

Bayesian multi-observation latent Gaussian models with inlabru

Statistics, LUSEM, Lund

Finn Lindgren, Finn.Lindgren@ed.ac.uk

School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh

2026-02-09

Traditional spatial covariance models vs RKHS inner products

- ▶ Gaussian random field: $u(\mathbf{s})$, $\mathbf{s} \in \mathcal{D}$ (subset of \mathbb{R}^d or a manifold such as \mathbb{S}^2)
- ▶ Moment characterisation:
 - ▶ Expectation $\mu(\mathbf{s}) = \mathbb{E}[u(\mathbf{s})]$
 - ▶ Covariance $\mathcal{R}(\mathbf{s}, \mathbf{s}') = \text{Cov}[u(\mathbf{s}), u(\mathbf{s}')]$, symmetric positive definite function.
- ▶ Precision operator; inverse covariance: $\mathcal{Q} = \mathcal{R}^{-1}$
In practice, easier conditions for valid models
- ▶ Reproducing Kernel Hilbert Space (RKHS) $H_{\mathcal{Q}}$: Inner product

$$\langle f, g \rangle_{H_{\mathcal{Q}}} = \langle f, \mathcal{Q}g \rangle_{\mathcal{D}}$$

and squared norm $\|f\|^2 = \langle f, f \rangle_{H_{\mathcal{Q}}}$

- ▶ $\mathbb{E}(u(\cdot) - \mu(\cdot) | \{u(\mathbf{s}_k)\}) \in H_{\mathcal{Q}}$ but $u(\cdot) - \mu(\cdot) \notin H_{\mathcal{Q}}$; the process is less smooth!

SPDEs and Gaussian random fields

- ▶ Spatial (and spatio-temporal) stochastic PDEs generate random field models:

$$\mathcal{L}u(\mathbf{s}) d\mathbf{s} = d\mathcal{W}(\mathbf{s})$$

$$Q_u = \mathcal{L}^* \mathcal{L}$$

$$\langle f, g \rangle_{H_Q} = \langle \mathcal{L}f, \mathcal{L}g \rangle_{\mathcal{D}}$$

Can work directly with the precision or inner product; no need to know the covariance.

- ▶ Regular Matérn field are the stationary solutions to a spatial SPDE (Lindgren et al, 2011, JRSS-B)

$$(\kappa^2 - \Delta)^{\alpha_s/2} u(\mathbf{s}) d\mathbf{s} = d\mathcal{W}(\mathbf{s})/\tau$$

- ▶ The iterated dampened heat equation is a simple non-separable space-time SPDE (Lindgren et al, 2024, SORT)

$$\left[\phi \frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha_s/2} \right]^{\alpha_t} u(\mathbf{s}, t) dt = d\mathcal{E}_{(\kappa^2 - \Delta)^{\alpha_e}}(\mathbf{s}, t)/\tau$$

Bayesian latent Gaussian process models

General latent Gaussian hierarchical model structure

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$$

$$\mathbf{x}|\boldsymbol{\theta} \sim \text{N}(\boldsymbol{\mu}_x(\boldsymbol{\theta}), \mathbf{Q}_x(\boldsymbol{\theta})^{-1})$$

$$\mathbf{y}|\mathbf{x}, \boldsymbol{\theta} \sim p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta})$$

Generalised additive models (GAMs) with Gaussian random fields (GRFs):

$$\mathbf{x} = (\boldsymbol{\beta}, \mathbf{u}_1, \dots, \mathbf{u}_K)$$

$$g(\text{E}[y_i|\mathbf{x}, \boldsymbol{\theta}]) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \sum_k f_k(z_{ik}; \mathbf{u}_k)$$

and e.g. $y_i|\mathbf{x}, \boldsymbol{\theta} \sim \text{N}(\boldsymbol{\eta}, \sigma_y^2)$ or $y_i|\mathbf{x}, \boldsymbol{\theta} \sim \text{Po}(\exp(\eta_i))$

We want to estimate the parameters of the GRFs, $\boldsymbol{\theta}$, the GRF processes values $f_k(\cdot)$ at observed and unobserved locations, and quantify the uncertainty in these estimates.

The Matérn-Whittle-Markov GRF/SPDE/GMRF connection

Each $f_k(\cdot)$ is a function of space, time, or a covariate, and is approximated by

$$f_k(\mathbf{z}_{ik}; \mathbf{u}_k) = \sum_j \psi_{kj}(\mathbf{z}_{ik}) u_{kj},$$

where $\psi_{kj}(\mathbf{z}_{ik})$ are basis functions, e.g. finite element basis functions. Matérn fields are solutions to the spatial SPDE

$$\begin{aligned} (\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} (\tau u(\mathbf{s})) &= \kappa^\gamma dW(\mathbf{s}) \\ u(\mathbf{s}) &\approx \sum_j \psi_j(\mathbf{s}) u_j, \mathbf{u} \sim N(\mathbf{0}, \mathbf{Q}_u^{-1}) \end{aligned}$$

where \mathbf{Q} is the precision matrix of the GRF/SPDE/GMRF representation.

When α is an integer, FEM yields a sparse matrix \mathbf{Q}_u , and $u(\mathbf{s})$ is a Markov random field (Lindgren et al, 2011).

For non-integers, $u(\mathbf{s})$ can be closely approximated by a sum of a few Markov processes (Bolin and Kirchner, 2020).

Parameter estimation and spatial prediction

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}) \approx \frac{p(\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x})}{p_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*}$$

where $p_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation to the conditional posterior density.

The INLA software uses numerical integration over $\boldsymbol{\theta}$ together with variational Bayes corrections $p_{GG}()$ of the Gaussian approximations to obtain the posterior marginal densities of \mathbf{x} :

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \int p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\ &\approx \sum_j p_{GG}(\mathbf{x}|\boldsymbol{\theta}^{(j)}, \mathbf{y})p(\boldsymbol{\theta}^{(j)}|\mathbf{y})w_j \end{aligned}$$

The inner core of the Integrated Nested Laplace method

- ▶ Latent Gaussian model structure (Bayesian GAMs with Gaussian process components)

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}) \quad (\text{precision parameters}) \quad \eta(\mathbf{s}, t) = \sum_{k=1}^m \psi_k(\mathbf{s}, t) u_k \quad (\text{predictor})$$

$$\mathbf{u} | \boldsymbol{\theta} \sim N[\boldsymbol{\mu}_u, \mathbf{Q}_u^{-1}] \quad (\text{latent field}) \quad \mathbf{y} | \boldsymbol{\theta}, \mathbf{u} \sim p(\mathbf{y} | \boldsymbol{\theta}, \eta) \quad (\text{observations})$$

- ▶ Conditional log-posterior mode ($\boldsymbol{\mu}_{u|y}$) and Hessian ($\mathbf{Q}_{u|y}$), for each $\boldsymbol{\theta}$, by iteration:

$$\mathbf{g}_y^* = - \left. \frac{d}{d\mathbf{u}} \log p(\mathbf{y} | \boldsymbol{\theta}, \eta) \right|_{\mathbf{u}=\mathbf{u}^*}$$

$$\mathbf{H}_y^* = - \left. \frac{d^2}{d\mathbf{u}d\mathbf{u}^\top} \log p(\mathbf{y} | \boldsymbol{\theta}, \eta) \right|_{\mathbf{u}=\mathbf{u}^*}$$

$$\mathbf{Q}_{u|y} = \mathbf{Q}_u + \mathbf{H}_y^*$$

$$\mathbf{Q}_{u|y}(\boldsymbol{\mu}_{u|y} - \boldsymbol{\mu}_u) = \mathbf{Q}_u(\mathbf{u}^* - \boldsymbol{\mu}_u) - \mathbf{g}_y^*$$

General observation models; rarely direct observations

- ▶ Point-referenced data; additive noise, counts, presence-absence, etc.
- ▶ Aggregated data; spatial averages/totals, counts, presence-absence, etc.
- ▶ Point process data. Poisson process log-likelihood function:

$$-\int \lambda(\mathbf{s}) d\mathbf{s} + \sum_i \log[\lambda(\mathbf{y}_i)] \approx -\sum_j w_j \exp[\eta(\mathbf{s}_j)] + \sum_i \eta(\mathbf{y}_i)$$

where $\{(\mathbf{s}_j, w_j)\}$ is a numerical integration scheme over the sampled region of space (Simpson et al, 2016, Biometrika)

- ▶ Semi-parametric densities, (animal) movement kernels, Hawkes processes, etc

Modern data analysis problems may involve multiple types of observations in a single model; each type may have a specialised method, but we need a general system for blending the information, which is provided by the general Bayesian framework.

