Embedding numerical stochastic PDE models in Bayesian inference for latent Gaussian models

Joint ProbAI/ACM Colloquium, Edinburgh

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

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

2025-11-05

Traditional spatial covariance models vs RKHS inner products

- Gaussian random field: $u(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{D}$ (subset of $\mathbb{R}^d$ or a manifold such as $\mathbb{S}^2$)
- Moment characterisation:
  - Expectation $\mu(\boldsymbol{s}) = \mathsf{E}[u(\boldsymbol{s})]$
  - Covariance $\mathcal{R}(\boldsymbol{s}, \boldsymbol{s}') = \mathsf{Cov}[u(\boldsymbol{s}), u(\boldsymbol{s}')]$, 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 $\|f\|^2 = \langle f, f \rangle_{H_{\mathcal{Q}}}$
- $\mathsf{E}(u(\cdot) - \mu(\cdot) | \{u(\boldsymbol{s}_k)\}) \in H_{\mathcal{Q}}$ 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:

$$\mathcal{L}u(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = \mathrm{d}\mathcal{W}(\boldsymbol{s})$$
$$\mathcal{Q}_u = \mathcal{L}^*\mathcal{L}$$
$$\langle f, g\rangle_{H_{\mathcal{Q}}} = \langle \mathcal{L}f, \mathcal{L}g\rangle_{\mathcal{D}}$$

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

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)

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

For constant parameters, $u(\boldsymbol{s}, t)$ 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):

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

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	$\min\left[\alpha_t - \frac{1}{2}, \frac{\nu_s}{\alpha_s}\right]$	$\alpha_e + \alpha_s(\alpha_t - \frac{1}{2}) - \frac{d}{2}$
α_t	0	α_e	Separable	$\alpha_t - \frac{1}{2}$	$\alpha_e - \frac{d}{2}$
α_t	α_s	$\frac{d}{2}$	Critical	$\alpha_t - \frac{1}{2}$	$\alpha_s(\alpha_t - \frac{1}{2})$
α_t	α_s	0	Fully non-separable	$\alpha_t - \frac{1}{2} - \frac{d}{2\alpha_s}$	$\alpha_s(\alpha_t - \frac{1}{2}) - \frac{d}{2}$
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	$\frac{3}{2} - \frac{d}{4}$	$3 - \frac{d}{2}$

Bayesian latent Gaussian process models

General latent Gaussian hierarchical model structure

$$
\begin{aligned}
\boldsymbol{\theta} &\sim p(\boldsymbol{\theta}) \\
\boldsymbol{x}|\boldsymbol{\theta} &\sim \mathrm{N}(\boldsymbol{\mu}_x(\boldsymbol{\theta}), \boldsymbol{Q}_x(\boldsymbol{\theta})^{-1}) \\
\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta} &\sim p(\boldsymbol{y} \mid \boldsymbol{x},\boldsymbol{\theta})
\end{aligned}
$$

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

$$
\begin{aligned}
\boldsymbol{x} &= (\beta, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_K) \\
g(\mathrm{E}[y_i|\boldsymbol{x},\boldsymbol{\theta}]) = \boldsymbol{\eta} &= \boldsymbol{X}\beta + \sum_k f_k(z_{ik}; \boldsymbol{u}_k)
\end{aligned}
$$

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

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

$$f_k(z_{ik}; \boldsymbol{u}_k) = \sum_j \psi_{kj}(z_{ik}) u_{kj},$$

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

$$(\kappa^2 - \nabla \cdot \nabla)^{\alpha/2}(\tau u(\boldsymbol{s})) = \kappa^\gamma \mathrm{d}W(\boldsymbol{s})$$

$$u(\boldsymbol{s}) \approx \sum_j \psi_j(\boldsymbol{s}) u_j, \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{Q}_u^{-1})$$

where $\boldsymbol{Q}$ is the precision matrix of the GRF/SPDE/GRMF representation.

