Combining all the pieces together to create an efficient full Bayesian geostatistical model: The SPDE method in Stan

Joaquin Cavieres

Instituto de Estadística, Universidad de Valparaíso

Overview

Motivation

- A Gaussian random field (GRF) is the main component of spatial modelling. It has been used in different areas of the research and lately has gained notoriety due to a special structure in a hierarchical framework, the latent Gaussian models.

- The problem arises when we want to evaluate non Gaussian likelihood with a dense covariance matrix ($\boldsymbol{\Sigma}$)

- A Gaussian random field (GRF) is the main component of spatial modelling. It has been used in different areas of the research and lately has gained notoriety due to a special structure in a hierarchical framework, the latent Gaussian models.

- The problem arises when we want to evaluate non Gaussian likelihood with a dense covariance matrix ($\boldsymbol{\Sigma}$)

If we use the classical definition of a GRF, then we need to consider:

$$\begin{pmatrix} u(\boldsymbol{s}_1) \\ u(\boldsymbol{s}_2) \\ \vdots \\ u(\boldsymbol{s}_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\boldsymbol{s}_1) \\ \mu(\boldsymbol{s}_2) \\ \vdots \\ \mu(\boldsymbol{s}_n) \end{pmatrix}, \begin{pmatrix} c(\boldsymbol{s}_1,\boldsymbol{s}_1) & c(\boldsymbol{s}_1,\boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_1,\boldsymbol{s}_n) \\ c(\boldsymbol{s}_2,\boldsymbol{s}_1) & c(\boldsymbol{s}_2,\boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_2,\boldsymbol{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\boldsymbol{s}_n,\boldsymbol{s}_1) & c(\boldsymbol{s}_n,\boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_n,\boldsymbol{s}_n) \end{pmatrix} \right), \tag{1}$$

where $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ are all of the distinct values of $\boldsymbol{s}_i$ in our spatial data.

- The storage scales quadratically in "n"
- The computation scales cubically in "n"

Rue and Held (2005) approximates this problem assuming that the GRF is a Gaussian Markov random field (GMRF).

If we use the classical definition of a GRF, then we need to consider:

$$\begin{pmatrix} u(\boldsymbol{s}_1) \\ u(\boldsymbol{s}_2) \\ \vdots \\ u(\boldsymbol{s}_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\boldsymbol{s}_1) \\ \mu(\boldsymbol{s}_2) \\ \vdots \\ \mu(\boldsymbol{s}_n) \end{pmatrix}, \begin{pmatrix} c(\boldsymbol{s}_1, \boldsymbol{s}_1) & c(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_1, \boldsymbol{s}_n) \\ c(\boldsymbol{s}_2, \boldsymbol{s}_1) & c(\boldsymbol{s}_2, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_2, \boldsymbol{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\boldsymbol{s}_n, \boldsymbol{s}_1) & c(\boldsymbol{s}_n, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_n, \boldsymbol{s}_n) \end{pmatrix} \right), \tag{1}$$

where $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ are all of the distinct values of $\boldsymbol{s}_i$ in our spatial data.

- The storage scales quadratically in "n"
- The computation scales cubically in "n"

Rue and Held (2005) approximates this problem assuming that the GRF is a Gaussian Markov random field (GMRF).

If we use the classical definition of a GRF, then we need to consider:

$$\begin{pmatrix} u(\boldsymbol{s}_1) \\ u(\boldsymbol{s}_2) \\ \vdots \\ u(\boldsymbol{s}_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\boldsymbol{s}_1) \\ \mu(\boldsymbol{s}_2) \\ \vdots \\ \mu(\boldsymbol{s}_n) \end{pmatrix}, \begin{pmatrix} c(\boldsymbol{s}_1, \boldsymbol{s}_1) & c(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_1, \boldsymbol{s}_n) \\ c(\boldsymbol{s}_2, \boldsymbol{s}_1) & c(\boldsymbol{s}_2, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_2, \boldsymbol{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\boldsymbol{s}_n, \boldsymbol{s}_1) & c(\boldsymbol{s}_n, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_n, \boldsymbol{s}_n) \end{pmatrix} \right), \tag{1}$$

where $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ are all of the distinct values of $\boldsymbol{s}_i$ in our spatial data.