Probabilistic latent Gaussian model specification

In plain INLA, the syntax mimics other R modelling packages such as `mgcv` and `lme4`, with formulae defining the predictor structure.

In `inlabru`, the latent components and observation models are defined separately, and combined into a full model definition. This allows a wide range of easy-to-specify extensions, model definitions, and flexible data wrangling.

- ▶ List of latent components, each with a Gaussian process definition, and a mapping between locations/covariate values, the latent variables, and the resulting "effect".
- ▶ List of observation models, each with a likelihood function linking a predictor expression to the observation distribution, and a mapping between the latent components/effects and the predictor.

The `inlabru` model specification is a type of probabilistic programming, and is increasingly using automatic differentiation techniques to compute Jacobians.

Basic joint model example; misaligned covariate/response measurements

$$(\tau_z, \mu_z, Q_z) \sim p(\tau_z)p(\mu_z)p(Q_z)$$

$$z(\cdot)\text{-coefficients} \sim N(\mu_z, Q_z^{-1})$$

$$\epsilon_i^z \sim N(0, \tau_z^{-1})$$

$$z_i^{\text{obs}} = z(\mathbf{s}_i) + \epsilon_i^z, \mathbf{s}_i \in \mathcal{S}_z$$

$$y_j^{\text{obs}} = \beta_0 + \beta_1 z(\mathbf{s}_j) + u(\mathbf{s}_j) + \epsilon_j^y, \mathbf{s}_j \in \mathcal{S}_y$$

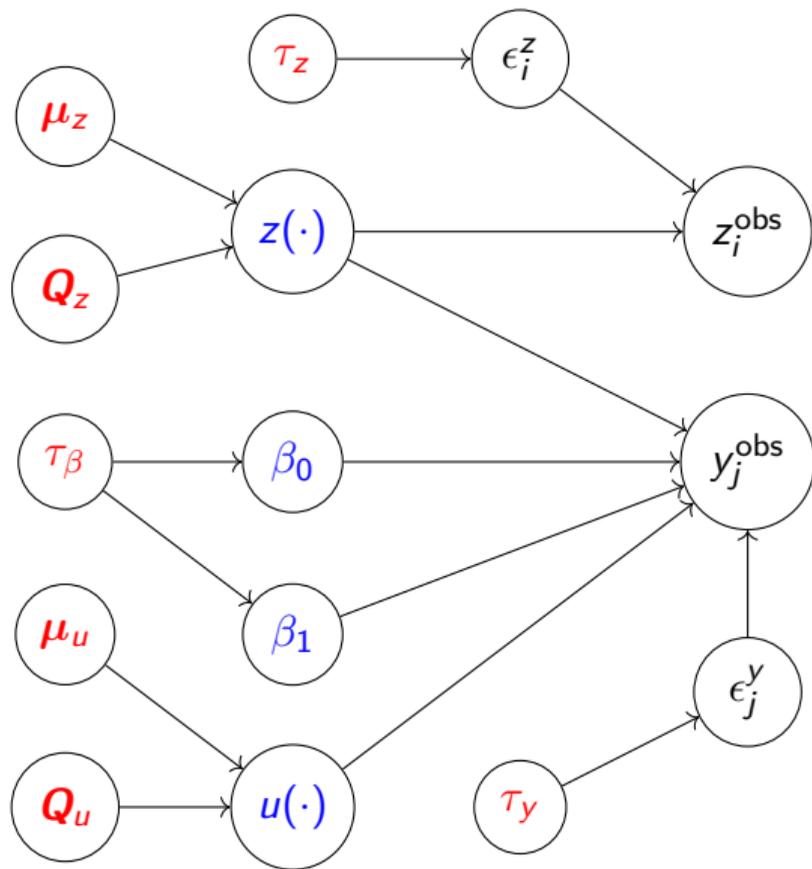
$$\epsilon_i^y \sim N(0, \tau_y^{-1})$$

$$u(\cdot)\text{-coefficients} \sim N(\mu_u, Q_u^{-1})$$

$$\beta_0, \beta_1 \sim N(0, \tau_\beta^{-1})$$

$$\tau_y \sim p(\tau_y), \tau_\beta \text{ fixed}$$

$$(\mu_u, Q_u) \sim p(\mu_u)p(Q_u)$$



Basic joint models and model mis-specification handling

- ▶ Joint models allow uncertainty propagation between sub-models:

$$z_i^{\text{obs}} = z(\mathbf{s}_i) + \epsilon_i^z$$

$$y_j^{\text{obs}} = \beta_0 + \beta_1 z(\mathbf{s}_j) + u(\mathbf{s}_j) + \epsilon_j^y$$

- ▶ Two-step approaches (e.g. fit covariate model, then fit main model) can avoid improper feedback problems, but need to propagate uncertainty:

$$z_i^{\text{obs}} = z(\mathbf{s}_i) + \epsilon_i^z$$

$$z^{\text{post}} \sim \text{N}(\boldsymbol{\mu}_{z|z^{\text{obs}}}, \mathbf{Q}_{z|z^{\text{obs}}}^{-1}) \quad (\text{convenient posterior approximation})$$

$$y_j^{\text{obs}} = \beta_0 + \beta_1 z^{\text{post}}(\mathbf{s}_j) + u(\mathbf{s}_j) + \epsilon_j^y$$

- ▶ Spatially consistent modelling of aggregated counts:

$$\lambda(\mathbf{s}) = \exp[\beta_0 + \beta_1 z(\mathbf{s}) + u(\mathbf{s})]$$

$$y_j^{\text{obs}} \sim \text{Poisson} \left(\int_{\mathcal{B}_j} \lambda(\mathbf{s}) \, d\mathbf{s} \right)$$

Non-linear predictors

The original motivation for the `inlabru` package was ecological transect distance sampling, requiring a model for imperfect detections:

$$\lambda_{\text{apparent}}(\mathbf{s}; \mathbf{u}, \mathbf{v}) = \lambda(\mathbf{s}; \mathbf{u})h(\mathbf{s}; \mathbf{v}),$$