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

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

Parameter estimation and spatial prediction

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

where $p_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ 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 $\boldsymbol{x}$:

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

The inner core of the Integrated Nested Laplace method

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

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

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

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

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

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

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

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

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:

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

  where $\{(\boldsymbol{s}_j, w_j)\}$ 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

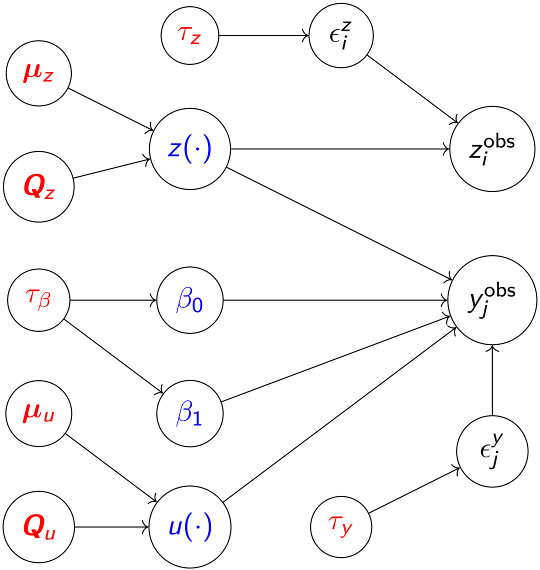
In plain INLA, the syntax mimics other R modelling packages such as `mgcv` and `lme4`, with formulae defining the predictor structure.

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



$$(\tau_z, \boldsymbol{\mu}_z, \boldsymbol{Q}_z) \sim p(\tau_z)p(\boldsymbol{\mu}_z)p(\boldsymbol{Q}_z)$$

$$z(\cdot)\text{-coefficients} \sim \text{N}(\boldsymbol{\mu}_z, \boldsymbol{Q}_z{}^{-1})$$

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

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

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

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

$$u(\cdot)\text{-coefficients} \sim \text{N}(\boldsymbol{\mu}_u, \boldsymbol{Q}_u{}^{-1})$$

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

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

$$(\boldsymbol{\mu}_u, \boldsymbol{Q}_u) \sim p(\boldsymbol{\mu}_u)p(\boldsymbol{Q}_u)$$

Basic joint models and model mis-specification handling

- Joint models allow uncertainty propagation between sub-models:

$$
\begin{aligned}
z_i^{\text{obs}} &= z(\boldsymbol{s}_i) + \epsilon_i^z \\
y_j^{\text{obs}} &= \beta_0 + \beta_1 z(\boldsymbol{s}_j) + u(\boldsymbol{s}_j) + \epsilon_j^y
\end{aligned}
$$

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

$$
\begin{aligned}
z_i^{\text{obs}} &= z(\boldsymbol{s}_i) + \epsilon_i^z \\
z^{\text{post}} &\sim \mathrm{N}(\boldsymbol{\mu}_{z|z^{\text{obs}}}, \boldsymbol{Q}_{z|z^{\text{obs}}}^{-1}) \quad \text{(convenient posterior approximation)} \\
y_j^{\text{obs}} &= \beta_0 + \beta_1 z^{\text{post}}(\boldsymbol{s}_j) + u(\boldsymbol{s}_j) + \epsilon_j^y
\end{aligned}
$$