- The storage scales quadratically in "n"
- The computation scales cubically in "n"

Rue and Held (2005) approximates this problem assuming that the GRF is a Gaussian Markov random field (GMRF).

Considering the above, and taking advantage of the computational efficiency of using a GMRF, Lindgren et al. (2011) created an explicit link to approximate the GRF by a GMRF. This is:

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{s}_2 - \boldsymbol{s}_1||)^{\nu} K_{\nu}(\kappa||\boldsymbol{s}_2 - \boldsymbol{s}_1||), \tag{2}$$

where $||\boldsymbol{s}_2 - \boldsymbol{s}_1||$ is the Euclidean distance between two geographical points $\boldsymbol{s}_1$ and $\boldsymbol{s}_2 \in \mathcal{R}^D$, K_ν is the modified Bessel function with $\nu > 0$, $\kappa > 0$ what controls the correlation through $\rho = \sqrt{8\nu}/\kappa$, and σ^2 is the marginal variance.

The authors noted that the GRF ($u(\boldsymbol{s})$) and Matérn function (2) has solution to the linear fractional SPDE

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau u(\boldsymbol{s})) = W(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{R}^D, \quad \text{with} \quad \alpha = \nu + d/2, \quad \kappa > 0, \nu > 0, \tag{3}$$

where W is a spatial Gaussian white noise (Whittle (1954), Whittle (1963)), Δ is the Laplacian operator and τ controls the marginal variance as:

$$\tau^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}\sigma^2} \tag{4}$$

So, to find $u(\boldsymbol{s})$ with Matérn function (2) then is necessary to solve (3).

With additional other mathematical calculus, Lindgren et al. (2011) used the elements finite method to represent $u(\boldsymbol{s})$ in a non structured triangulation as:

$$u_h(\boldsymbol{s}) = \sum_{k=1}^{n} w_k \psi_k(\boldsymbol{s}), \tag{5}$$

where $\{\psi_k\}_{k=1}^{n}$ are piecewise linear basis functions. Finally, they showed that the Gaussian coefficients $\{w_k\}_{k=1}^{n}$ are GMRF when $\alpha = 1$ and can be approximated with a GMRF when $\alpha = 2$ (Liu et al. (2016)).

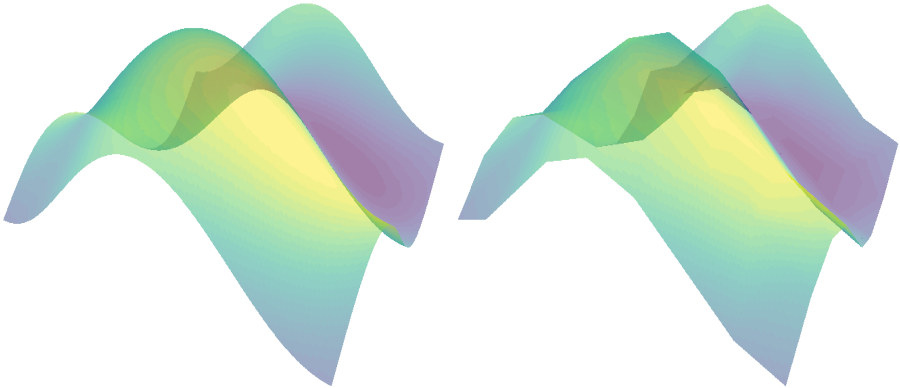

Figure 1: A GRF (left) approximated by a GMRF (right), (Krainski et al. (2018))

Lindgren et al. (2021) did a list with recent applications of the SPDE method in different areas of the research, for example:

- Astronomy (Levis et al. (2021))
- Health (Moraga et al. (2021), INLA et al. (2021))
- Engineering (Zhang et al. (2021))
- Theory (Ghattas and Willcox (2021))
- Environmetrics (Hough et al. (2021))
- Imaging (Aquino et al. (2021))
- Fisheries (Cavieres et al. (2021))
- .....

More of this references in Lindgren et al. (2021)

How we can use the approximate GRF ($\sim$ GMRF) in a Bayesian spatial (spatio-temporal) model?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in R.

Stan

Stan (Gelman et al. (2015), Carpenter et al. (2017)) is a probabilistic programming language that provides full Bayesian inference for continuous-variable models through Markov chain Monte Carlo method with an adaptive form of Hamiltonian Monte Carlo algorithm.

