Modelling, computing, and understanding uncertainty in spatial statistics

Finn Lindgren

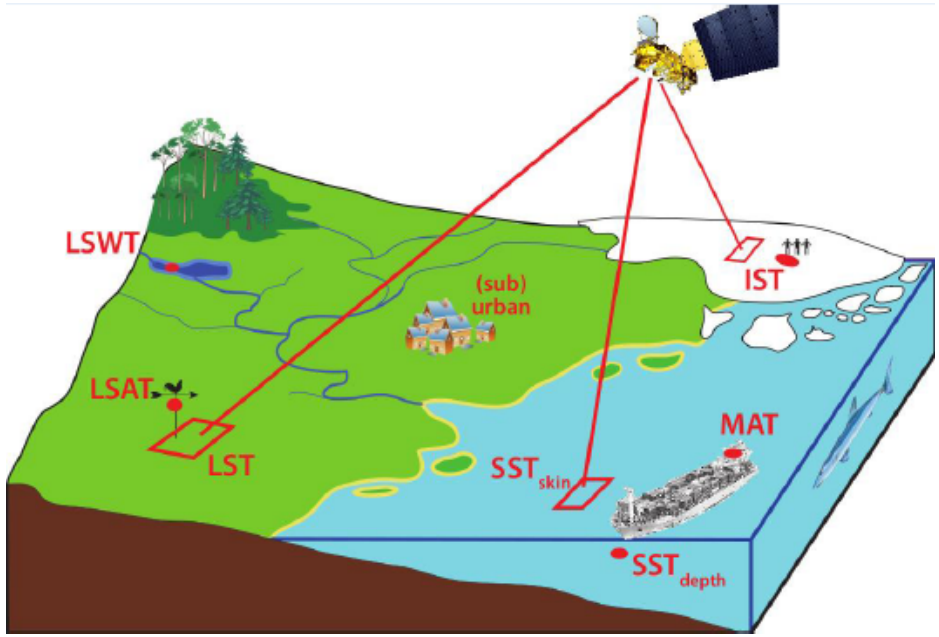
University of Edinburgh, 30 March 2016

EUSTACE (*EU Surface Temperatures for All Corners of Earth*)

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

EUSTACE

EUSTACE will give publicly available daily estimates of surface air temperature since 1850 across the globe for the first time by combining surface and satellite data using novel statistical techniques.



Nick A. Rayner, Renate Auchmann, Janette Bessembinder, Stefan Brönnimann, Yuri Brugnara, Laura Carrea, Darren Ghent, Elizabeth Good, Katie Herring, Jacob Høyer, John Kennedy, Albert Klein Tank, Finn Lindgren, Colin Morice, Chris Merchant, John Remedios, Ag Stephens and Rasmus Tonboe

contact: nick.rayner@metoffice.gov.uk

Royal Netherlands Meteorological Institute *Ministry of Infrastructure and the Environment*

Science & Technology Facilities Council

University of Leicester

Met Office Hadley Centre

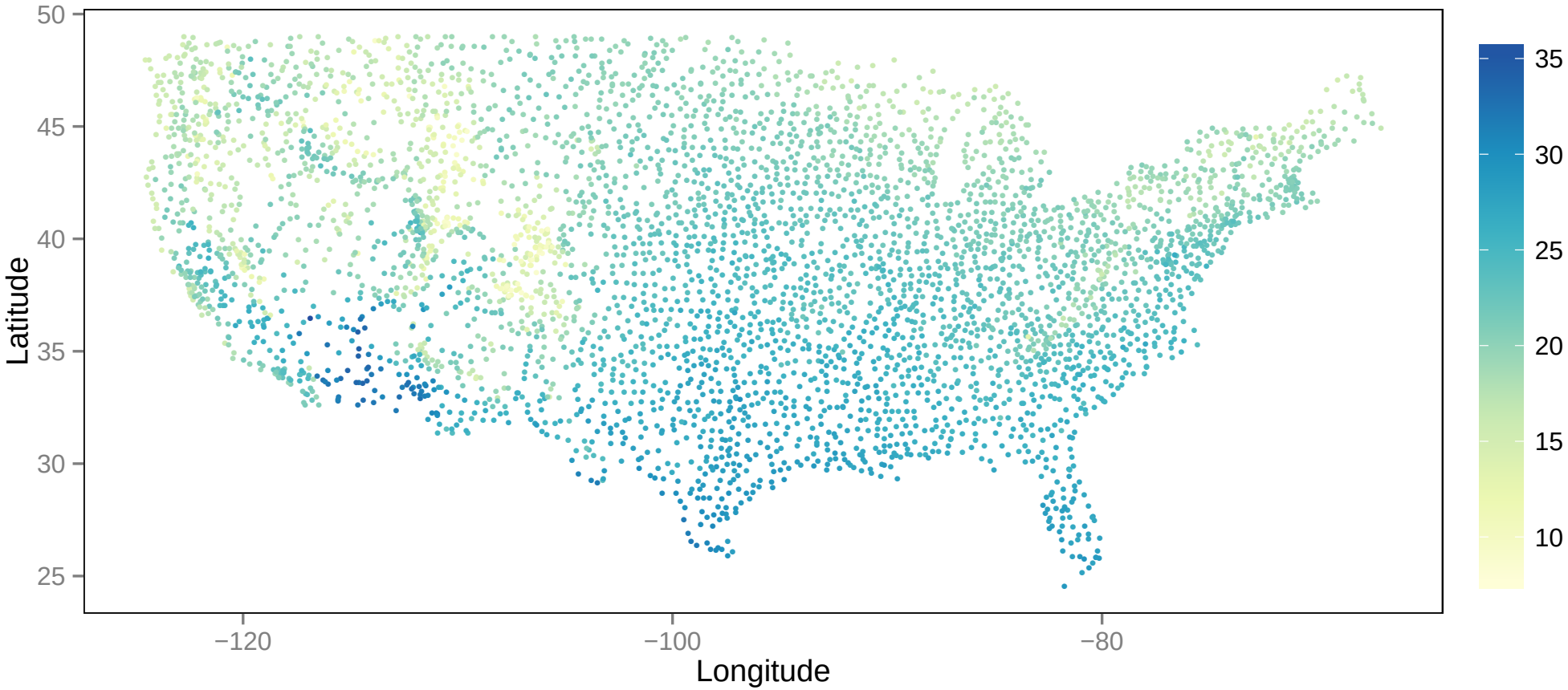
University of Reading

University of Bath

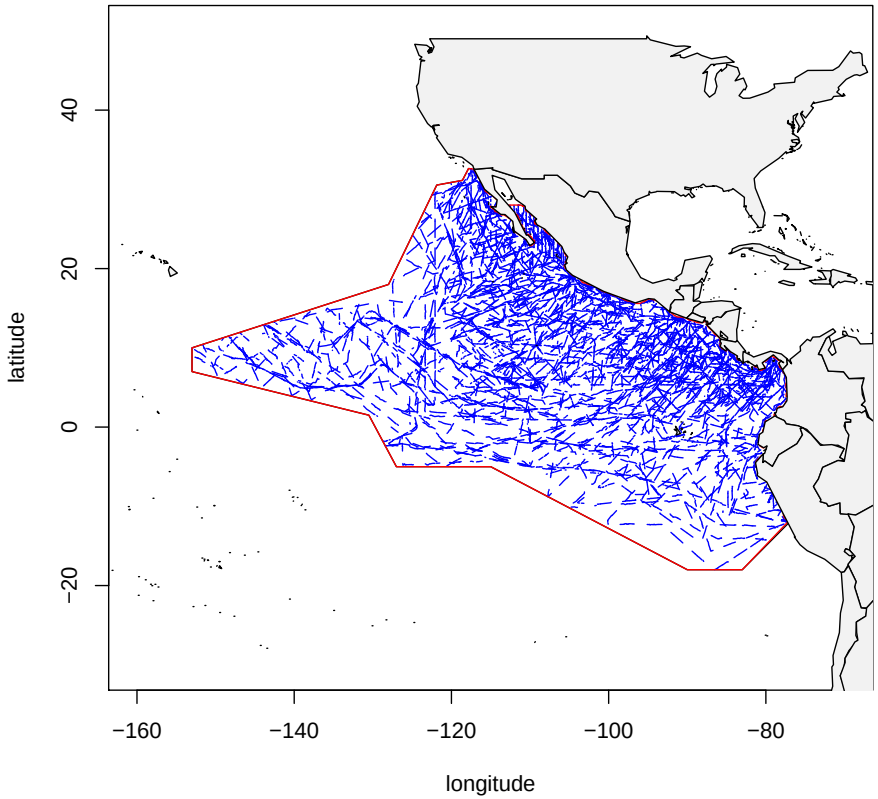
Universität Bern Oeschger Centre Climate Change Research

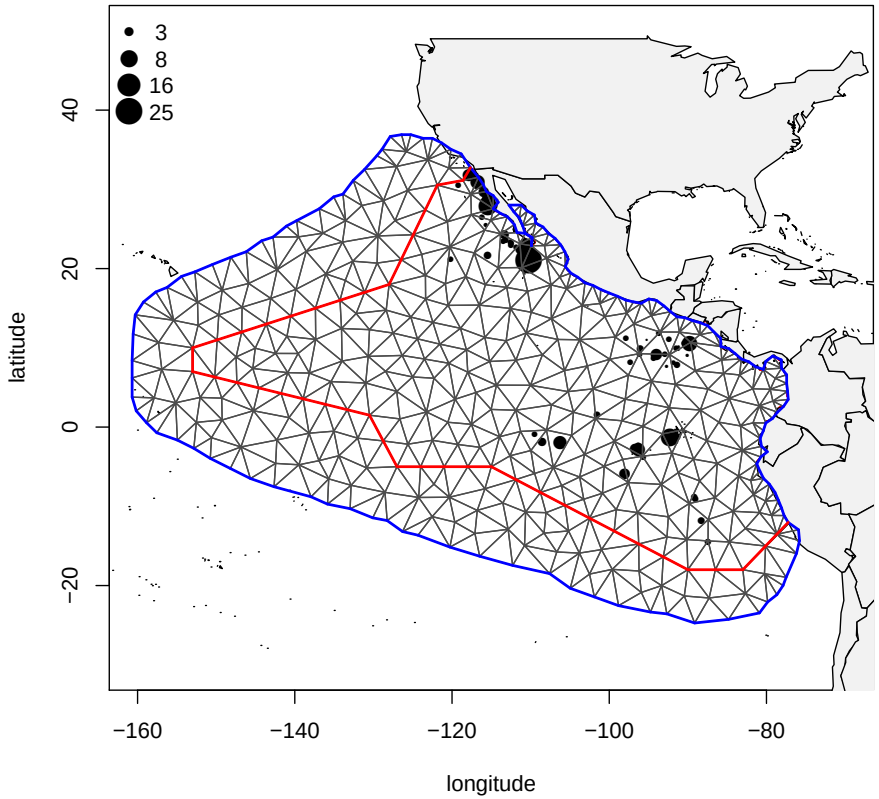
DMI

Mean summer temperature measurements over mainland USA for 1997



Animal abundance estimation for distance sampling





The probability of detecting a group of animals depends on the distance to the ship and the size of the group. The abundance varies across space and time.