Non-linear predictors

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

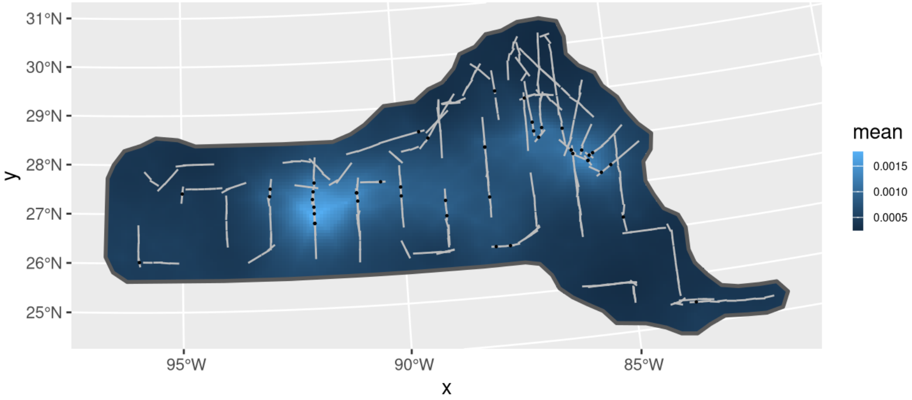
$$\lambda_{\text{apparent}}(\boldsymbol{s};\boldsymbol{u},\boldsymbol{v}) = \lambda(\boldsymbol{s};\boldsymbol{u})h(\boldsymbol{s};\boldsymbol{v}),$$

where $h(\boldsymbol{s};\boldsymbol{v})$ is the detection probability for a point located at $\boldsymbol{s}$, and $\boldsymbol{v}$ is a vector of parameters for the detection function.

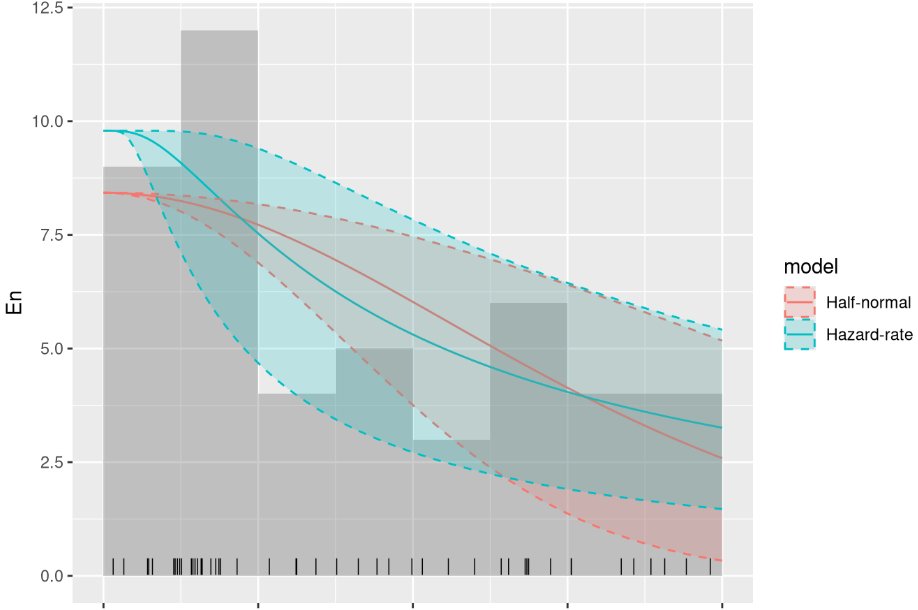
The `inlabru` package solves this by iterating the INLA method on a linearisation of the non-linear predictor