How we can use the approximate GRF ($\sim$ GMRF) in a Bayesian spatial (spatio-temporal) model?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in `R`.

Stan

`Stan` (Gelman et al. (2015), Carpenter et al. (2017)) is a probabilistic programming language that provides full Bayesian inference for continuous-variable models through Markov chain Monte Carlo method with an adaptive form of Hamiltonian Monte Carlo algorithm.

How we can use the approximate GRF ($\sim$ GMRF) in a Bayesian spatial (spatio-temporal) model?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in R.

Stan

Stan (Gelman et al. (2015), Carpenter et al. (2017)) is a probabilistic programming language that provides full Bayesian inference for continuous-variable models through Markov chain Monte Carlo method with an adaptive form of Hamiltonian Monte Carlo algorithm.

A very quick explanation of how works INLA in the spatial context:

- The GRF is parameterized by the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$.
- We don't built a discrete model for the GRF on a grid, we construct an approximation of the GRF in a spatial continuous space defined on the entire study area.
- INLA done the inference for univariate posterior densities for the parameters of $u(\boldsymbol{s})$, and the joint posterior of the hyperparameters of the model.

some additional INLA references Rue et al. (2009), Lindgren et al. (2015), Blangiardo and Cameletti (2015), Bakka et al. (2018), Moraga (2019).

A very quick explanation of how works `Stan` in the spatial context:

- The GRF is parameterized by the covariance matrix $\boldsymbol{\Sigma}$.
- It use the Hamiltonian Monte Carlo (Neal (2011)), which itself is a generalization of the familiar Metropolis algorithm, to conduct sampling more efficiently through the posterior distribution by performing multiple steps per iteration.
- The sampling is done from a target density proportional to the product of marginal likelihood and the pior density specified by the user.

INLA uses the Laplace ratio approximation (LRA) described in Tierney and Kadane (1986) in an "nested" way, however, it remains an approximation method. So, how can we do an efficient (SPDE/GMRF) full Bayesian inference (MCMC method) for a spatial model in `Stan` ?

Template Model Builder (`TMB`, Kristensen et al. (2015) and `tmbstan`, Monnahan and Kristensen (2018)).

INLA uses the Laplace ratio approximation (LRA) described in Tierney and Kadane (1986) in an "nested" way, however, it remains an approximation method. So, how can we do an efficient (SPDE/GMRF) full Bayesian inference (MCMC method) for a spatial model in `Stan` ?

Template Model Builder (`TMB`, Kristensen et al. (2015) and `tmbstan`, Monnahan and Kristensen (2018)).

Methodology

What is TMB?

- Template Model Builder (TMB) is an open source R package that enables quick implementation of complex nonlinear random effects (latent variables) (Kristensen et al. (2015)).
- TMB is a frequentist statistical platform that woks with Automatic differentiation (AD) to obtain the first and second derivatives of the a function (e.g. loglikelihood, Skaug and Fournier (2006))
- Offers an easy access to parallel computations and the user can define the joint likelihood for the data and the random effects in a C++ template function, while all the other operations are done in R .

Why use TMB?

- The objective function (and its derivatives) can be called from `R`, hence, the parameter optimization can be done via e.g. `nlminb()`.
- The user can specify use the Laplace approximation to obtain the marginal likelihood of the latent variables (random effects).
- Compute the standard deviations of any parameter, or derived parameters by the Delta method.

Why use TMB?

- The objective function (and its derivatives) can be called from R, hence, the parameter optimization can be done via e.g. `nlminb()`.
- The user can specify use the Laplace approximation to obtain the marginal likelihood of the latent variables (random effects).
- Compute the standard deviations of any parameter, or derived parameters by the Delta method.

So, how can we express a spatial model in TMB? → Hierarchical model!

$$\boldsymbol{\theta} \sim \boldsymbol{\theta} \quad \text{Hyperparameters} \tag{6}$$

$$\boldsymbol{u} \mid \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}) \quad \text{Latent Gaussian field} \tag{7}$$

$$y_i \mid \boldsymbol{u}, \boldsymbol{\theta} \sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) \quad \text{Observations} \tag{8}$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix, $\boldsymbol{u}$ is the latent Gaussian field and $\eta_i = log(\mu_i) = \text{intercept} + f(\boldsymbol{X}_i) + \boldsymbol{u}_i$, where the matrix $\boldsymbol{X}$ is a set of covariates and $\boldsymbol{u} \sim \text{GMRF}(0, \boldsymbol{Q}^{-1})$