Spatial fields, observations, an stochastic models

- Partially observed spatial functions (temperature) or objects related to *latent* spatial functions (animal intensity)
- Wanted: estimates of the true values at observed and unobserved locations
- Wanted: quantified uncertainty about those values
- Measurement errors and lack of observational coverage can be modeled using random variables

Spatial hierarchical model framework

- Observations $\boldsymbol{y} = \{y_i, i = 1, \ldots, n_y\}$
- Latent random field $x(\boldsymbol{s})$, $\boldsymbol{s} \in \Omega$
- Model parameters $\boldsymbol{\theta} = \{\theta_j, j = 1, \ldots, n_\theta\}$

Bayesian modelling and estimation

In a *Bayesian* setting the behaviour of $\boldsymbol{y}$, $\boldsymbol{x}$, and $\boldsymbol{\theta}$ is modelled via probability density functions:

- $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ (data likelihood)
- $p(\boldsymbol{x}|\boldsymbol{\theta})$ (latent random field prior)
- $p(\boldsymbol{\theta})$ (parameter prior)

Given observed $\boldsymbol{y}$, all statements about x and $\boldsymbol{\theta}$ are based on the *posteror densities*

$$p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{p(\boldsymbol{y})} \propto p(\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})$$

$$p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\theta}$$

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \int p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}$$

A *Gaussian random field* $x : D \mapsto \mathbb{R}$ is defined via

$$\begin{aligned}
\mathsf{E}(x(\boldsymbol{s})) &= m(\boldsymbol{s}), \\
\mathsf{Cov}(x(\boldsymbol{s}), x(\boldsymbol{s}')) &= K(\boldsymbol{s}, \boldsymbol{s}'), \\
\left[x(\boldsymbol{s}_i), i = 1, \ldots, n\right] \sim \mathsf{N}(\boldsymbol{m} &= \left[m(\boldsymbol{s}_i), i = 1, \ldots, n\right], \\
\boldsymbol{\Sigma} &= \left[K(\boldsymbol{s}_i, \boldsymbol{s}_j), i, j = 1, \ldots, n\right])
\end{aligned}$$

for all finite location sets $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, and $K(\cdot, \cdot)$ symmetric positive definite.

A *generalised Gaussian random field* $x : D \mapsto \mathbb{R}$ is defined via a random measure, $\langle f, x \rangle_D = x^*(f) : H_{\mathcal{R}}(D) \mapsto \mathbb{R}$,

$$\begin{aligned}
\mathsf{E}(\langle f, x \rangle_D) &= \langle f, m \rangle_D = \int_D f(\boldsymbol{s}) m(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s}, \\
\mathsf{Cov}(\langle f, x \rangle_D, \langle g, x \rangle_D) &= \langle f, \mathcal{R} g \rangle_D \equiv \iint_{D \times D} f(\boldsymbol{s}) K(\boldsymbol{s}, \boldsymbol{s}') g(\boldsymbol{s}') \, \mathrm{d}\boldsymbol{s} \, \mathrm{d}\boldsymbol{s}', \\
\langle f, x \rangle_D &\sim \mathsf{N}(\langle f, m \rangle_D, \langle f, \mathcal{R} f \rangle_D)
\end{aligned}$$

for all $f, g \in H_{\mathcal{R}}(D) \equiv \{f : D \mapsto \mathbb{R}; \langle f, \mathcal{R} f \rangle_D < \infty\}$. This allows for singular covariance kernels $K(\cdot, \cdot)$.

Covariance functions and stochastic PDEs

The Matérn covariance family on $\mathbb{R}^d$

$$\mathrm{Cov}(x(\mathbf{0}), x(\boldsymbol{s})) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|\boldsymbol{s}\|)^\nu K_\nu(\kappa \|\boldsymbol{s}\|)$$

Scale $\kappa > 0$, smoothness $\nu > 0$, variance $\sigma^2 > 0$

Whittle (1954, 1963): Matérn as SPDE solution

Matérn fields are the stationary solutions to the SPDE

$$\left(\kappa^2 - \nabla \cdot \nabla\right)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \alpha = \nu + d/2$$

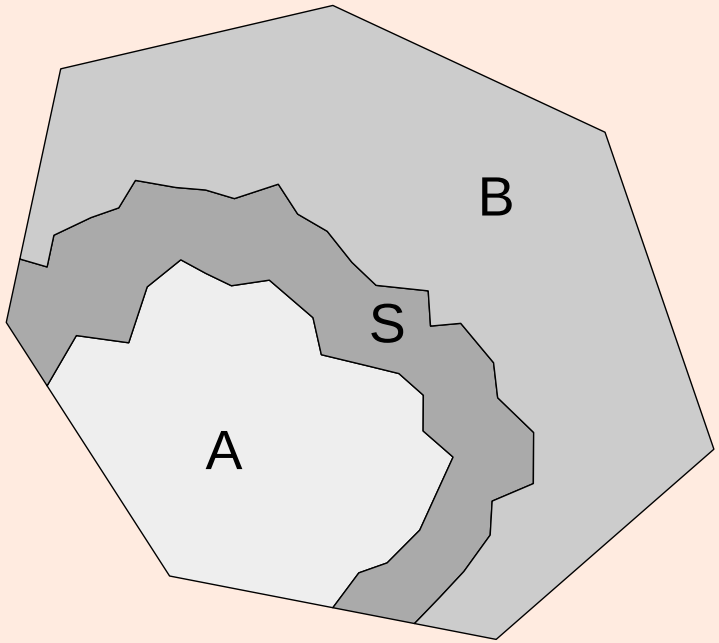
$\mathcal{W}(\cdot)$ white noise, $\nabla \cdot \nabla = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$, $\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)\kappa^{2\nu}(4\pi)^{d/2}}$

White noise has $K(\boldsymbol{s}, \boldsymbol{s}') = \delta(\boldsymbol{s} - \boldsymbol{s}')$.

Markov in space

Markov properties

S is a separating set for A and B: $x(A) \perp x(B) \mid x(S)$



Solutions to

$$\left(\kappa^2 - \nabla \cdot \nabla\right)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s})$$

are Markov when α is an integer.

(Generally, when the reciprocal of the spectral density is a polynomial, Rozanov, 1977)

Discrete representations ($\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$):

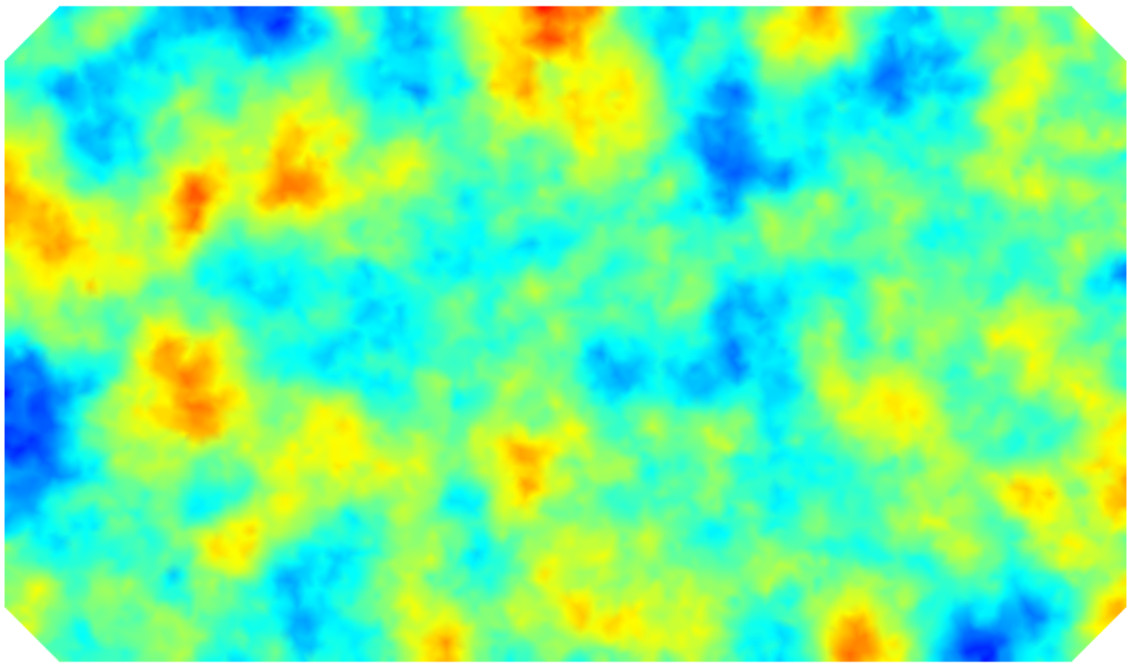
$$\boldsymbol{Q}_{AB} = \boldsymbol{0}$$
$$\boldsymbol{Q}_{A|S,B} = \boldsymbol{Q}_{AA}$$
$$\boldsymbol{\mu}_{A|S,B} = \boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AS}(\boldsymbol{u}_S - \boldsymbol{\mu}_S)$$

If we use local basis function expansions, we can exploit the continuous Markov property as sparse numerical matrix algebra.

GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

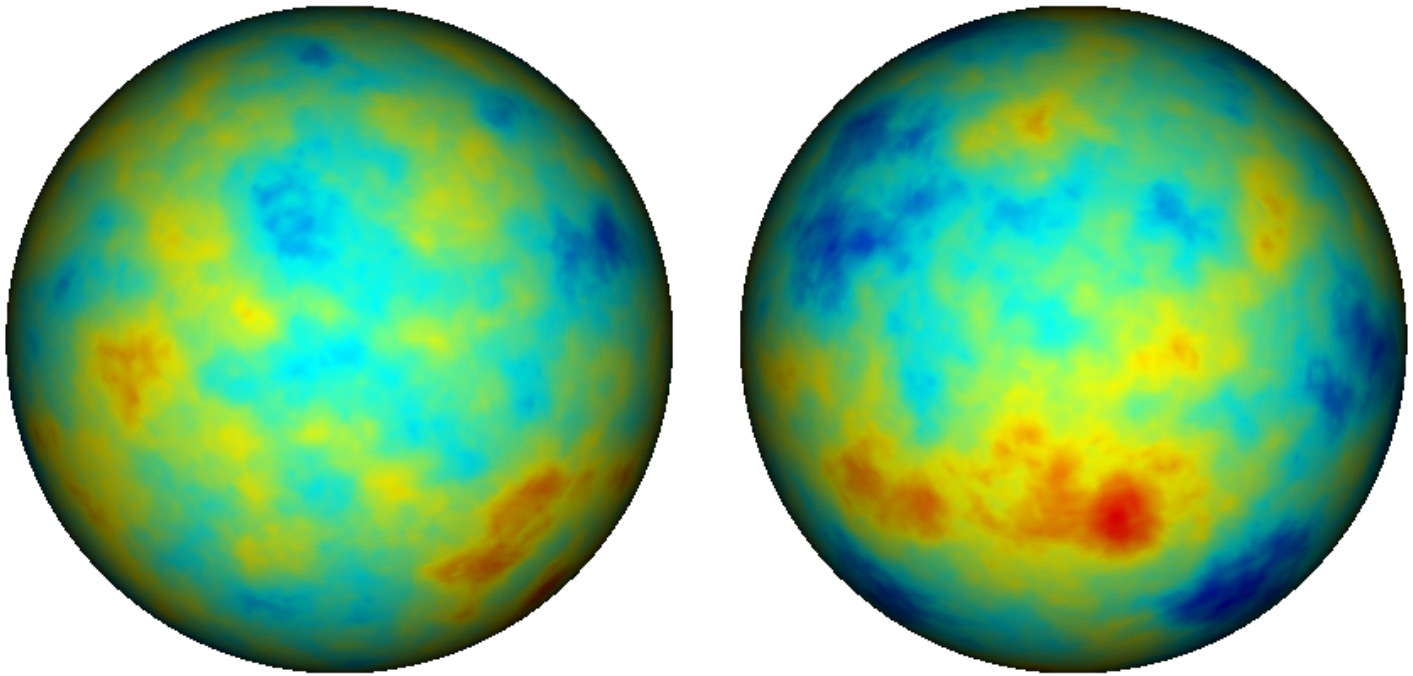
$$(\kappa^2 - \Delta)(\tau\, x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

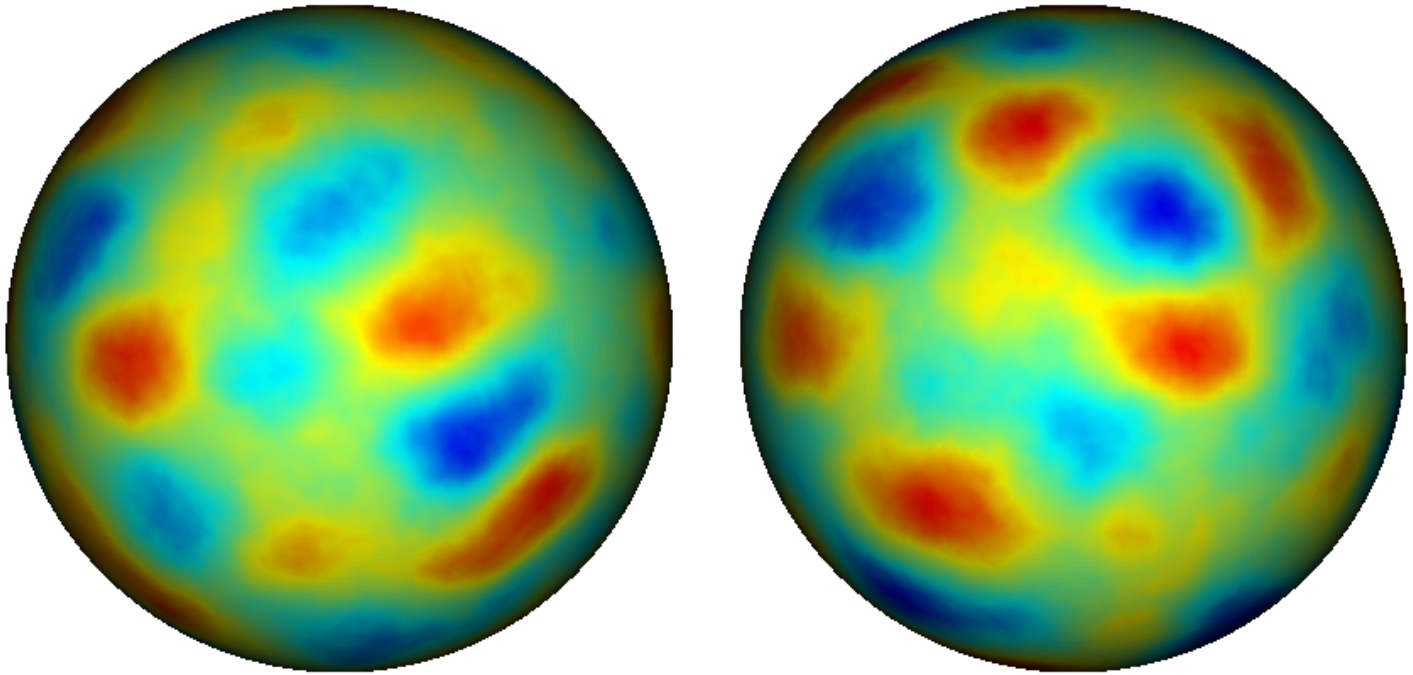
$$(\kappa^2 - \Delta)(\tau\, x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

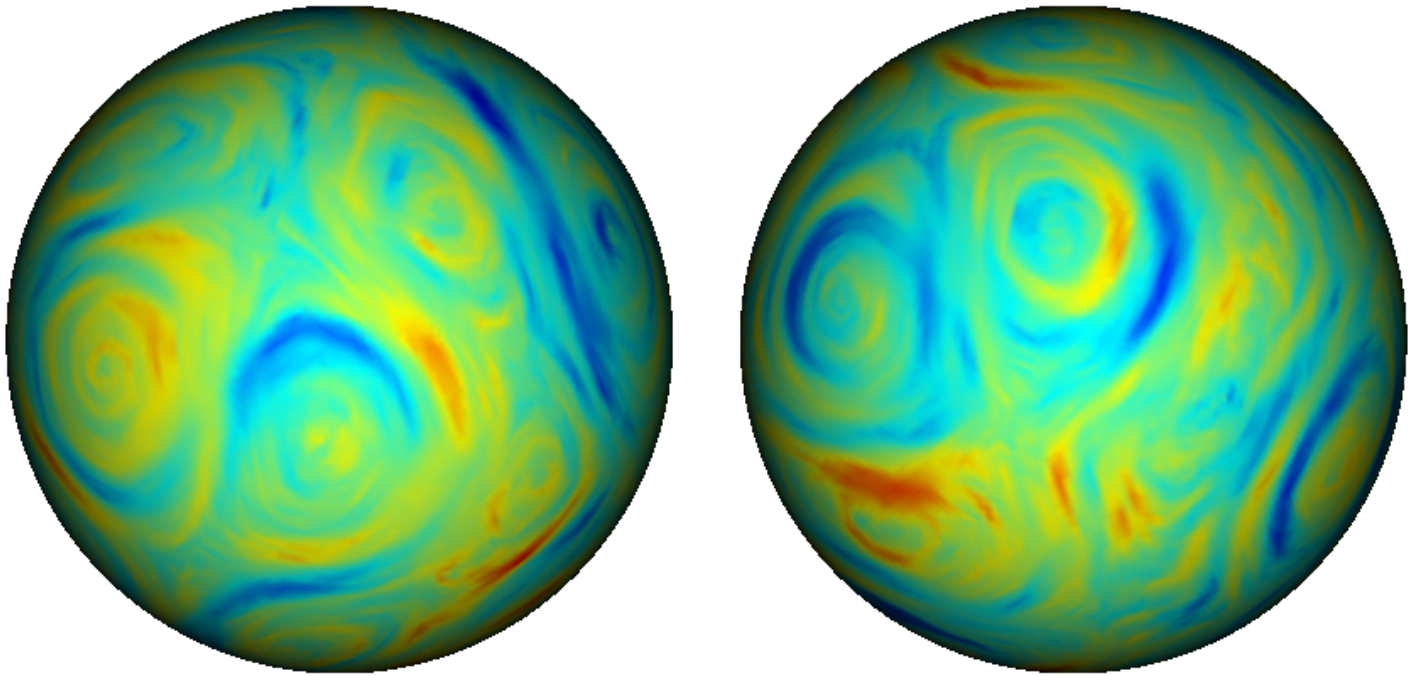
$$(\kappa^2 e^{i\pi\theta} - \Delta)(\tau\, x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