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

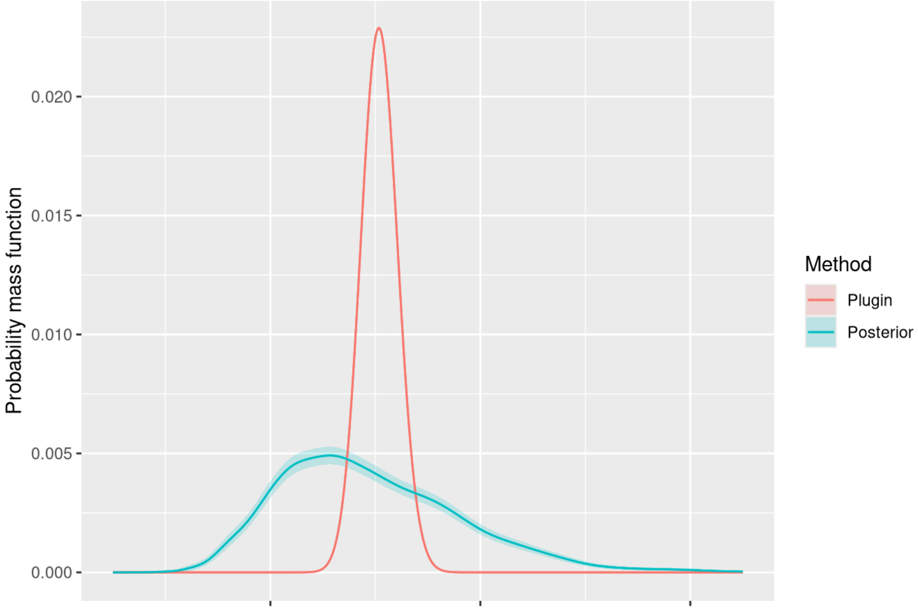
Dolphin group detection; estimated density field



Dolphin group detection; estimated detection probabilities



Dolphin group detection; estimated total count



Numerical challenges

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

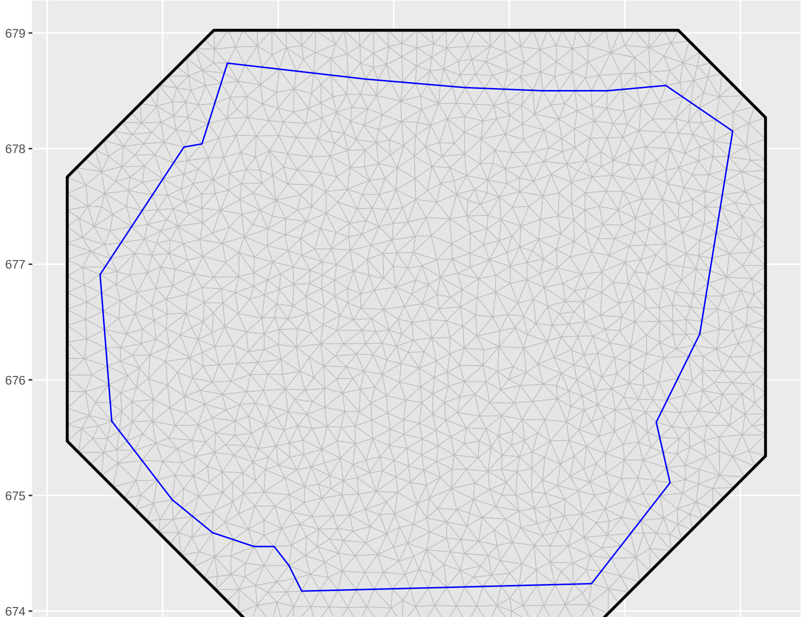
Modelling and computational challenges

- 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 ($\sim 10^{1}1$ 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)

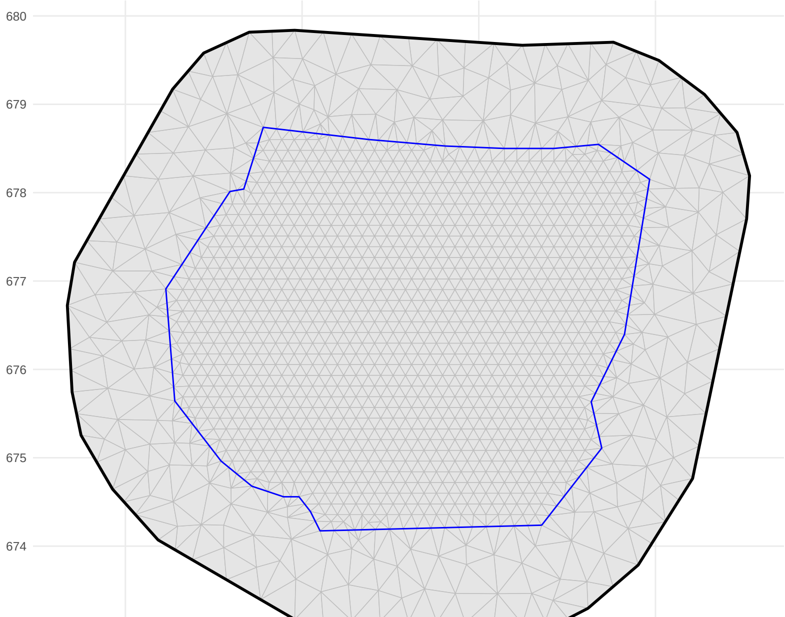
Partial inversion beyond Takahashi recursions