So, how can we express a spatial model in `TMB`? $\rightarrow$ Hierarchical model!

$$
\begin{aligned}
\boldsymbol{\theta} &\sim \boldsymbol{\theta} && \text{Hyperparameters} && (6)\\
\boldsymbol{u} \mid \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}) && \text{Latent Gaussian field} && (7)\\
y_i \mid \boldsymbol{u}, \boldsymbol{\theta} &\sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) && \text{Observations} && (8)
\end{aligned}
$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix, $\boldsymbol{u}$ is the latent Gaussian field and $\eta_i = log(\mu_i) = \text{intercept} + f(\boldsymbol{X}_i) + \boldsymbol{u}_i$, where the matrix $\boldsymbol{X}$ is a set of covariates and $\boldsymbol{u} \sim \text{GMRF}(0, \boldsymbol{Q}^{-1})$

So, how can we express a spatial model in `TMB`? $\rightarrow$ Hierarchical model!

$$
\begin{aligned}
\boldsymbol{\theta} &\sim \boldsymbol{\theta} && \text{Hyperparameters} && (6)\\
\boldsymbol{u} \mid \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}) && \text{Latent Gaussian field} && (7)\\
y_i \mid \boldsymbol{u}, \boldsymbol{\theta} &\sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) && \text{Observations} && (8)
\end{aligned}
$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix, $\boldsymbol{u}$ is the latent Gaussian field and $\eta_i = log(\mu_i) = \text{intercept} + f(\boldsymbol{X}_i) + \boldsymbol{u}_i$, where the matrix $\boldsymbol{X}$ is a set of covariates and $\boldsymbol{u} \sim \text{GMRF}(0, \boldsymbol{Q}^{-1})$

Consider to $\boldsymbol{y} = (y_1, ..., y_n)^T$ a vector of observations and $\boldsymbol{u} = (u_1, ..., u_q)$ a vector of latent random variables (random effects) influencing the value of $\boldsymbol{y}$. The conditional density of $\boldsymbol{y}$ given $\boldsymbol{u}$ is denoted by $(\boldsymbol{y} \mid \boldsymbol{u})$, and the marginal density of $\boldsymbol{u}$ is denoted by $h(\boldsymbol{y})$. Generally f and h depend on unknown parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)$ thus, we denoted this dependency by f_θ and h_θ respectively.

For estimation purposes the likelihood function for $\boldsymbol{\theta}$ must be expressed on function of the marginal distribution of $\boldsymbol{y}$, which is obtained by integrating out the random effects $\boldsymbol{u}$ from the joint density $f_\theta(\boldsymbol{y} \mid \boldsymbol{u})h_\theta(\boldsymbol{u})$, then the marginal likelihood can be written as:

$$\mathcal{L}(\boldsymbol{\theta}) = \int f_\theta(\boldsymbol{y} \mid \boldsymbol{u})h_\theta(\boldsymbol{u})d\boldsymbol{u} = \int \exp\{g(\boldsymbol{u}, \boldsymbol{\theta})\}d\boldsymbol{u} \tag{9}$$

where

$$g(\boldsymbol{u}, \boldsymbol{\theta}) = \log\{f_\theta(\boldsymbol{y} \mid \boldsymbol{u}) + \log\{h_\theta(\boldsymbol{u})\} \tag{10}$$

Solving this integral means a challenge and intensive computational work with the purpose of maximizing $\mathcal{L}(\boldsymbol{\theta})$. For the above, numerical methods can be performed to approximate the solution and one of them is the Laplace approximation. How $(\boldsymbol{\theta})$ is the vector of parameters, then we assume the function $g(\boldsymbol{u}, \boldsymbol{\theta})$ is such that:

$$\hat{\boldsymbol{u}}(\boldsymbol{\theta}) = \underset{u}{\operatorname{argmax}}\, g(\boldsymbol{u}, \boldsymbol{\theta}) \tag{11}$$

and

$$H(\theta) = \frac{\partial^2}{\partial \boldsymbol{u}^2} = g(\boldsymbol{u}, \boldsymbol{\theta})\Big|_{\boldsymbol{u}=\hat{\boldsymbol{u}}(\boldsymbol{\theta})} \tag{12}$$

are well defined in the range of $\boldsymbol{\theta}$. Then, we can derive an efficient algorithm for maximizing the Laplace approximation as:

$$\mathcal{L}^\star(\boldsymbol{\theta}) = |\mathrm{det}\{H(\theta)\}|^{-1/2}\mathrm{exp}\{g\{\hat{\boldsymbol{u}}, \theta\}\} \tag{13}$$

respect to $\boldsymbol{\theta}$. The critical point here is the numerical evaluation of the Hessian ($\boldsymbol{H(\theta)}$) but this can be solved by Automatic Differentiation.