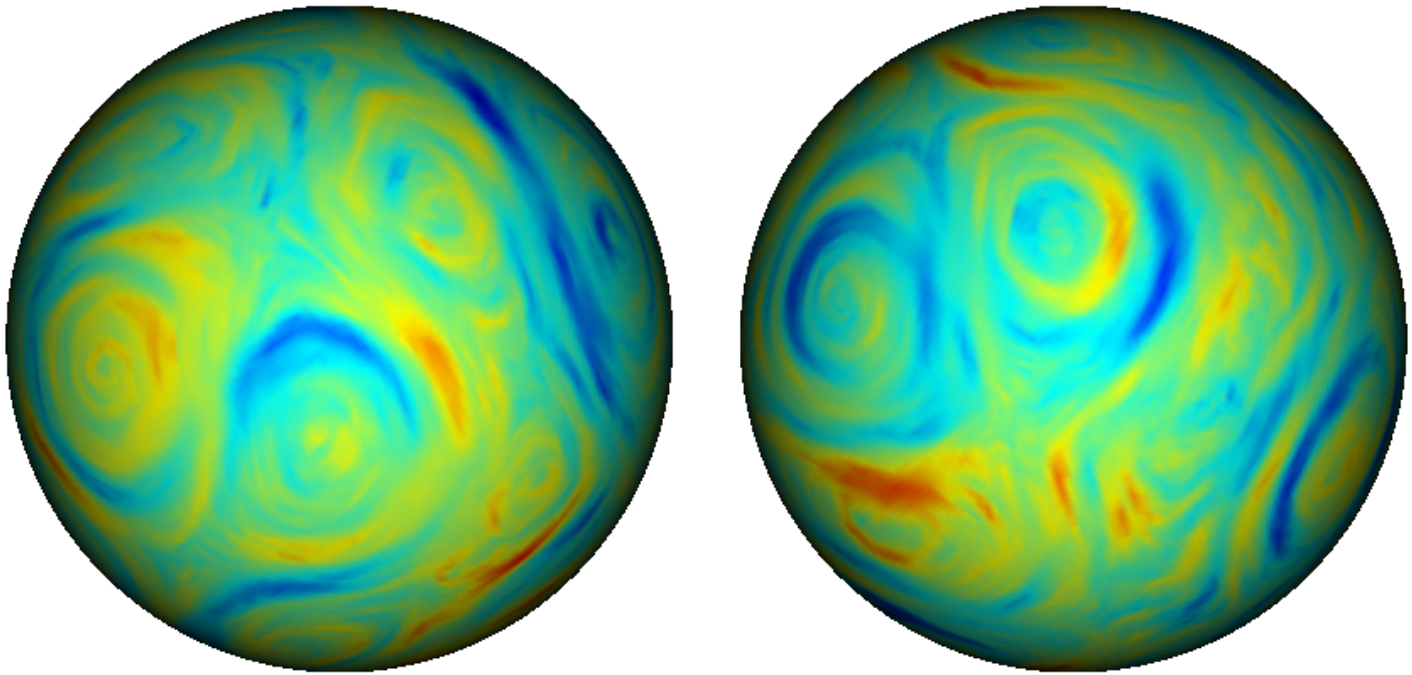
$$(\kappa_{\boldsymbol{s}}^2 + \nabla \cdot \boldsymbol{m}_{\boldsymbol{s}} - \nabla \cdot \boldsymbol{M}_{\boldsymbol{s}} \nabla)(\tau_{\boldsymbol{s}} x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

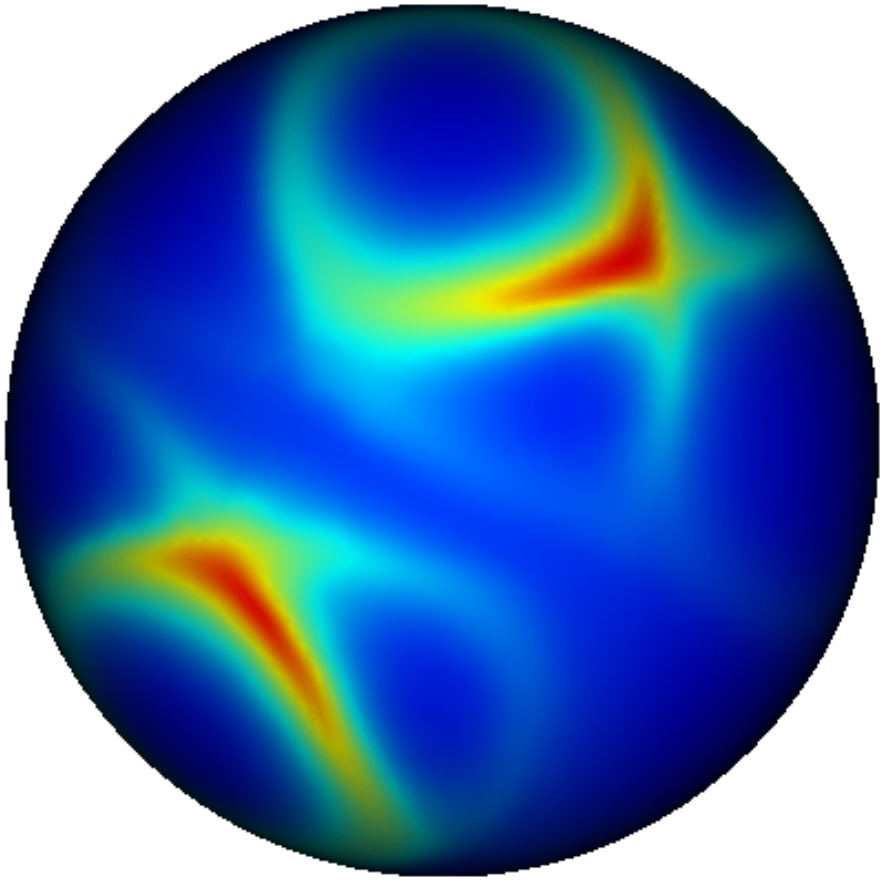
GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

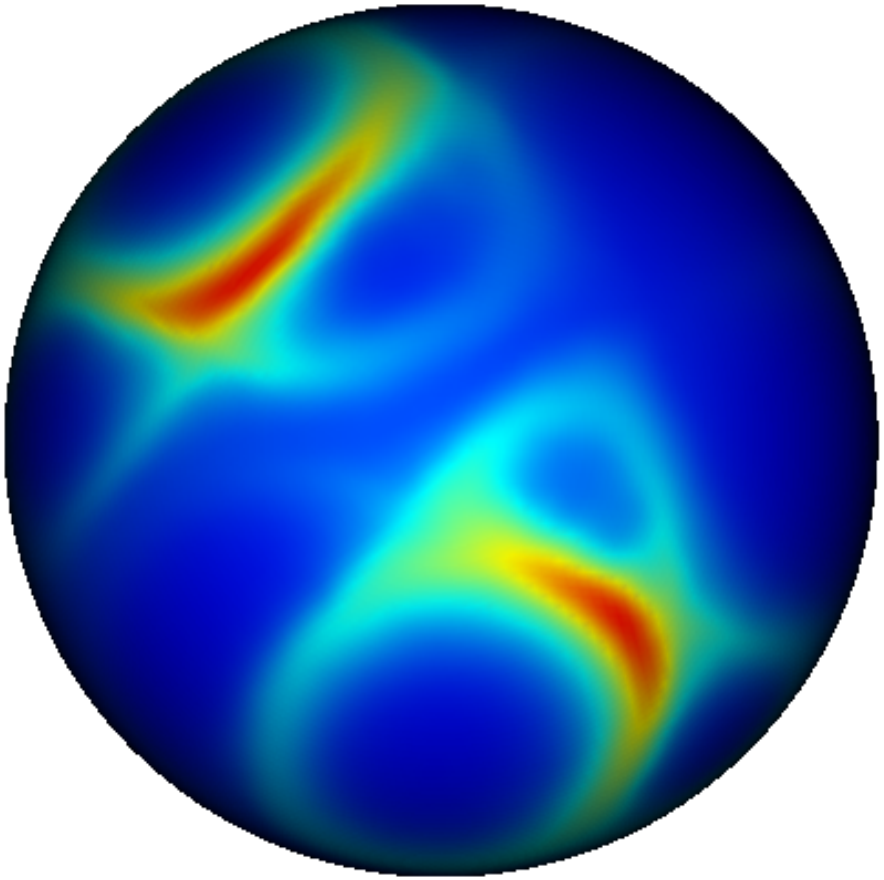
$$\left(\frac{\partial}{\partial t} + \kappa^2_{\boldsymbol{s},t} + \nabla \cdot \boldsymbol{m}_{\boldsymbol{s},t} - \nabla \cdot \boldsymbol{M}_{\boldsymbol{s},t} \nabla\right) (\tau_{\boldsymbol{s},t} x(\boldsymbol{s},t)) = \mathcal{E}(\boldsymbol{s},t), \quad (\boldsymbol{s},t) \in \Omega \times \mathbb{R}$$



Covariances for four reference points

$$\left(\frac{\partial}{\partial t} + \kappa^2_{\boldsymbol{s},t} + \nabla \cdot \boldsymbol{m}_{\boldsymbol{s},t} - \nabla \cdot \boldsymbol{M}_{\boldsymbol{s},t} \nabla\right) \left(\tau_{\boldsymbol{s},t} x(\boldsymbol{s},t)\right) = \mathcal{E}(\boldsymbol{s},t), \quad (\boldsymbol{s},t) \in \Omega \times \mathbb{R}$$





Hilbert space approximation

We want to construct finite dimensional approximations to the distribution of $x(\cdot)$, where

$$\left[\left\langle f_i, (\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} x(\cdot)\right\rangle_D, \, i = 1, \ldots, m\right] =_d \left[\left\langle f_i, \mathcal{W}(\cdot)\right\rangle_D, \, i = 1, \ldots, m\right]$$

for all finite collections of test functions $f_i \in H_{\mathcal{R}}(D)$.

A finite basis expansion

$$x(\boldsymbol{s}) = \sum_{j=1}^{n} \psi_j(\boldsymbol{s}) x_j$$

can only hope to achieve this for a subspace of size n.

Two main approaches:

- Galerkin: $\{f_i = \psi_i, \, i = 1, \ldots, n\}$
- Least squares: $\{f_i = (\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} \psi_i, \, i = 1, \ldots, n\}$

We use least squares for $\alpha = 1$, Galerkin for $\alpha = 2$, and a recursion for $\alpha \geq 3$.

Stochastic Green's first identity

On any sufficiently smooth manifold domain D,

$$\langle f, -\nabla \cdot \nabla g \rangle_D = \langle \nabla f, \nabla g \rangle_D - \langle f, \partial_n g \rangle_{\partial D}$$

holds, even if either ∇f or $-\nabla \cdot \nabla g$ are as generalised as white noise.

For now, we'll impose deterministic Neumann boundary conditions, informally $\partial_n x(\boldsymbol{s}) = 0$ for all $\boldsymbol{s} \in \partial D$. For $\alpha = 2$ and Galerkin,

$$\begin{aligned}\left[\left\langle \psi_i, (\kappa^2 - \nabla \cdot \nabla) \sum_j \psi_j x_j \right\rangle_D\right] &= \left[\sum_j \left\{\kappa^2 \langle \psi_i, \psi_j \rangle_D + \langle \nabla \psi_i, \nabla \psi_j \rangle_D\right\} x_j\right] \\ &= (\kappa^2 \boldsymbol{C} + \boldsymbol{G})\boldsymbol{x}\end{aligned}$$

The covariance for the RHS of the SPDE is

$$\left[\mathsf{Cov}(\langle \psi_i, \mathcal{W} \rangle_D, \langle \psi_j, \mathcal{W} \rangle_D\right] = \left[\langle \psi_i, \psi_j \rangle_D\right] = \boldsymbol{C}$$

by the definition of $\mathcal{W}$.