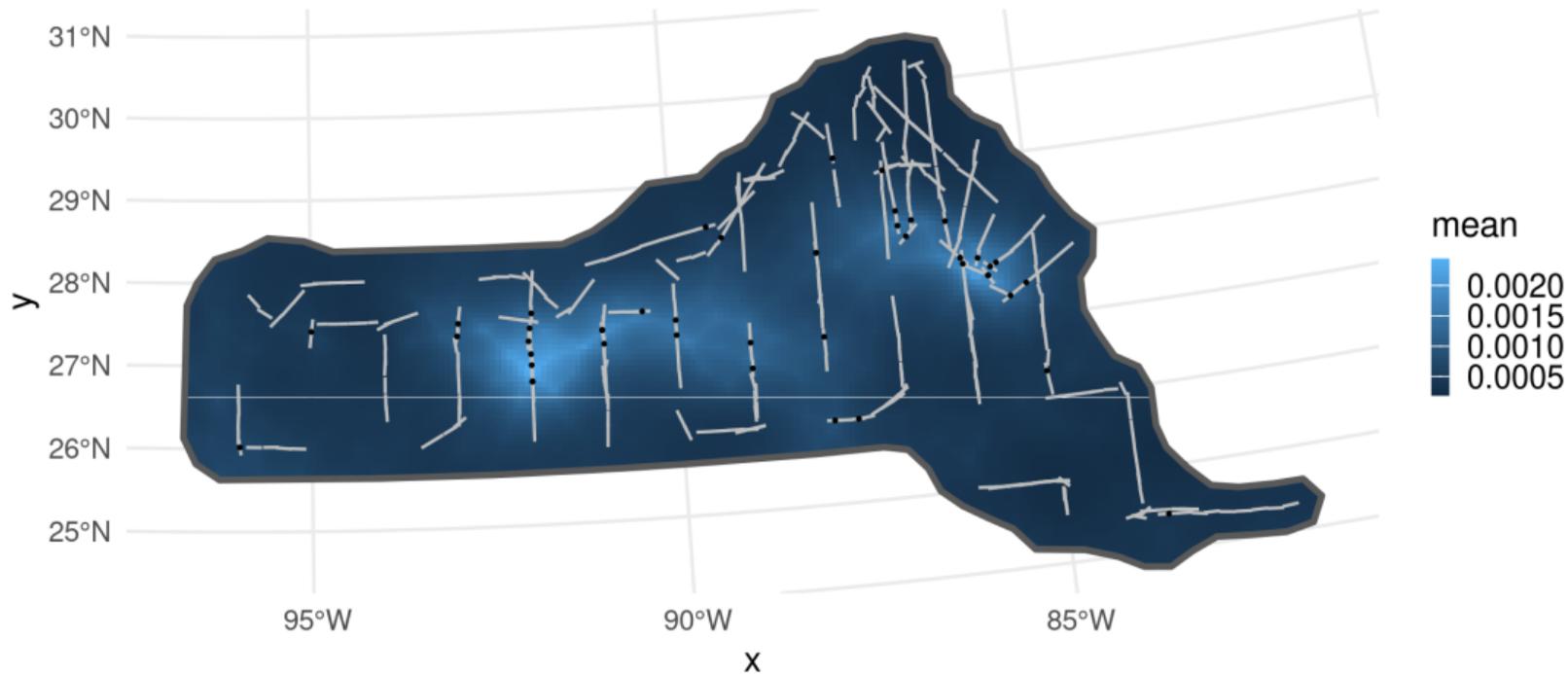
where $h(\mathbf{s}; \mathbf{v})$ is the detection probability for a point located at \mathbf{s} , and \mathbf{v} is a vector of parameters for the detection function, such as the *hazard rate* model:

$$h(\mathbf{s}; \mathbf{v}) = 1 - \exp \left[- \left(\frac{\text{distance}[\mathbf{s}, \text{transect}]}{\phi_1(v_1)} \right)^{-\phi_2(v_2)} \right] \quad (\text{quantile transformations } \phi_1 \text{ and } \phi_2)$$

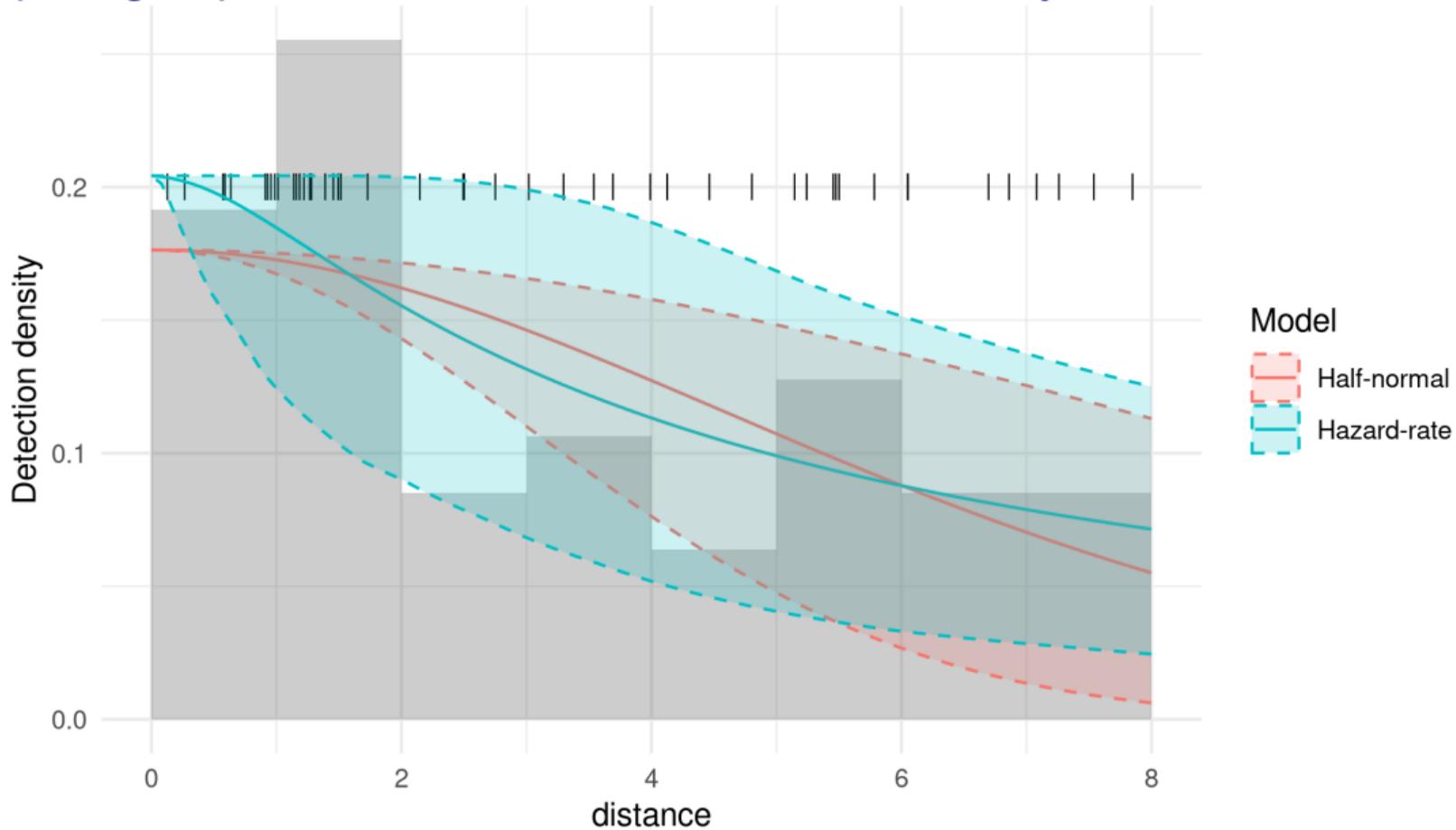
The `inlabru` package solves this by iterating the INLA method on a linearisation w.r.t. \mathbf{u} and \mathbf{v} of the non-linear predictor

$$\eta(\mathbf{s}; \mathbf{u}, \mathbf{v}) = \log[\lambda(\mathbf{s}; \mathbf{u})] + \log[h(\mathbf{s}; \mathbf{v})].$$