How to solve in an efficient way a (spatial) Hierarchical Bayesian model?

- `R-INLA` $\Longrightarrow$ Rue et al. (2009), Blangiardo et al. (2013), Lindgren et al. (2015), Bakka et al. (2018)
- MCMC ... `Stan`? How? $\Longrightarrow$ `tmbstan` Monnahan and Kristensen (2018)

`tmbstan`

The models built in `TMB` pass their log-density and gradients calculations to the Bayesian samplers in `Stan` through a R interface. You only need to add the prior distribution on the parameters in your `TMB` code!

How to solve in an efficient way a (spatial) Hierarchical Bayesian model?

- `R-INLA` $\Longrightarrow$ Rue et al. (2009), Blangiardo et al. (2013), Lindgren et al. (2015), Bakka et al. (2018)
- MCMC ... `Stan`? How? $\Longrightarrow$ `tmbstan` Monnahan and Kristensen (2018)

`tmbstan`

The models built in `TMB` pass their log-density and gradients calculations to the Bayesian samplers in `Stan` through a R interface. You only need to add the prior distribution on the parameters in your `TMB` code!

How to solve in an efficient way a (spatial) Hierarchical Bayesian model?

- R-INLA $\Longrightarrow$ Rue et al. (2009), Blangiardo et al. (2013), Lindgren et al. (2015), Bakka et al. (2018)
- MCMC ... Stan? How? $\Longrightarrow$ tmbstan Monnahan and Kristensen (2018)

tmbstan

The models built in TMB pass their log-density and gradients calculations to the Bayesian samplers in Stan through a R interface. You only need to add the prior distribution on the parameters in your TMB code!

How to solve in an efficient way a (spatial) Hierarchical Bayesian model?

- `R-INLA` $\Longrightarrow$ Rue et al. (2009), Blangiardo et al. (2013), Lindgren et al. (2015), Bakka et al. (2018)
- MCMC ... `Stan`? How? $\Longrightarrow$ `tmbstan` Monnahan and Kristensen (2018)

`tmbstan`

The models built in TMB pass their log-density and gradients calculations to the Bayesian samplers in `Stan` through a R interface. You only need to add the prior distribution on the parameters in your TMB code!

How to solve in an efficient way a (spatial) Hierarchical Bayesian model?

- `R-INLA` $\Longrightarrow$ Rue et al. (2009), Blangiardo et al. (2013), Lindgren et al. (2015), Bakka et al. (2018)
- MCMC ... `Stan`? How? $\Longrightarrow$ `tmbstan` Monnahan and Kristensen (2018)

`tmbstan`

The models built in `TMB` pass their log-density and gradients calculations to the Bayesian samplers in `Stan` through a R interface. You only need to add the prior distribution on the parameters in your `TMB` code!

Application

Sea urchin (*Loxechinus albus*) is one of the most important benthic resource in Chile (Guisado (1987), Moreno et al. (2011)). Due to their large-scale spatial metapopulation structure, sea urchin subpopulations are interconnected by larval dispersion, so the recovery of local abundance depends on the distance and hydrodynamic characteristics of their spatial domain.

Figure 2: Common structure of the Sea urchin in a "patch". Reference: http://cocinafuturo.net/

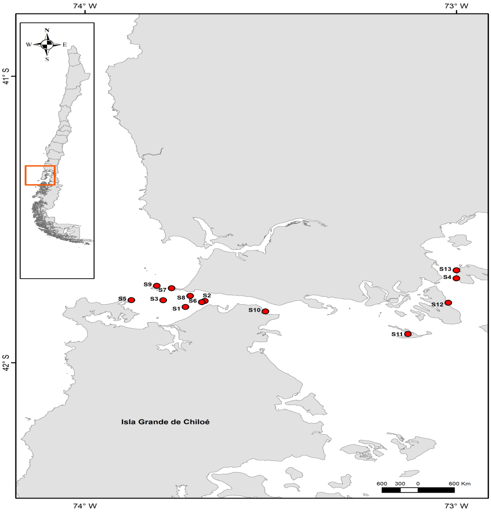

Figure 3: Sites of fishing of the Sea urchin (*Loxechinus albus*) in the south of Chile

What is the problem?