We seek $\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ such that $\mathsf{Var}\{(\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{x}\} = \boldsymbol{C}$:

$$(\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{\Sigma}(\kappa^2\boldsymbol{C} + \boldsymbol{G}) = \boldsymbol{C}$$

$$\boldsymbol{\Sigma} = (\kappa^2\boldsymbol{C} + \boldsymbol{G})^{-1}\boldsymbol{C}(\kappa^2\boldsymbol{C} + \boldsymbol{G})^{-1}$$

If ψ_i are piecewise linear on a triangulation of D, then $\boldsymbol{C}$ and $\boldsymbol{G}$ are both very sparse, and in addition, $\boldsymbol{C} = \mathrm{diag}(\langle \psi_i, 1 \rangle_D)$ is a valid approximation. Then, the *precision* matrix is also sparse,

$$\boldsymbol{Q} = (\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1}(\kappa^2\boldsymbol{C} + \boldsymbol{G})$$

and $\boldsymbol{x}$ is Markov on the adjacency graph given by the non-zero structure of $\boldsymbol{Q}$.

Least squares and Galerkin recursion gives precisions for all $\alpha = 1, 2, \ldots$:

- $\boldsymbol{Q}_1 = (\kappa^2\boldsymbol{C} + \boldsymbol{G})$
- $\boldsymbol{Q}_2 = (\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1}(\kappa^2\boldsymbol{C} + \boldsymbol{G}) = \kappa^4\boldsymbol{C} + 2\kappa^2\boldsymbol{G} + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}$
- $\boldsymbol{Q}_\alpha = (\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1}\boldsymbol{Q}_{\alpha-2}\boldsymbol{C}^{-1}(\kappa^2\boldsymbol{C} + \boldsymbol{G})$
- Any $\alpha \geq 0$: $\boldsymbol{Q}_\alpha = \boldsymbol{C}^{1/2}\left\{\boldsymbol{C}^{-1/2}(\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1/2}\right\}^\alpha \boldsymbol{C}^{1/2}$
  (non-sparse for non-integer α)

Basis function representations for Gaussian Matérn fields

Basis definitions

	Finite basis set ($k = 1, \ldots, n$)
Karhunen-Loève	$(\kappa^2 - \nabla \cdot \nabla)^{-\alpha} e_{\kappa,k}(\boldsymbol{s}) = \lambda_{\kappa,k} e_{\kappa,k}(\boldsymbol{s})$
Fourier	$-\nabla \cdot \nabla e_k(\boldsymbol{s}) = \lambda_k e_k(\boldsymbol{s})$
Convolution	$(\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} g_\kappa(\boldsymbol{s}) = \delta(\boldsymbol{s})$
General	$\psi_k(\boldsymbol{s})$

Field representations

	Field $x(\boldsymbol{s})$	Weights
Karhunen-Loève	$\propto \sum_k e_{\kappa,k}(\boldsymbol{s}) z_k$	$z_k \sim \mathsf{N}(0, \lambda_{\kappa,k})$
Fourier	$\propto \sum_k e_k(\boldsymbol{s}) z_k$	$z_k \sim \mathsf{N}(0, (\kappa^2 + \lambda_k)^{-\alpha})$
Convolution	$\propto \sum_k g_\kappa(\boldsymbol{s} - \boldsymbol{s}_k) z_k$	$z_k \sim \mathsf{N}(0, \lvert\mathsf{cell}_k\rvert)$
General	$\propto \sum_k \psi_k(\boldsymbol{s}) x_k$	$\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{Q}_\kappa^{-1})$

Note: Harmonic basis functions (as in the Fourier approach) give a diagonal $\boldsymbol{Q}_\kappa$, but lead to dense posterior precision matrices.

Hierarchical models

Continuous Markovian spatial models (Lindgren et al, 2011)

Local basis: $x(\boldsymbol{s}) = \sum_k \psi_k(\boldsymbol{s}) x_k$, (compact, piecewise linear)

Basis weights: $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$, sparse $\boldsymbol{Q}$ based on an SPDE

Special case: $(\kappa^2 - \nabla \cdot \nabla) x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$

Precision: $\boldsymbol{Q} = \kappa^4 \boldsymbol{C} + 2\kappa^2 \boldsymbol{G} + \boldsymbol{G}_2$ $\quad (\kappa^4 + 2\kappa^2 |\boldsymbol{\omega}|^2 + |\boldsymbol{\omega}|^4)$

Conditional distribution in a jointly Gaussian model

$$\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}), \quad \boldsymbol{y}|\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{Q}_{y|x}^{-1}) \qquad (A_{ij} = \psi_j(\boldsymbol{s}_i))$$

$$\boldsymbol{x}|\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{\mu}_{x|y}, \boldsymbol{Q}_{x|y}^{-1})$$

$$\boldsymbol{Q}_{x|y} = \boldsymbol{Q}_x + \boldsymbol{A}^T \boldsymbol{Q}_{y|x} \boldsymbol{A} \quad (\sim\text{"Sparse iff } \psi_k \text{ have compact support"})$$

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{Q}_{x|y}^{-1} \boldsymbol{A}^T \boldsymbol{Q}_{y|x} (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_x)$$

The computational GMRF work-horse

Cholesky decomposition (Cholesky, 1924)

$$\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^\top, \quad \boldsymbol{L} \text{ lower triangular } (\sim \mathcal{O}(n^{(d+1)/2}) \text{ for } d = 1, 2, 3)$$

$$\boldsymbol{Q}^{-1}\boldsymbol{x} = \boldsymbol{L}^{-\top}\boldsymbol{L}^{-1}\boldsymbol{x}, \quad \text{via forward/backward substitution}$$

$$\log\det \boldsymbol{Q} = 2\log\det \boldsymbol{L} = 2\sum_i \log L_{ii}$$

André-Louis Cholesky (1875–1918)

"He invented, for the solution of the condition equations in the method of least squares, a very ingenious computational procedure which immediately proved extremely useful, and which most assuredly would have great benefits for all geodesists, if it were published some day." (Euology by Commandant Benoit, 1922)

Laplace approximations for non-Gaussian observations

Quadratic posterior log-likelihood approximation

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathsf{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}), \quad \boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta} \sim p(\boldsymbol{y} \mid \boldsymbol{x}\boldsymbol{\theta})$$

$$p_G(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \sim \mathsf{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1})$$

$$\mathbf{0} = \nabla_{\boldsymbol{x}} \left\{ \ln p(\boldsymbol{x} \mid \boldsymbol{\theta}) + \ln p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \right\} \big|_{\boldsymbol{x}=\widetilde{\boldsymbol{\mu}}}$$

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_x - \nabla_{\boldsymbol{x}}^2 \ln p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \big|_{\boldsymbol{x}=\widetilde{\boldsymbol{\mu}}}$$

Direct Bayesian inference with INLA (`r-inla.org`)

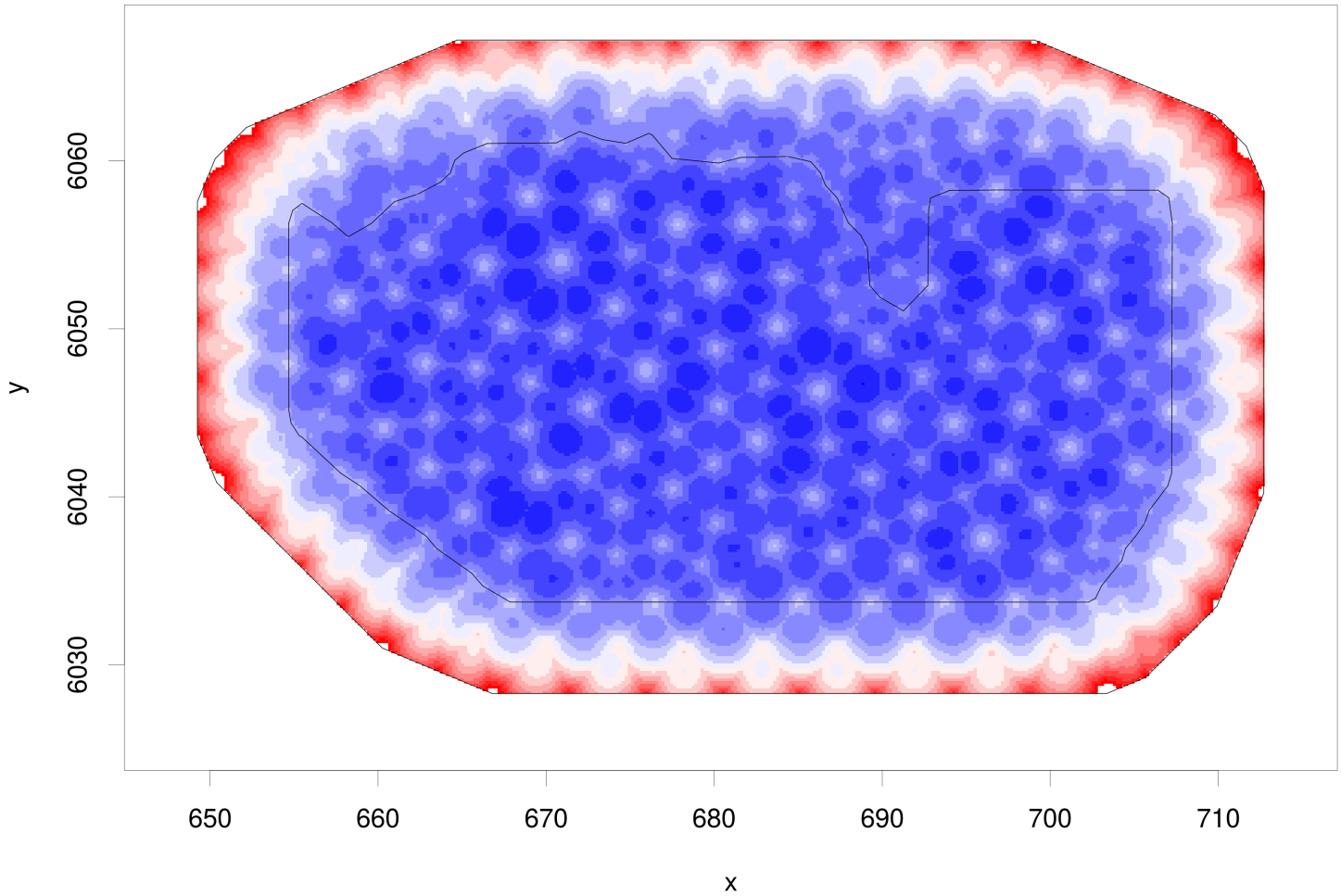
$$\widetilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \left. \frac{p(\boldsymbol{\theta}) p(\boldsymbol{x} \mid \boldsymbol{\theta}) p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta})}{p_G(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta})} \right|_{\boldsymbol{x}=\widetilde{\boldsymbol{\mu}}}$$

$$\widetilde{p}(x_i \mid \boldsymbol{y}) \propto \int p_{GG}(x_i \mid \boldsymbol{y}, \boldsymbol{\theta}) \widetilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y}) \, d\boldsymbol{\theta}$$

The main practical limiting factors for the INLA method are the number of latent variables and the number model parameters.