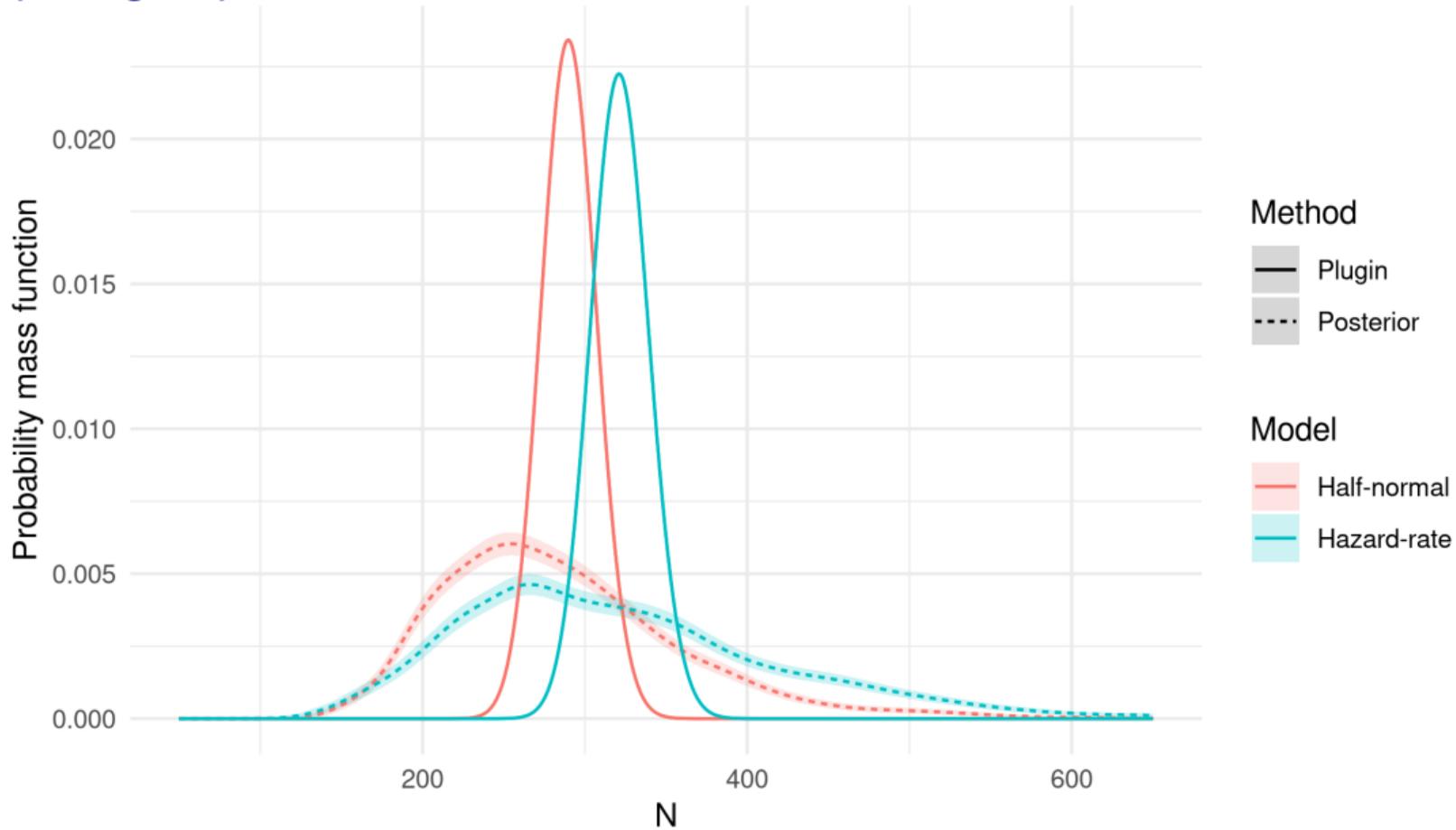
Dolphin group detection; estimated density field



Dolphin group detection; estimated detection density



Dolphin group detection; estimated total count



Numerical challenges

- ▶ $\mathbf{Q}_{x|\theta,y}$ is a large, (usually) sparse matrix
- ▶ Need to solve linear systems of the form $\mathbf{Q}_{x|\theta,y}\mathbf{x} = \mathbf{b}$
- ▶ Need to evaluate marginal variances $\left[\mathbf{Q}_{x|\theta,y}^{-1}\right]_{ii}$
(Cholesky plus Takahashi recursions, but what about large problems where Cholesky is unavailable?)
- ▶ Need to evaluate log-determinants $\log|\mathbf{Q}_{x|\theta}|$ and $\log|\mathbf{Q}_{x|\theta,y}|$
- ▶ Gradient descent methods can make use of the log-determinant derivative $\text{tr}\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\theta}\right)$

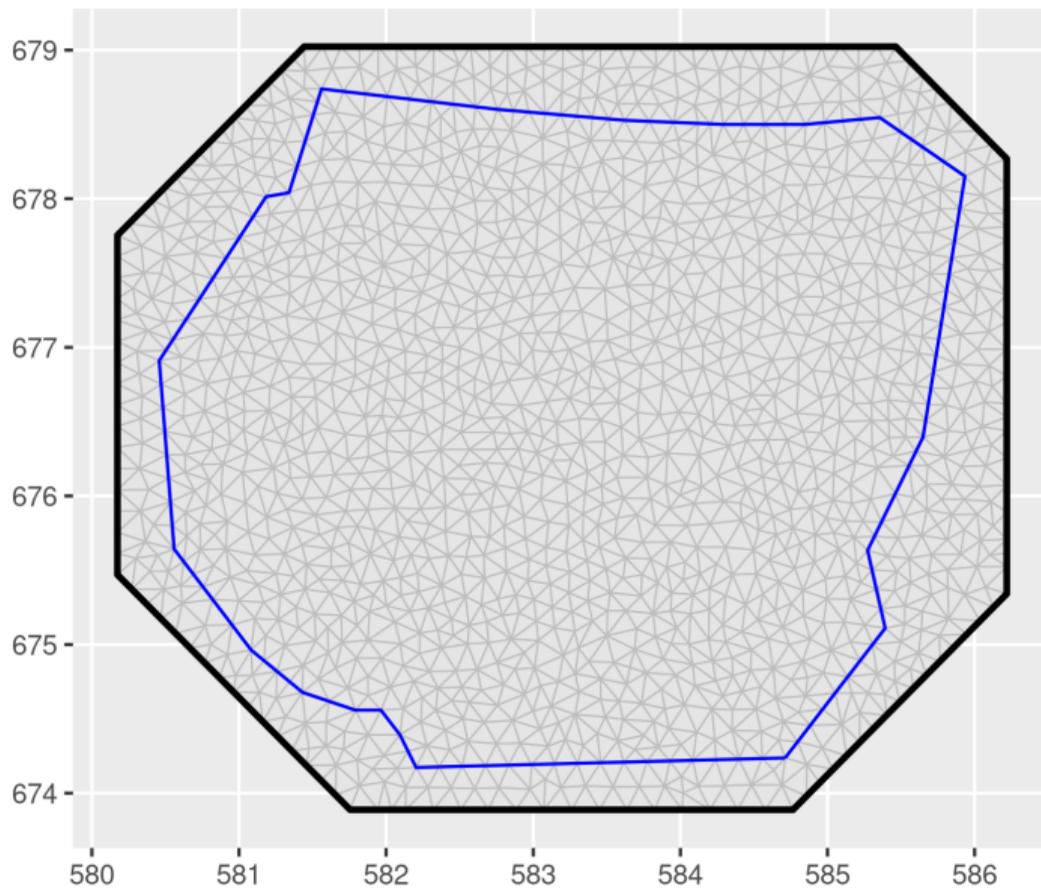
Modelling and computational challenges

- ▶ How to automate uncertainty propagation in multi-step approaches, e.g. for aggregated data problems (current work with Man Ho Suen and Stephen Jun Villejo)
- ▶ How to construct sensible/interpretable prior distributions for anisotropy and non-stationarity (current work with Liam Llamazares-Elias: Penalised complexity priors for anisotropy and non-stationarity)
- ▶ Scaling things up to large ($\sim 10^{11}$ unknowns) space-time problems with complex observation models; observations involve sums of several processes on different time-scales, systematic biases, and irregular observation patterns; (past work in the EUSTACE project, still relevant)