- 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.

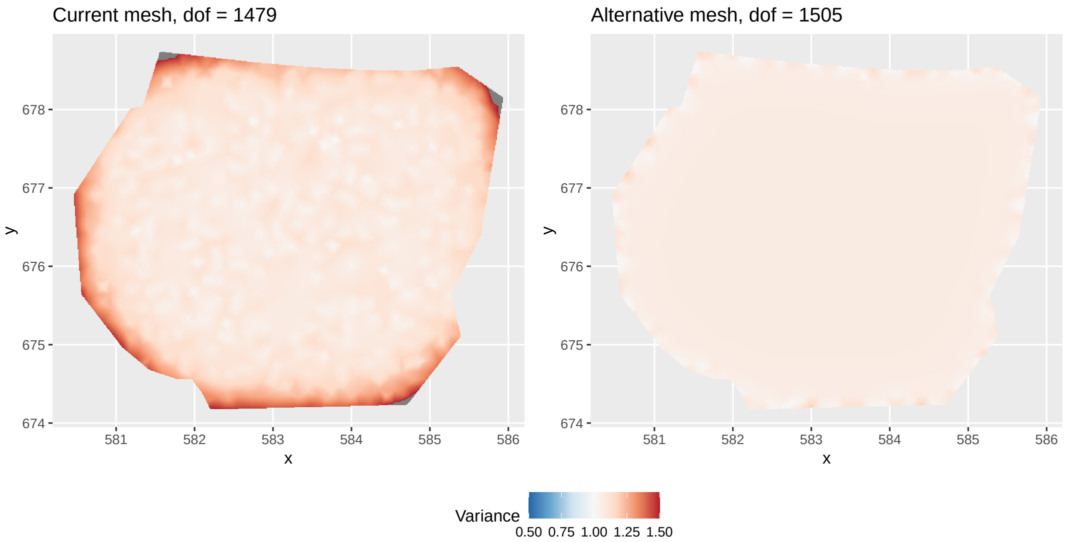
Challenge: explaining FEM meshes to non-PDE scientists