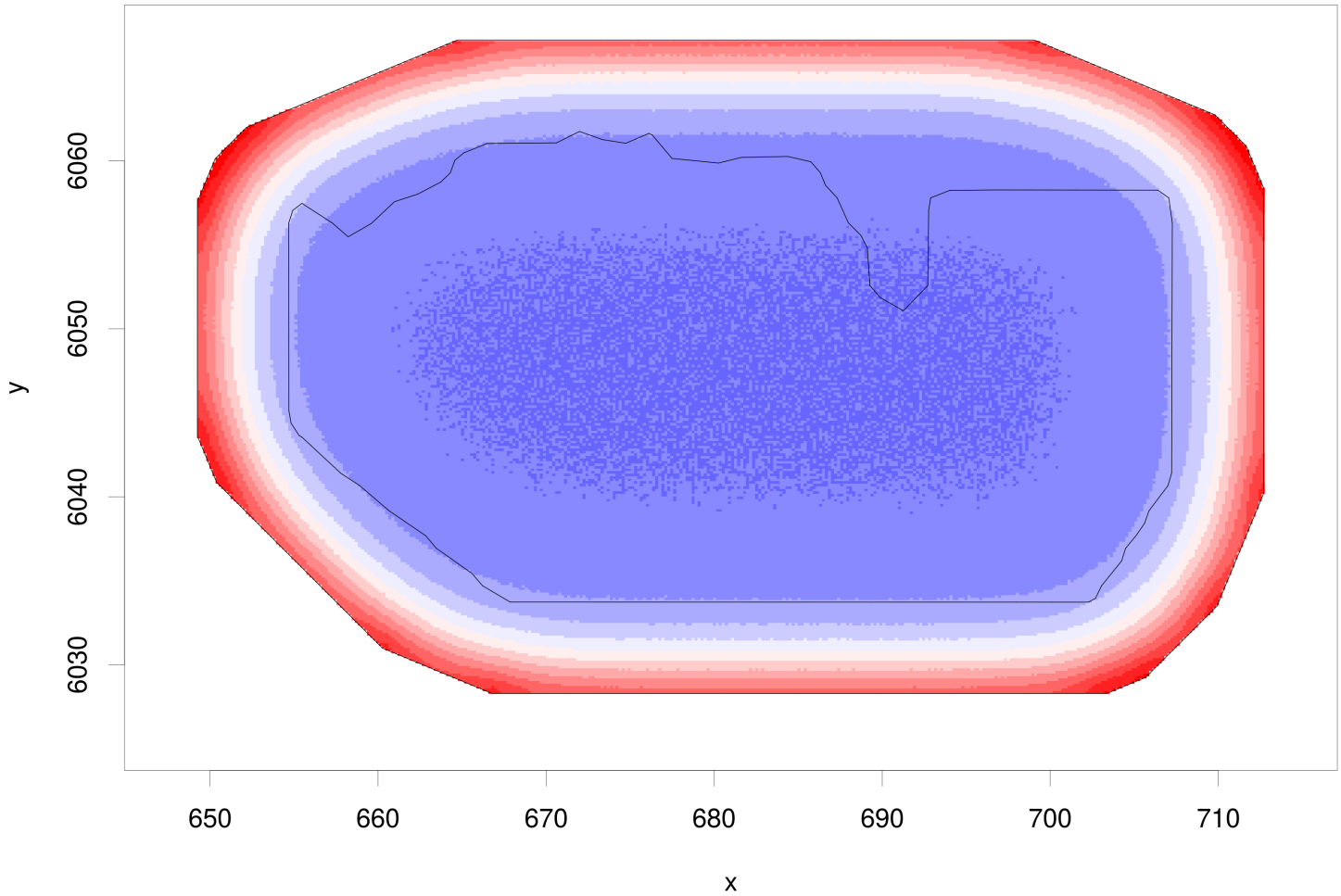
A coarse mesh: boundary and discretisation effects

The standard deviation at each location depends on the geometry



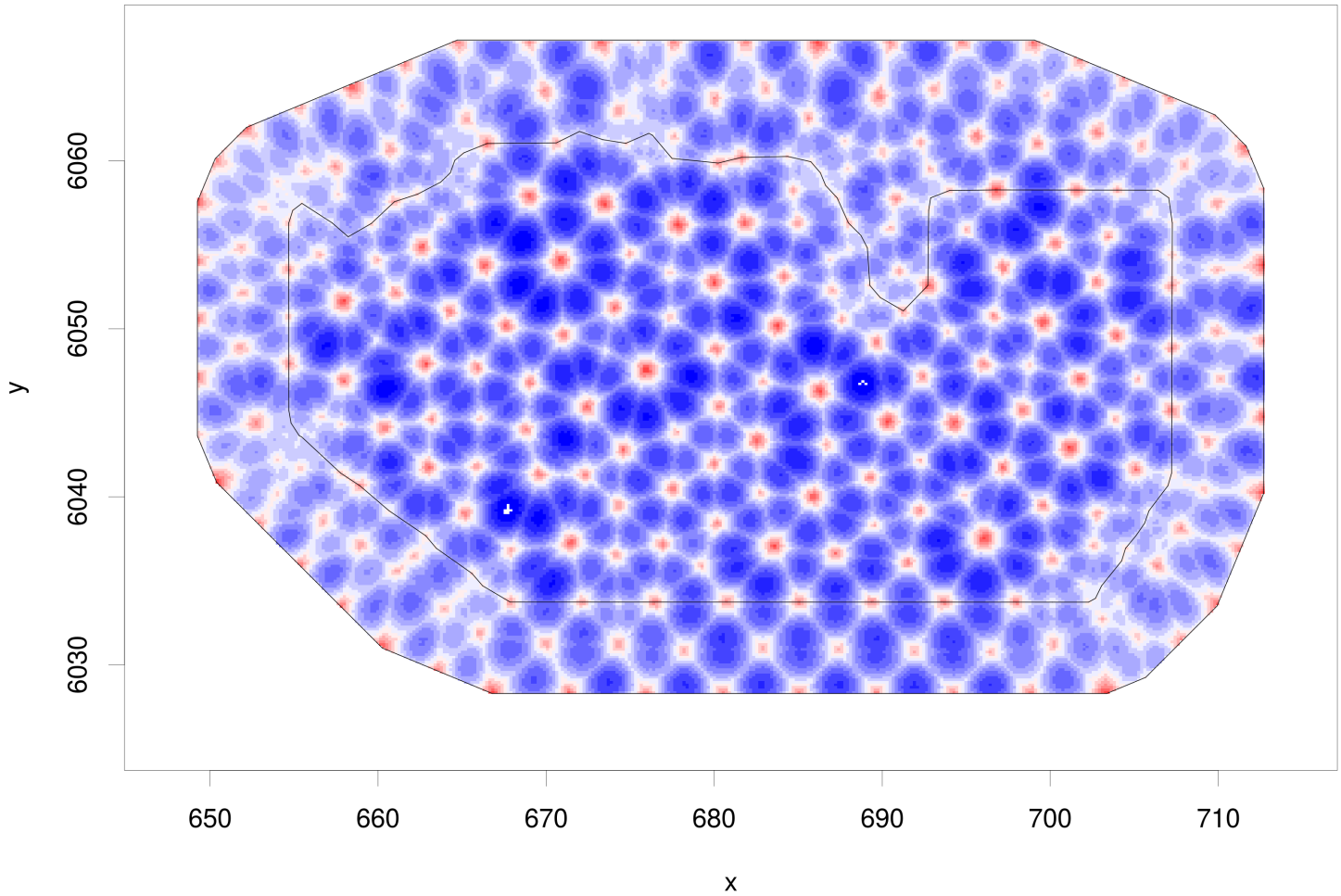
A coarse mesh: boundary and discretisation effects