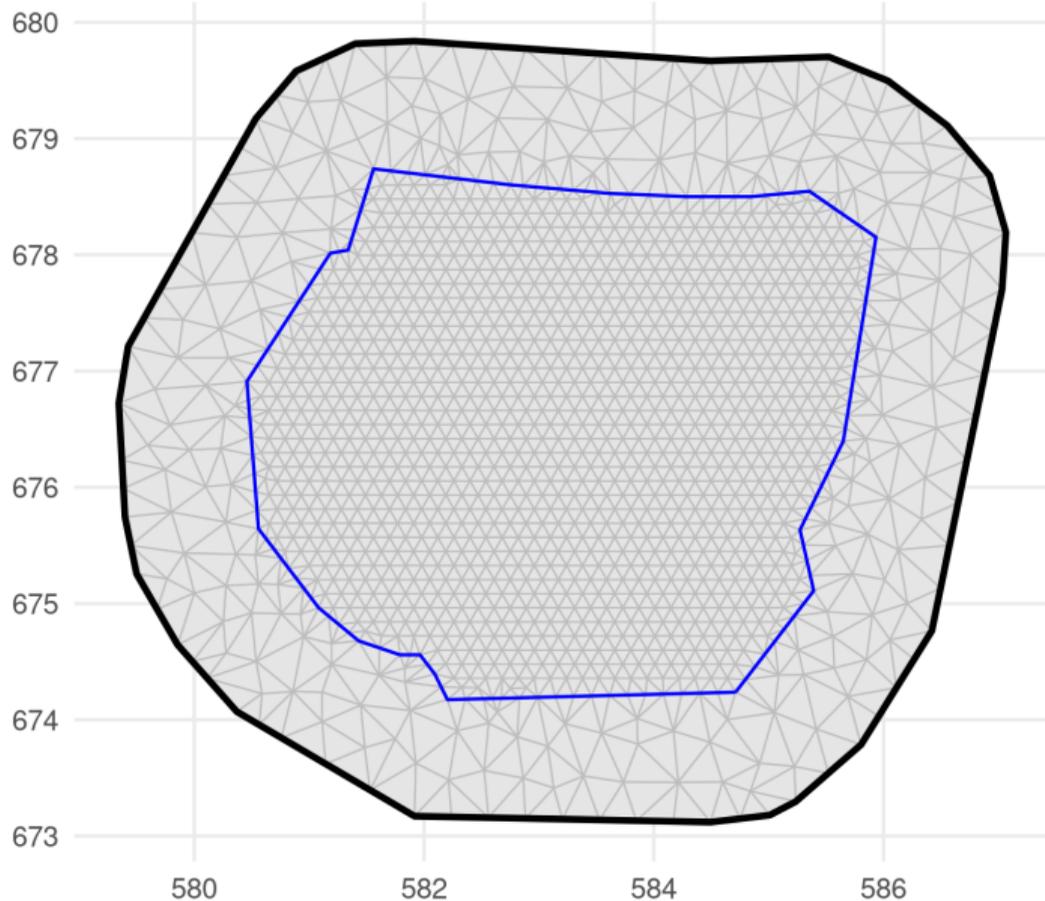
Partial inversion beyond Takahashi recursions

- ▶ Monte Carlo estimation; expensive, as may need to use iterative methods to construct each sample
- ▶ Iterative combinations of MC and local exact partial inversion; not as nice as we would like.
- ▶ Idea for space and space-time models: Need to jointly solve for the (posterior) marginal variances *and* the local shape of the correlation function. There appears to be a way to formulate this problem as a multidimensional (likely non-linear) PDE, which might then be solvable using a single run of an iterative PDE solver.

Challenge: explaining FEM meshes to non-PDE scientists

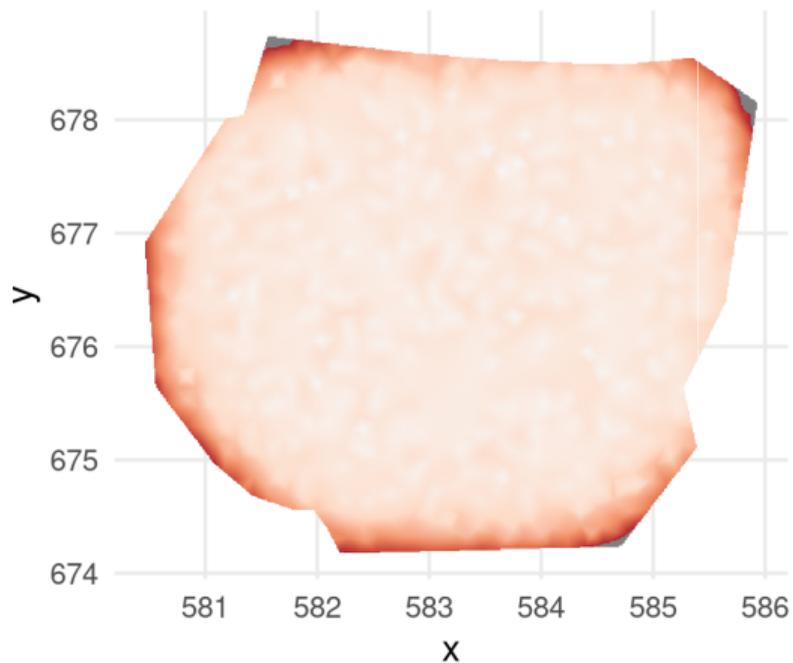


The methods are flexible, but some choices are generally better

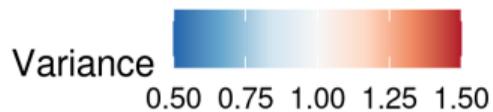
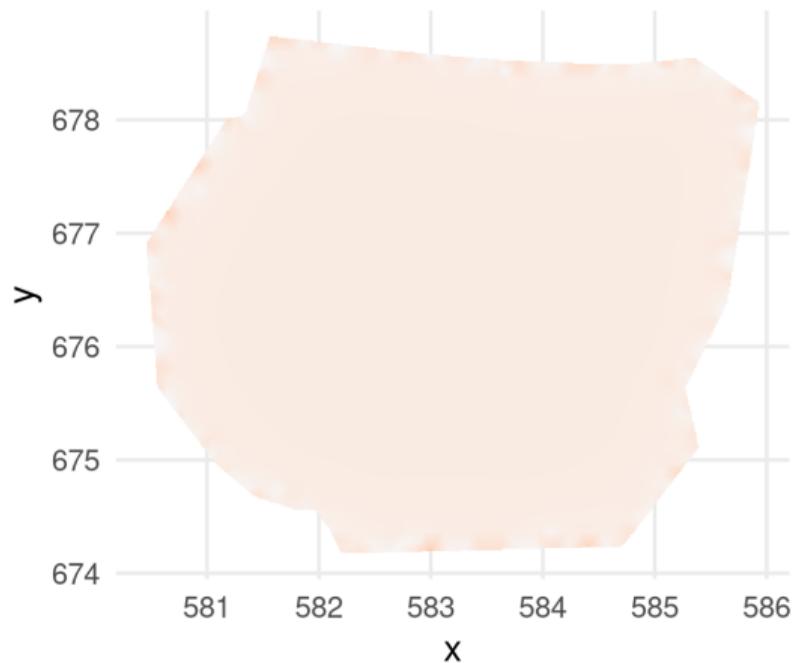


