbakka.bitbucket.io/organisedtopics.html
- http://www.flutterbys.com.au/stats/tut/tut12.13.html