The standard deviation at each location depends on the geometry



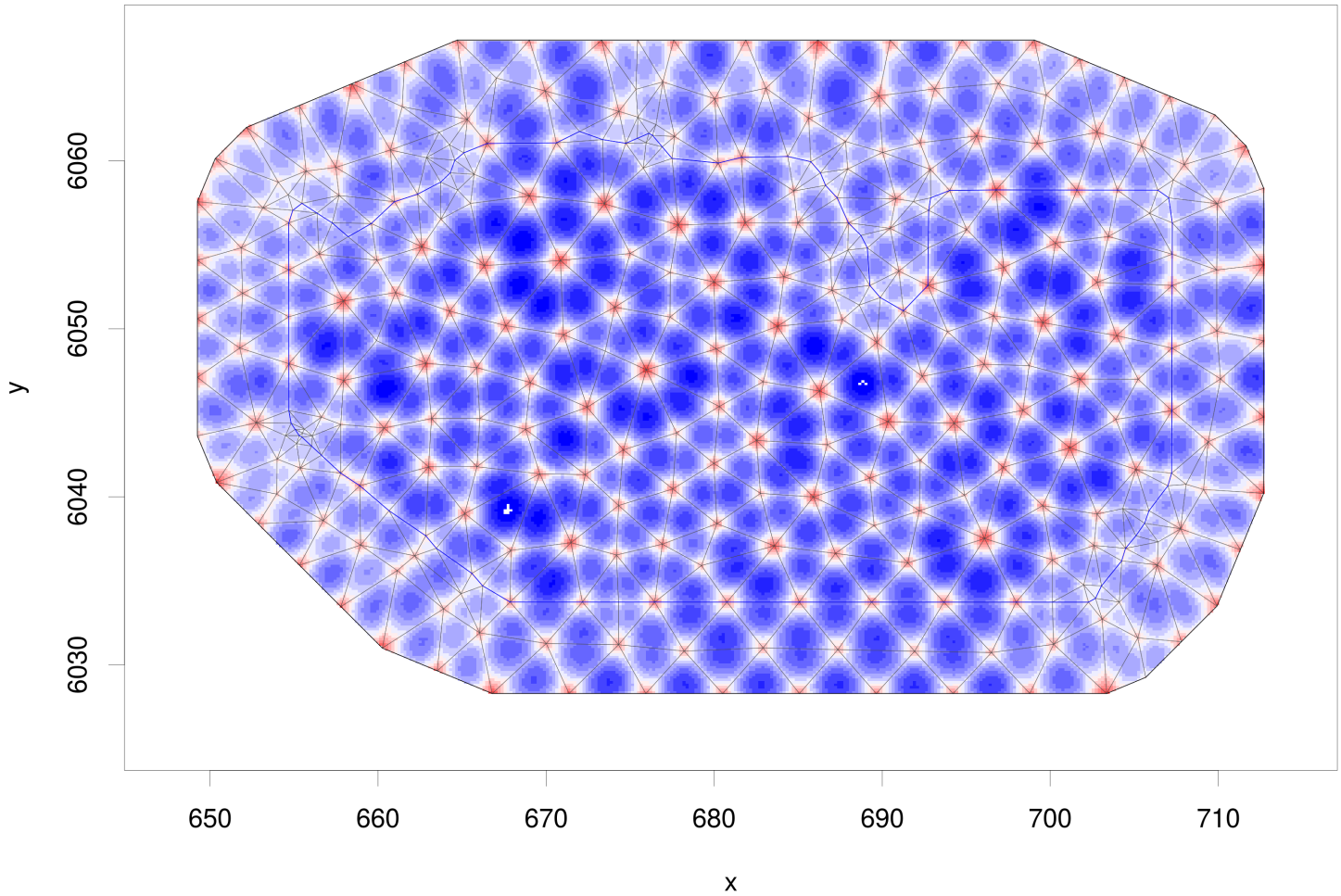
A coarse mesh: boundary and discretisation effects

The standard deviation at each location depends on the geometry



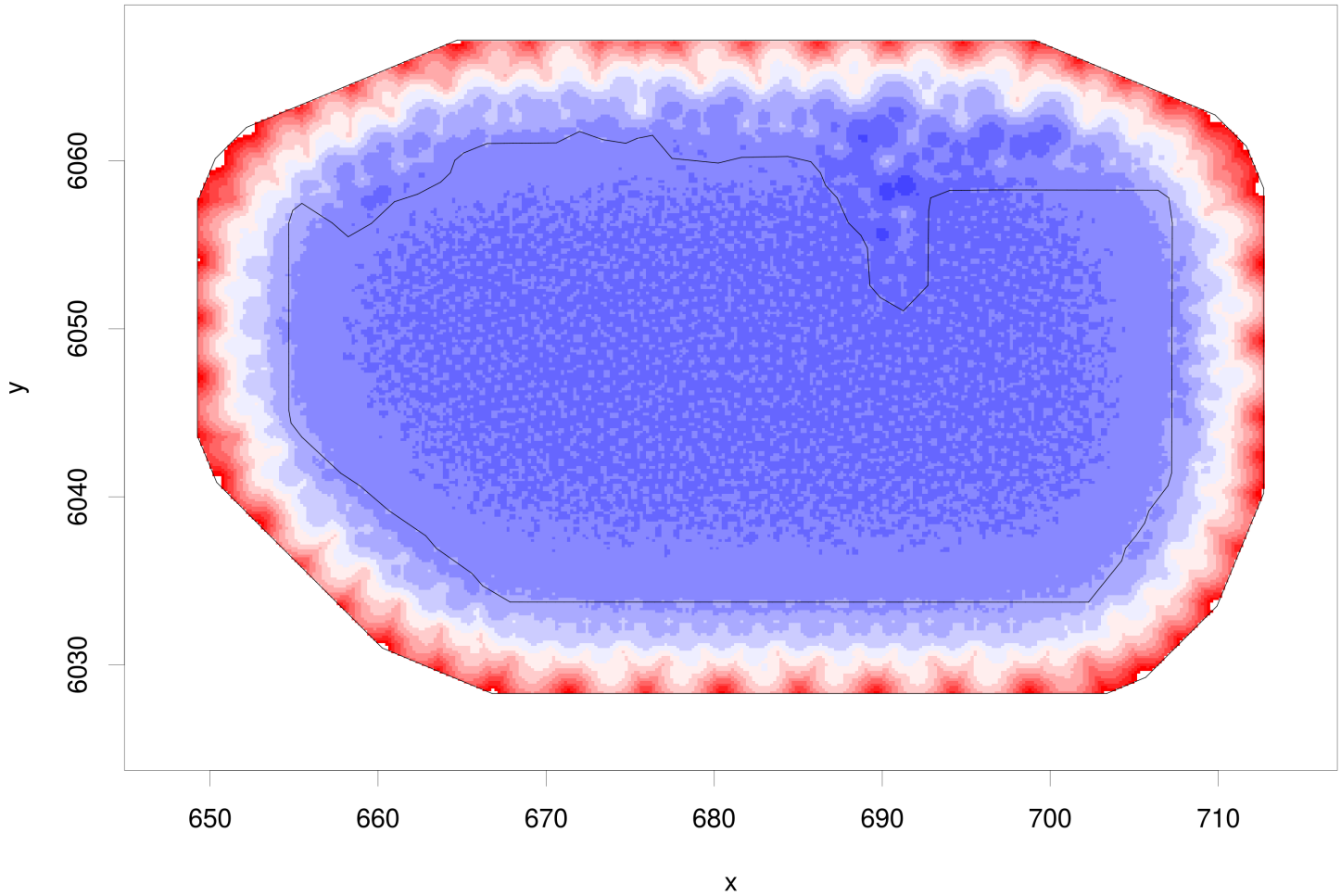
A coarse mesh: boundary and discretisation effects