Currently, this resource is evaluated with classical stock assessment models, using standardized catch per unit effort (an index of relative abundance) as a key piece of information to determine catch quotas and achieve sustainability.

.... and what is a standardized catch per unit effort?

Catch per unit effort (CPUE) is a crucial variable in fishery science and often assumed proportional to the abundance for a particular fishery resource over time

What is the problem?

Currently, this resource is evaluated with classical stock assessment models, using standardized catch per unit effort (an index of relative abundance) as a key piece of information to determine catch quotas and achieve sustainability.

.... and what is a standardized catch per unit effort?

Catch per unit effort (CPUE) is a crucial variable in fishery science and often assumed proportional to the abundance for a particular fishery resource over time

What is the problem?

Currently, this resource is evaluated with classical stock assessment models, using standardized catch per unit effort (an index of relative abundance) as a key piece of information to determine catch quotas and achieve sustainability.

.... and what is a standardized catch per unit effort?

Catch per unit effort (CPUE) is a crucial variable in fishery science and often assumed proportional to the abundance for a particular fishery resource over time

What is the problem?

Currently, this resource is evaluated with classical stock assessment models, using standardized catch per unit effort (an index of relative abundance) as a key piece of information to determine catch quotas and achieve sustainability.

.... and what is a standardized catch per unit effort?

Catch per unit effort (CPUE) is a crucial variable in fishery science and often assumed proportional to the abundance for a particular fishery resource over time

Thus, we need to estimate the CPUE (index of relative abundance)!

Data

- Temporal observations: from 1996 to 2016 ("Year" variable as a "factor").
- Spatial observations: 13 sites of fishing ("sites" as a spatial random effects)
- Covariates: "Depth" (average depth of catches), "Quarter" (season of the year), and the variable "Market" (1 or 2).

Models

Table 1: Proposed models to obtain a relative abundance index

Models	Structure
Lognormal	$(y_i \mid \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$
Spatial Lognormal	$(y_i \mid \boldsymbol{u}, \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$
Gamma	$(y_i \mid \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$
Spatial Gamma	$(y_i \mid \boldsymbol{u}, \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$

Thus, we need to estimate the CPUE (index of relative abundance)!

Data

- Temporal observations: from 1996 to 2016 ("Year" variable as a "factor").
- Spatial observations: 13 sites of fishing ("sites" as a spatial random effects)
- Covariates: "Depth" (average depth of catches), "Quarter" (season of the year), and the variable "Market" (1 or 2).

Models

Table 1: Proposed models to obtain a relative abundance index

Models	Structure
Lognormal	$(y_i \mid \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$
Spatial Lognormal	$(y_i \mid \boldsymbol{u}, \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$
Gamma	$(y_i \mid \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$
Spatial Gamma	$(y_i \mid \boldsymbol{u}, \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta})$

Results

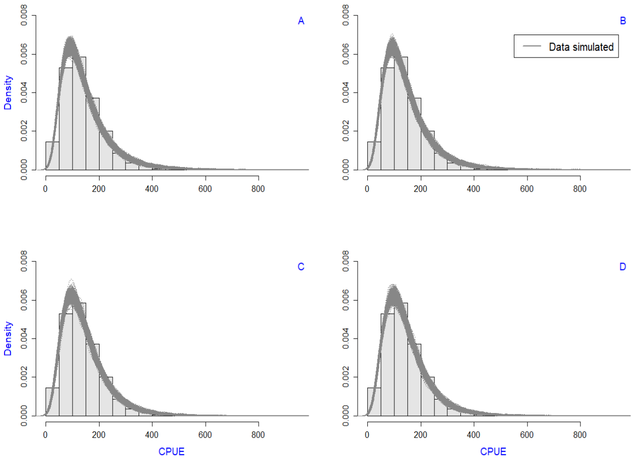

Figure 4: Observed data (histogram) and density estimates (lines) of 1,000 posterior predictive data simulated from Lognormal model, spatial Lognormal model, Gamma model and spatial Gamma model