Variances

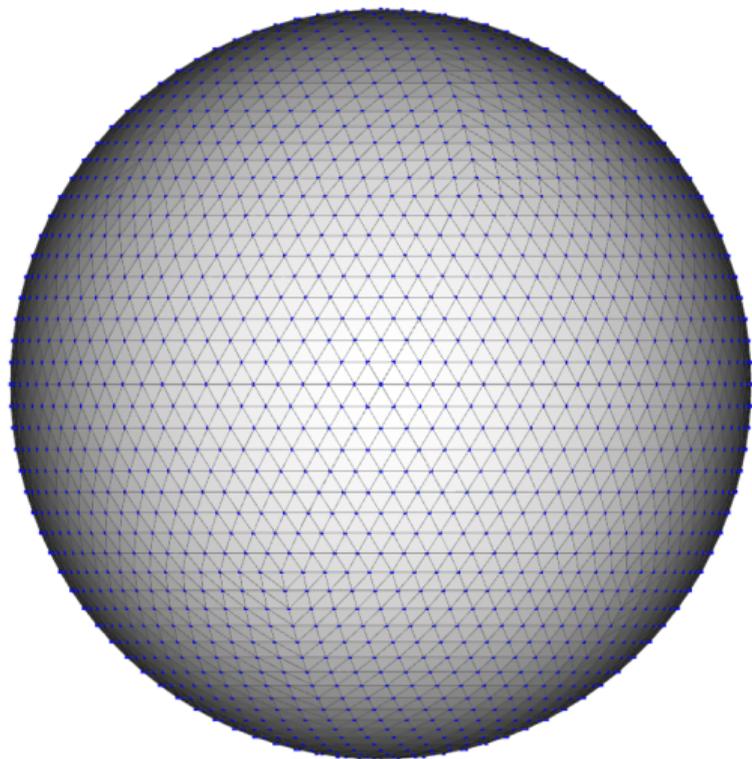
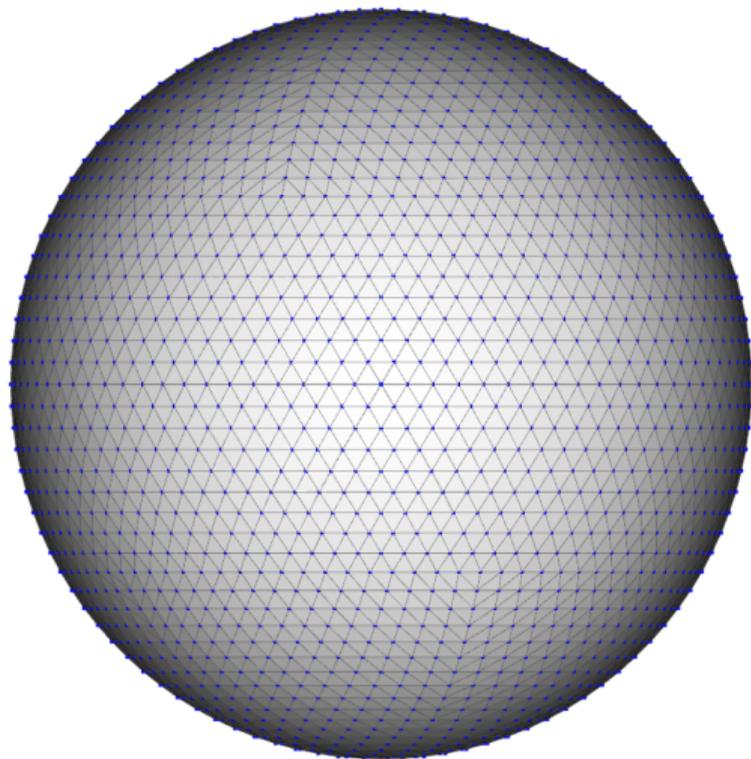
Current mesh, dof = 1479



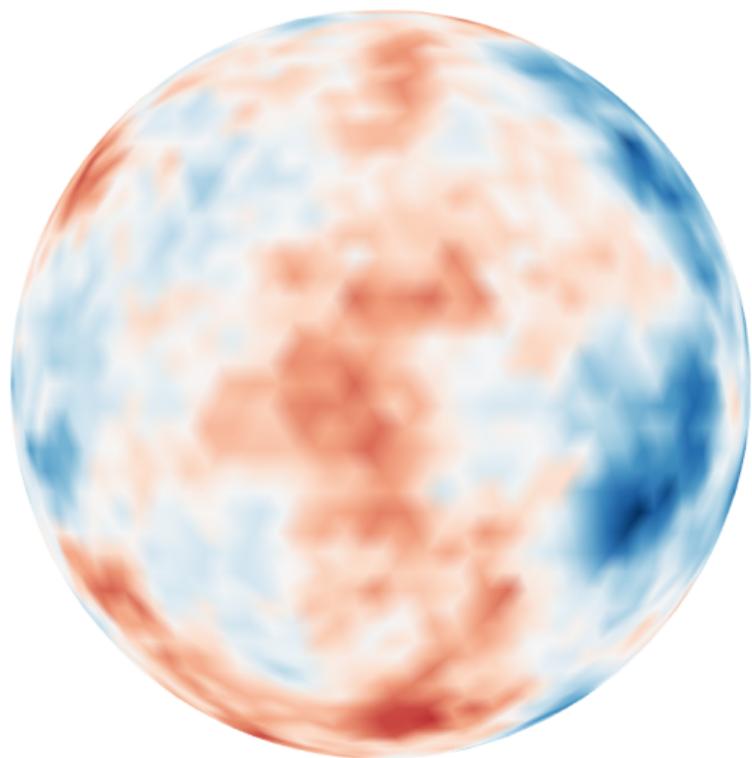
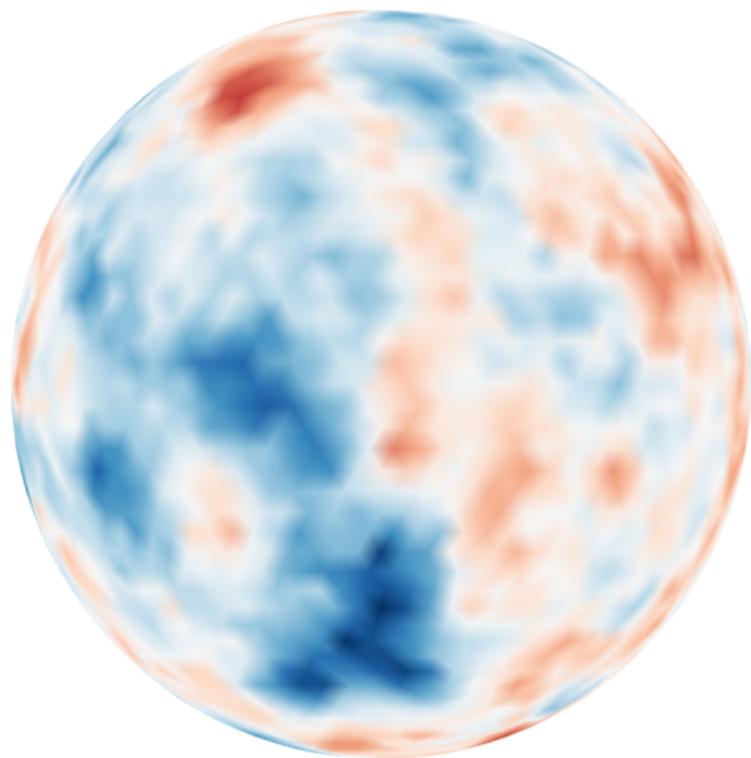
Alternative mesh, dof = 1505