The methods are flexible, but some choices are generally better

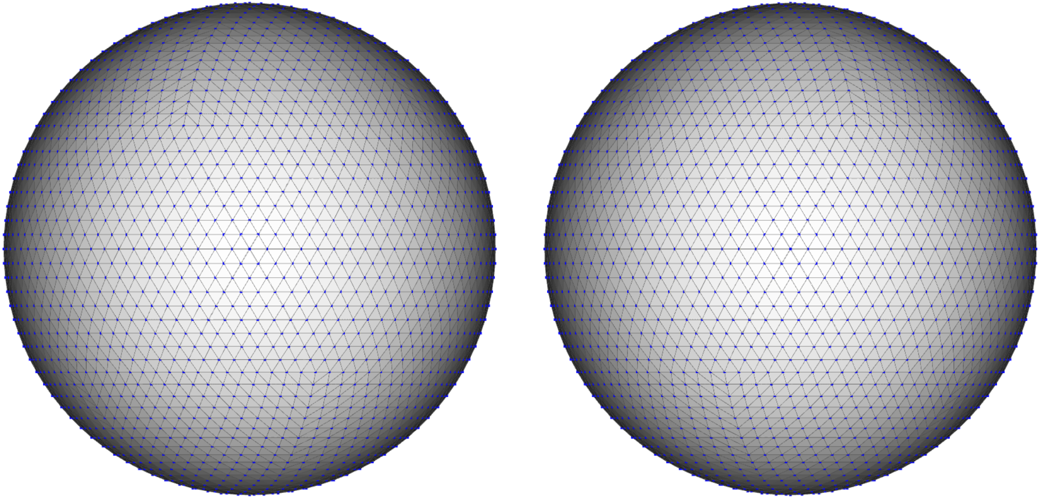


Variances

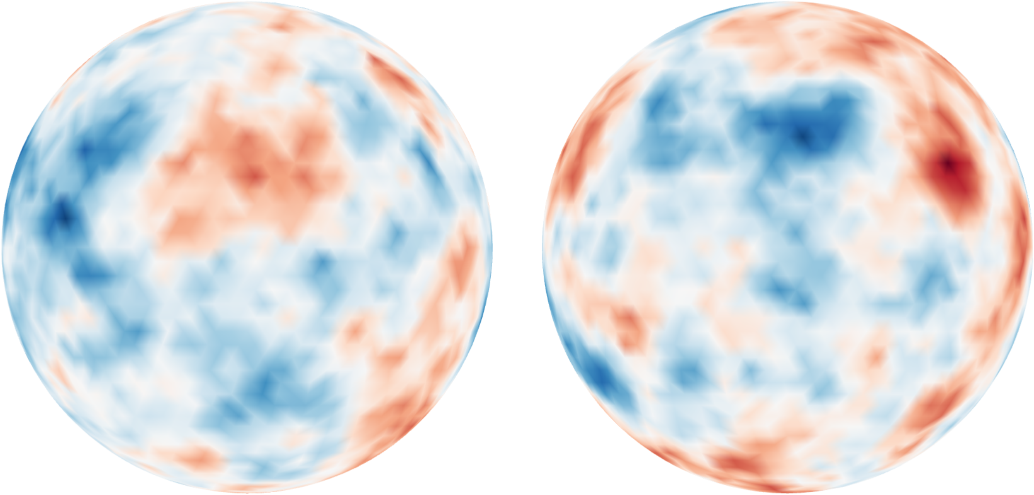


Current mesh, dof = 1479
Alternative mesh, dof = 1505
678
677
676
675
674
y
581
582
583
584
585
586
x
Variance
0.50
0.75
1.00
1.25
1.50