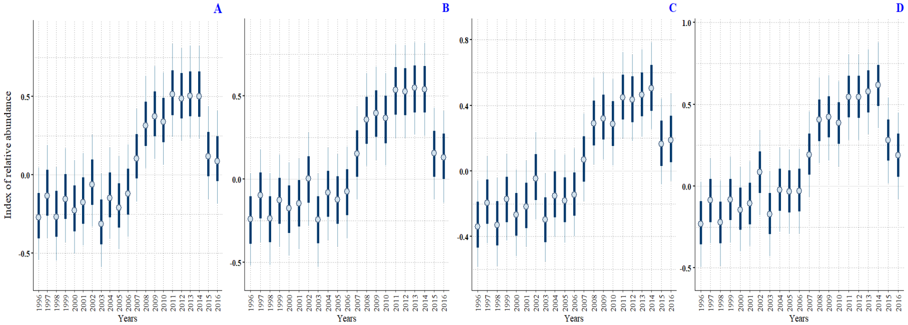

Figure 5: Comparison of the coefficients (relative abundance indices) estimated with Lognormal model, spatial Lognormal model, Gamma model and spatial Gamma model. The points are the values for the coefficients and the thick bars are uncertainty intervals computed from posterior draws with all chains merged in the MCMC method (90% credible interval).

Table 2: Comparison with LOO criterion for each proposed model. The elpd_diff measures the difference between each model relative to the best $\widehat{elpd}_n$ (the model in the first row) and se_diff is the standard error of the difference in elpd_diff.

Models	elpd_diff	se_diff
Spatial Gamma	0	0
Spatial Lognormal	-42	23
Lognormal	-88	18
Gamma	-125	28

To assess the potential effects of including sites with only one year of observations we made two additional model comparisons.

Table 3: Additional comparisons with LOO

Comparison excluding site 1		
Models	elpd_diff	se_diff
Spatial Gamma	0	0
Spatial Lognormal	-15	23
Gamma	-89	18
Lognormal	-95	28

Comparison excluding sites 1, 2, 3 and 8		
Models	elpd_diff	se_diff
Spatial Lognormal	0	0
Spatial Gamma	-5	22
Lognormal	-70	16
Gamma	-81	28

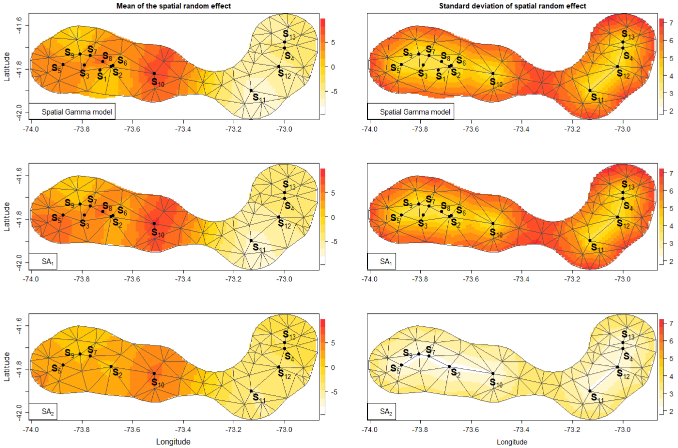

Figure 6: Mean (left) and standard deviation (right) of the spatial field estimated by the spatial Gamma model, and for the first (SA_1) and second (SA_2) sensitivity analysis respectively.

Conclusions

- Incorporating a spatial random effect to obtain a relative abundance index for sea urchin (L. albus) enabled better statistical performance than models without spatial dependence.
- Although the trends of the estimated indices for our case study with and without spatial effects were similar, statistical diagnostics clearly indicated that the spatial model outperformed the non-spatial version and fit the data better.
- This difference could have important impacts on the estimated status and trend of the stock, and ultimately the catch quota, so assessing the stock with both indices would be valuable.