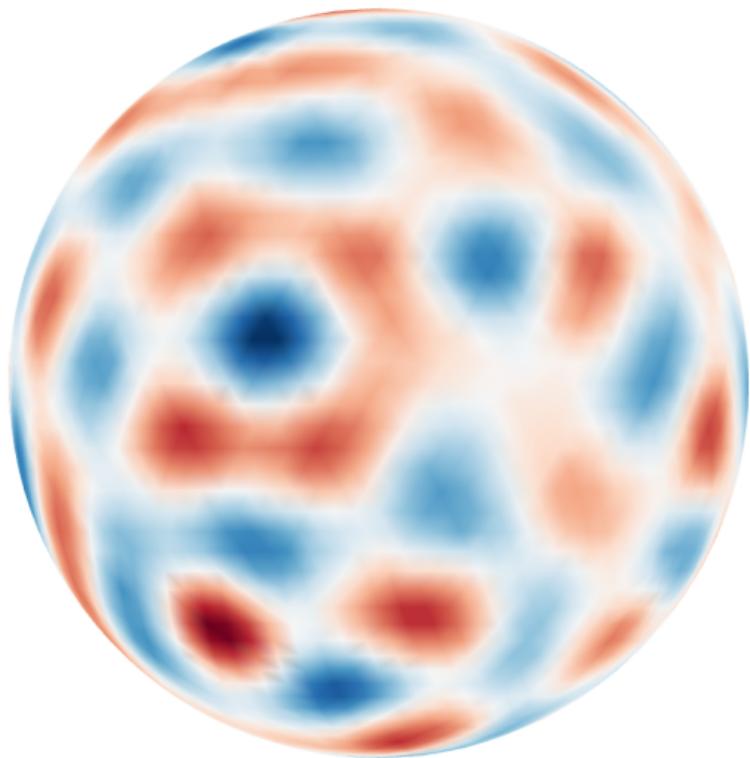
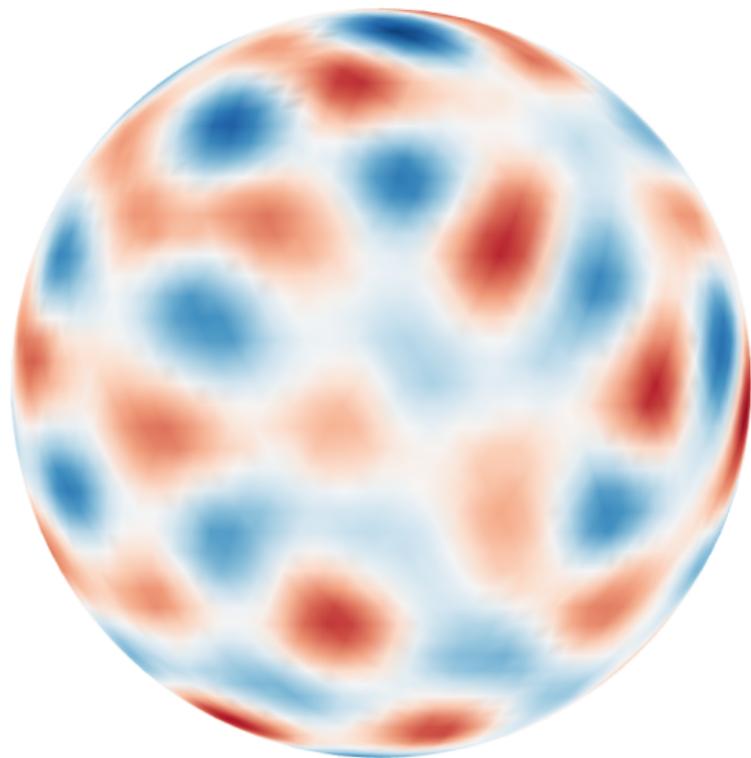
Nearly regular mesh on the unit sphere



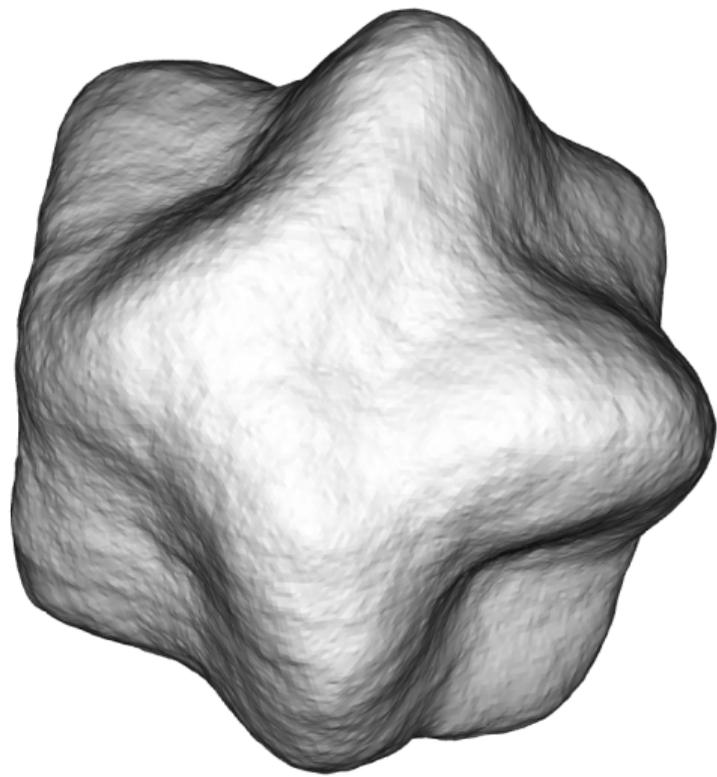
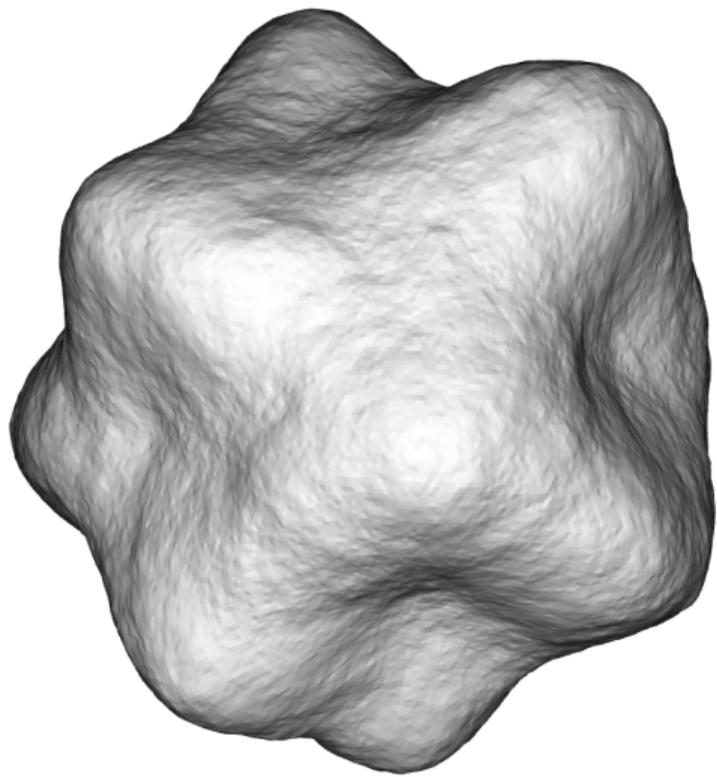
Whittle-Matérn field on the unit sphere