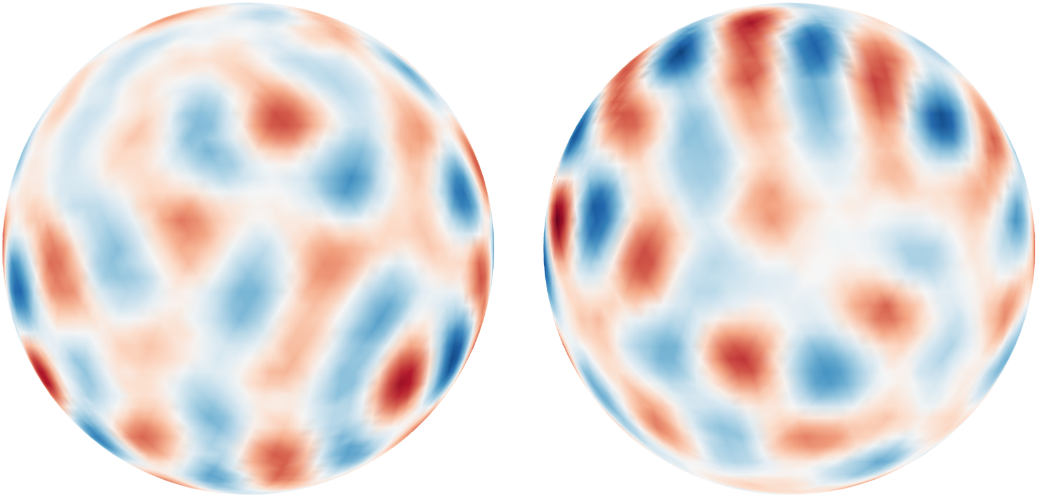
Nearly regular mesh on the unit sphere



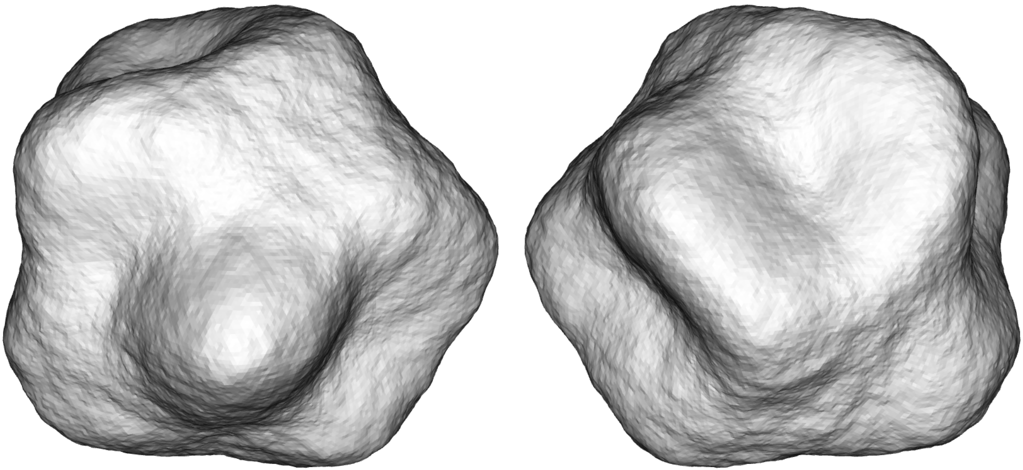
Whittle-Matérn field on the unit sphere



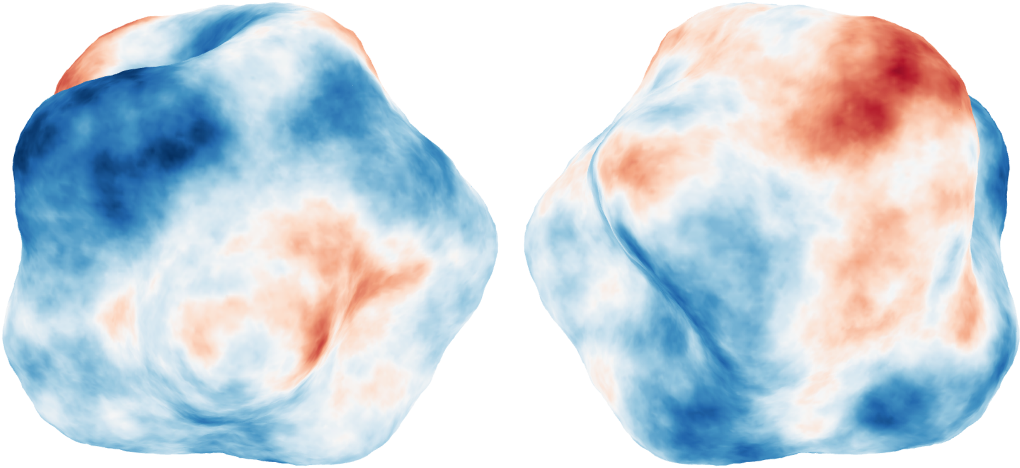
Oscillatory field on the unit sphere (modified Whittle SPDE)



‘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
- `R-INLA` documentation and examples: https://www.r-inla.org/
- `fmesher` / `inlabru` Mesh handling and model estimation: https://inlabru-org.github.io/fmesher/ and `.../inlabru/`
- `INLAspacetime` non-separable space-time: https://eliaskrainski.github.io/INLAspacetime/