Acknowledgments

- This research was done at Aalto University (Internship) supervised by Aki Vehtari
- Cole C Monnahan (NOAA)

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

ELSEVIER

Accounting for spatial dependence improves relative abundance estimates in a benthic marine species structured as a metapopulation

Joaquin Cavieres[a,*], Cole C. Monnahan[b], Aki Vehtari[c]

[a] Instituto de Estadística, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso, Chile
[b] Resource Ecology and Fisheries Management, National Marine Fisheries Service (NOAA), Seattle, WA, United States
[c] Department of Computer Science, Aalto University, Finland

Thank You

References I

Aquino, B., Castruccio, S., Gupta, V., and Howard, S. (2021). Spatial modeling of mid-infrared spectral data with thermal compensation using integrated nested laplace approximation. *Applied Optics*, 60(27):8609–8615.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with r-inla: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1443.

Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.

Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with r-inla. *Spatial and spatio-temporal epidemiology*, 4:33–49.

References II

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1):1–32.

Cavieres, J., Monnahan, C. C., and Vehtari, A. (2021). Accounting for spatial dependence improves relative abundance estimates in a benthic marine species structured as a metapopulation. *Fisheries Research*, 240:105960.

Gelman, A., Lee, D., and Guo, J. (2015). Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5):530–543.

Universidad de Valparaíso CHILE

References III

Ghattas, O. and Willcox, K. (2021). Learning physics-based models from data: perspectives from inverse problems and model reduction. *Acta Numerica*, 30:445–554.

Guisado, C. (1987). Historia de vida, reproducción y avances en el cultivo del erizo comestible chileno l. albus (molina, 1782)(echinoidea; echinidae). *Manejo y desarrollo pesquero*, pages 59–68.

Hough, I., Sarafian, R., Shtein, A., Zhou, B., Lepeule, J., and Kloog, I. (2021). Gaussian markov random fields improve ensemble predictions of daily 1 km pm2. 5 and pm10 across france. *Atmospheric Environment*, 264:118693.

References IV

INLA, U. D. L., LESTIMATION, S. P., and ALGERIE, P. E. (2021). Using inla/spde approach for estimating a spatial model for lung cancer mortality in algeria 2016. *Revue dEconomie et de Statistique Appliquée*, 18(1).

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. (2015). Tmb: automatic differentiation and laplace approximation. *arXiv preprint arXiv:1509.00660*.

References V

Levis, A., Lee, D., Tropp, J. A., Gammie, C. F., and Bouman, K. L. (2021). Inference of black hole fluid-dynamics from sparse interferometric measurements. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2340–2349.

Lindgren, F., Bolin, D., and Rue, H. (2021). The spde approach for gaussian and non-gaussian fields: 10 years and still running. *arXiv preprint arXiv:2111.01084*.

Lindgren, F., Rue, H., et al. (2015). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63(19):1–25.

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

References VI

Liu, X., Guillas, S., and Lai, M.-J. (2016). Efficient spatial modeling using the spde approach with bivariate splines. *Journal of Computational and Graphical Statistics*, 25(4):1176–1194.

Monnahan, C. C. and Kristensen, K. (2018). No-u-turn sampling for fast bayesian inference in admb and tmb: Introducing the adnuts and tmbstan r packages. *PloS one*, 13(5):e0197954.

Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. CRC Press.

Moraga, P., Dean, C., Inoue, J., Morawiecki, P., Noureen, S. R., and Wang, F. (2021). Bayesian spatial modelling of geostatistical data using inla and spde methods: A case study predicting malaria risk in mozambique. *Spatial and Spatio-temporal Epidemiology*, 39:100440.

References VII

Moreno, C. A., Molinet, C., Díaz, P., Díaz, M., Codjambassis, J., and Arévalo, A. (2011). Bathymetric distribution of the chilean red sea urchin (loxechinus albus, molina) in the inner seas of northwest patagonia: Implications for management. *Fisheries Research*, 110(2):305–311.

Neal, R. M. (2011). Mcmc using hamiltonian dynamics (handbook of markov chain monte carlo) ed s brooks et al. *A Gelman, G Jones XL Meng*.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

References VIII

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

Skaug, H. J. and Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics & Data Analysis*, 51(2):699–709.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 81(393):82–86.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449.

References IX

Whittle, P. (1963). Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994.

Zhang, H., Guilleminot, J., and Gomez, L. J. (2021). Stochastic modeling of geometrical uncertainties on complex domains, with application to additive manufacturing and brain interface geometries. *Computer methods in applied mechanics and engineering*, 385:114014.