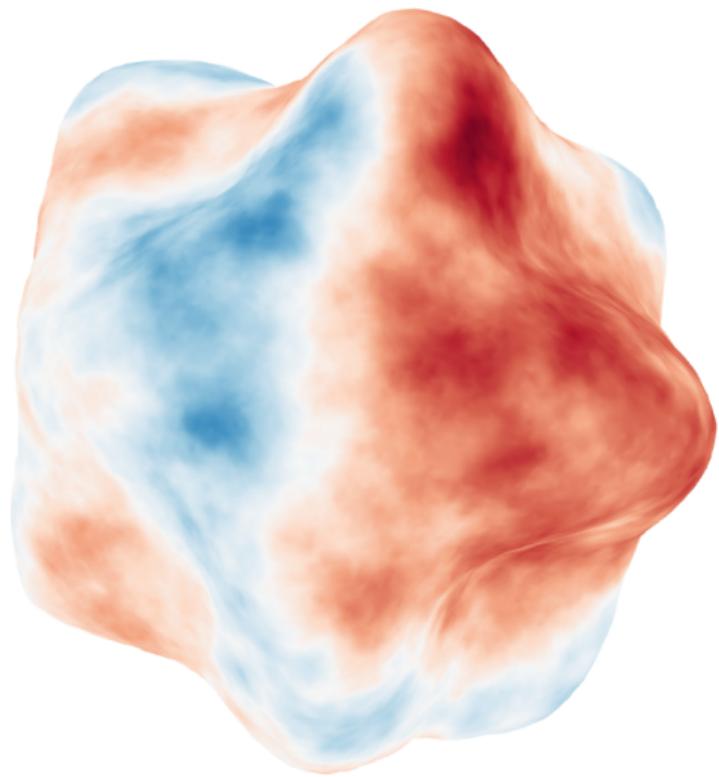
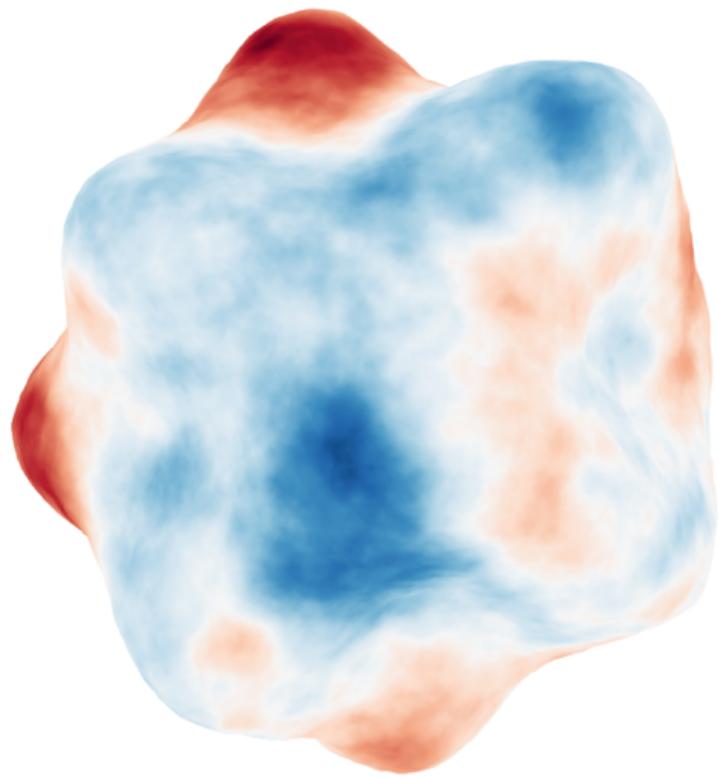
Oscillatory field on the unit sphere (modified Whittle SPDE)



'Potato'



Potato field (application to atrial manifolds in Coveney et al, 2020)



References

- ▶ The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running (Lindgren et al, 2022, Spatial Statistics)
<https://doi.org/10.1016/j.spasta.2022.100599>
- ▶ Going off grid: computationally efficient inference for log-Gaussian Cox processes (Simpson et al, 2016, Biometrika) <https://doi.org/10.1093/biomet/asv064>
- ▶ A diffusion-based spatio-temporal extension of Gaussian Matérn fields (Lindgren et al, 2024, SORT) <https://doi.org/10.57645/20.8080.02.13>
- ▶ A data fusion model for meteorological data using the INLA-SPDE method (Stephen Jun Villejo et al, 2025, JRSS-C)
- ▶ Coherent Disaggregation and Uncertainty Quantification for Spatially Misaligned Data (Man Ho Suen et al, 2026, to appear in Environmetrics)
<https://arxiv.org/abs/2502.10584>
- ▶ Documentation and examples: R-INLA (<https://www.r-inla.org/>), `inlabru`, `fmesher` (<https://inlabru-org.github.io/inlabru/> and [.../fmesher/](https://inlabru-org.github.io/fmesher/))

Non-separable space-time: Matérn driven heat equation

The iterated dampened heat equation is a simple non-separable space-time SPDE (Lindgren et al, 2024, SORT)

$$\left[\phi \frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha_s/2} \right]^{\alpha_t} u(\mathbf{s}, t) dt = d\mathcal{E}_{(\kappa^2 - \Delta)^{\alpha_e}}(\mathbf{s}, t) / \tau$$

For constant parameters, $u(\mathbf{s}, t)$ has spatial Matérn covariance (for each t) on \mathbb{R}^d and a generalised Matérn-Whittle covariance on \mathbb{S}^2 .

Smoothness properties (can be derived from the spectra):

$$\begin{cases} \nu_t = \min \left[\alpha_t - \frac{1}{2}, \frac{\nu_s}{\alpha_s} \right], \\ \nu_s = \alpha_e + \alpha_s \left(\alpha_t - \frac{1}{2} \right) - \frac{d}{2}, \\ \beta_s = 1 - \frac{\alpha_e}{\nu_s + d/2}, \end{cases} \quad \begin{cases} \alpha_t = \nu_t \max \left(1, \frac{\beta_s}{\beta_*(\nu_s, d)} \right) + \frac{1}{2}, \\ \alpha_s = \frac{\nu_s}{\nu_t} \min \left(\frac{\beta_s}{\beta_*(\nu_s, d)}, 1 \right), \\ \alpha_e = \frac{1 - \beta_s}{\beta_*(\nu_s, d)} \nu_s = (\nu_s + d/2)(1 - \beta_s), \end{cases}$$

where $\beta_*(\nu_s, d) = \frac{\nu_s}{\nu_s + d/2}$, and $\beta_s \in [0, 1]$ is a non-separability parameter.

Smoothness properties

α_t	α_s	α_e	Type	ν_t	ν_s
α_t	α_s	α_e	General	$\min \left[\alpha_t - \frac{1}{2}, \frac{\nu_s}{\alpha_s} \right]$	$\alpha_e + \alpha_s \left(\alpha_t - \frac{1}{2} \right) - \frac{d}{2}$
α_t	0	α_e	Separable	$\alpha_t - \frac{1}{2}$	$\alpha_e - \frac{d}{2}$
α_t	α_s	$\frac{d}{2}$	Critical	$\alpha_t - \frac{1}{2}$	$\alpha_s \left(\alpha_t - \frac{1}{2} \right)$
α_t	α_s	0	Fully non-separable	$\alpha_t - \frac{1}{2} - \frac{d}{2\alpha_s}$	$\alpha_s \left(\alpha_t - \frac{1}{2} \right) - \frac{d}{2}$
1	2	$\alpha_e > \frac{d}{2}$	Sub-critical diffusion	1/2	$\alpha_e + 1 - \frac{d}{2}$
1	2	$\frac{d}{2}$	Critical diffusion	1/2	1
1	2	$\frac{d}{2} - 1 < \alpha_e < \frac{d}{2}$	Super-critical diffusion	$\nu_s/2$	$\alpha_e + 1 - \frac{d}{2}$
1	0	2	Separable	1/2	$2 - \frac{d}{2}$
3/2	2	0	Fractional diffusion	$1 - \frac{d}{4}$	$2 - \frac{d}{2}$
2	2	0	Iterated diffusion	$\frac{3}{2} - \frac{d}{4}$	$3 - \frac{d}{2}$