The standard deviation at each location depends on the geometry



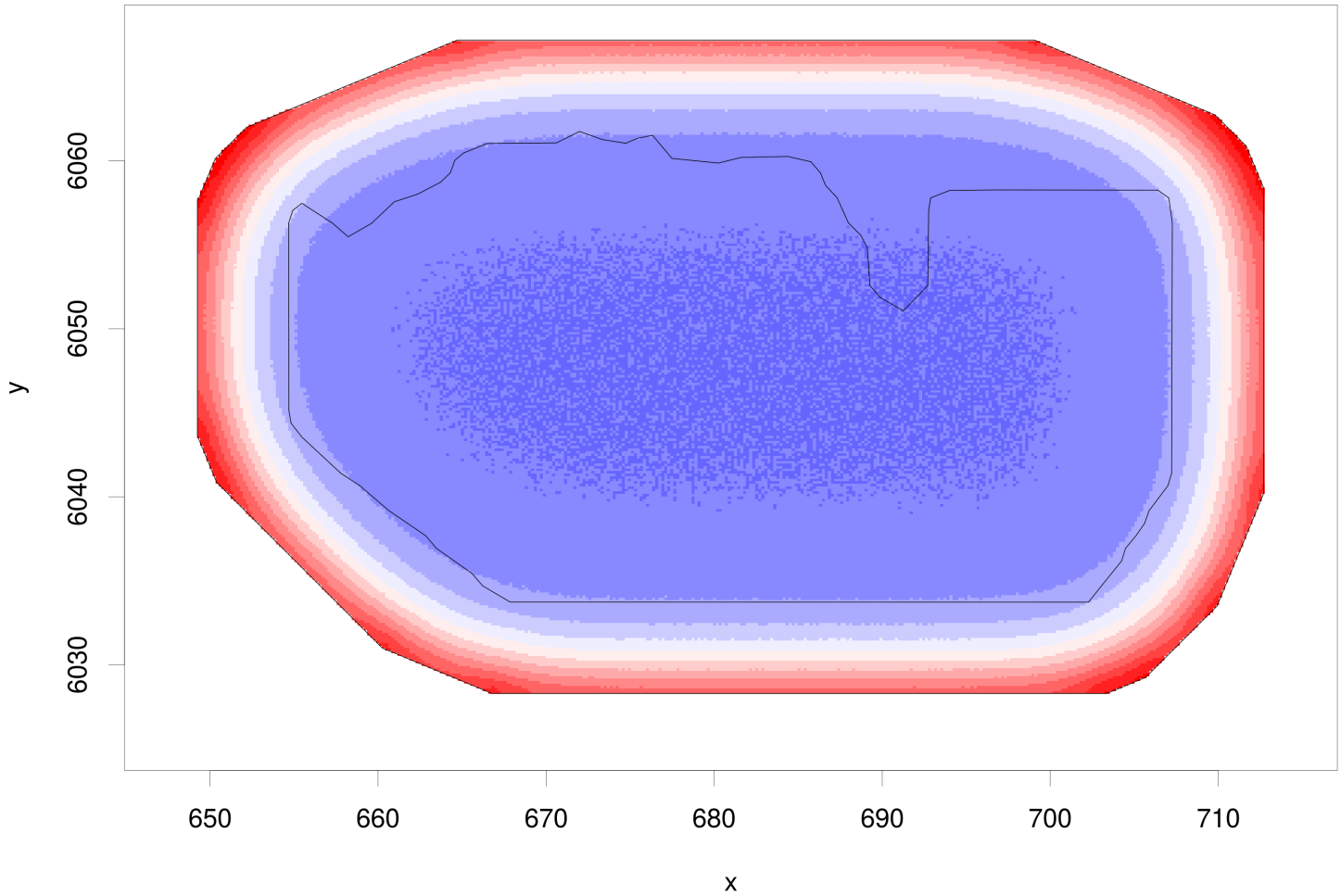
A refined mesh

Refining the mesh in the domain of interest



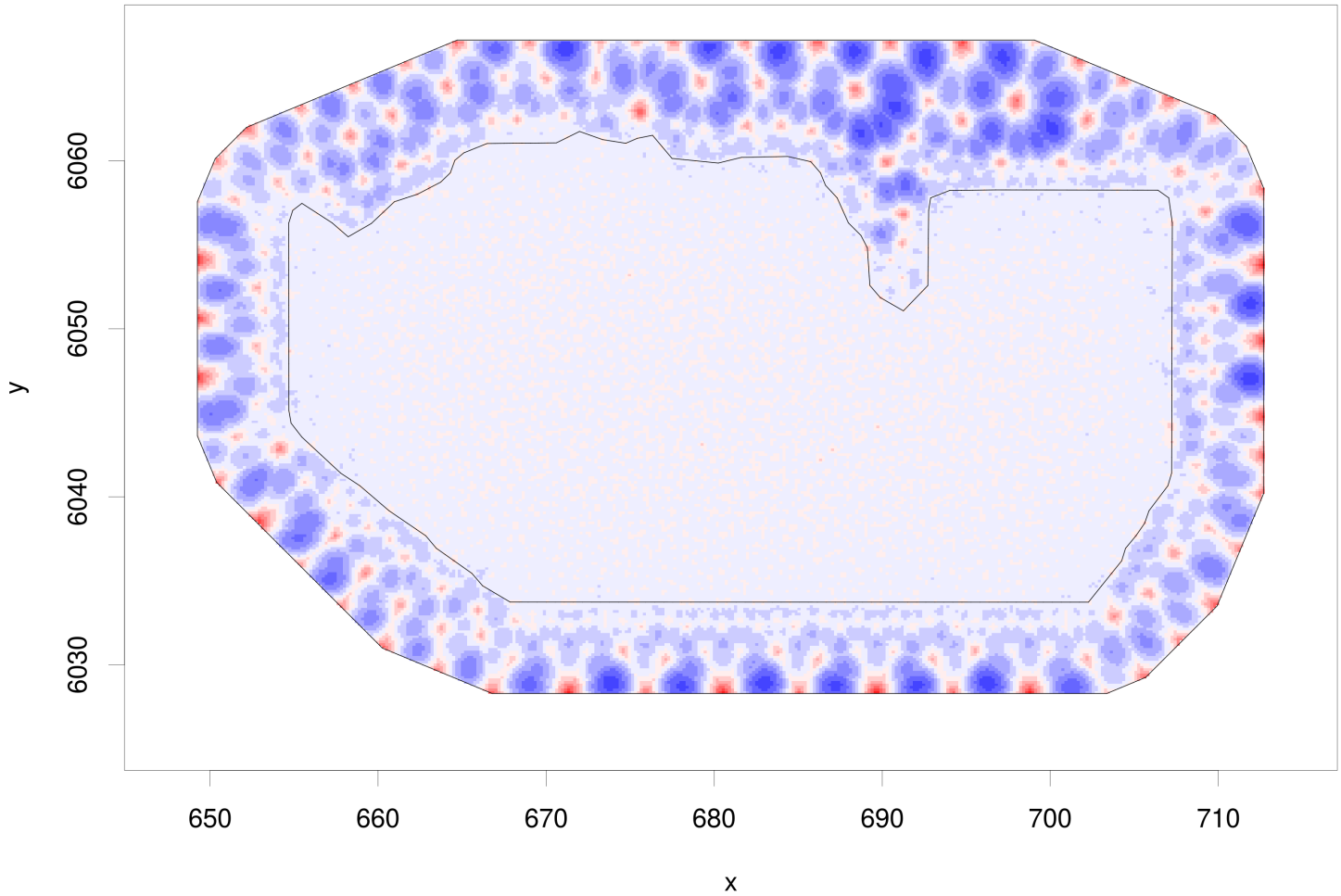
A refined mesh

Refining the mesh in the domain of interest



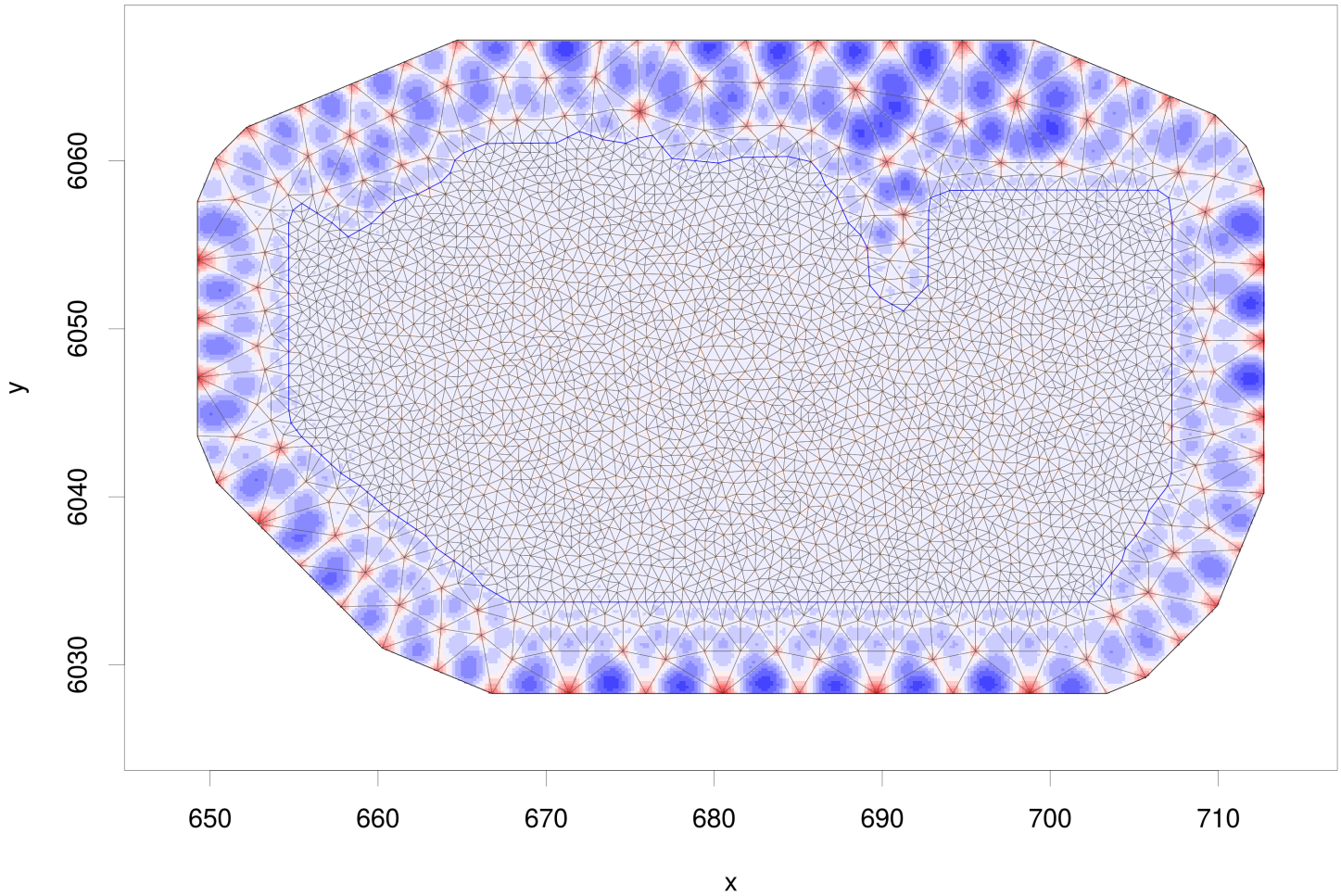
A refined mesh

Refining the mesh in the domain of interest



A refined mesh

Refining the mesh in the domain of interest

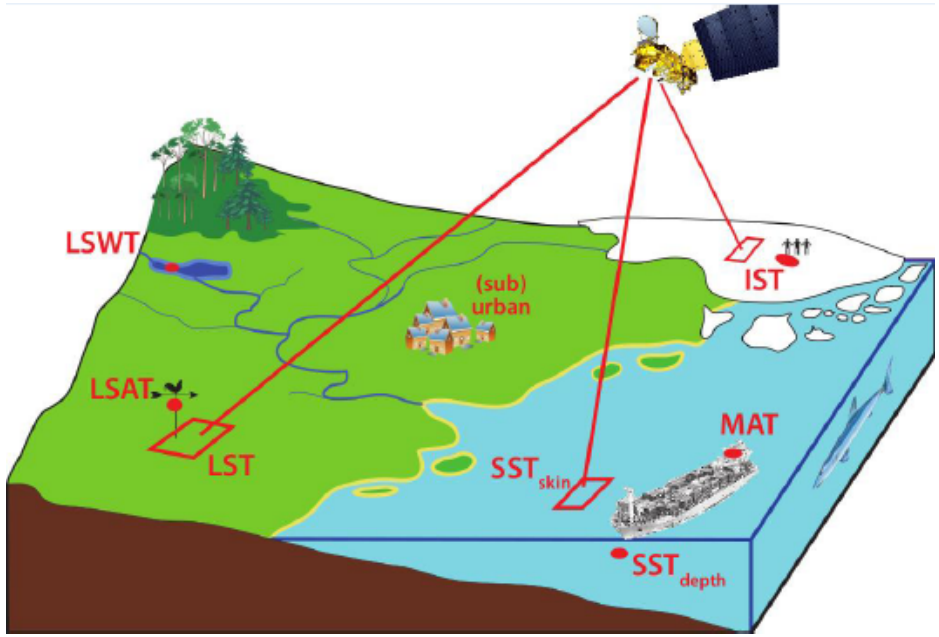


EUSTACE (*EU Surface Temperatures for All Corners of Earth*)

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

EUSTACE

EUSTACE will give publicly available daily estimates of surface air temperature since 1850 across the globe for the first time by combining surface and satellite data using novel statistical techniques.



Nick A. Rayner, Renate Auchmann, Janette Bessembinder, Stefan Brönnimann, Yuri Brugnara, Laura Carrea, Darren Ghent, Elizabeth Good, Katie Herring, Jacob Høyer, John Kennedy, Albert Klein Tank, Finn Lindgren, Colin Morice, Chris Merchant, John Remedios, Ag Stephens and Rasmus Tonboe

contact: nick.rayner@metoffice.gov.uk

Royal Netherlands Meteorological Institute *Ministry of Infrastructure and the Environment*

Science & Technology Facilities Council

University of Leicester

Met Office Hadley Centre

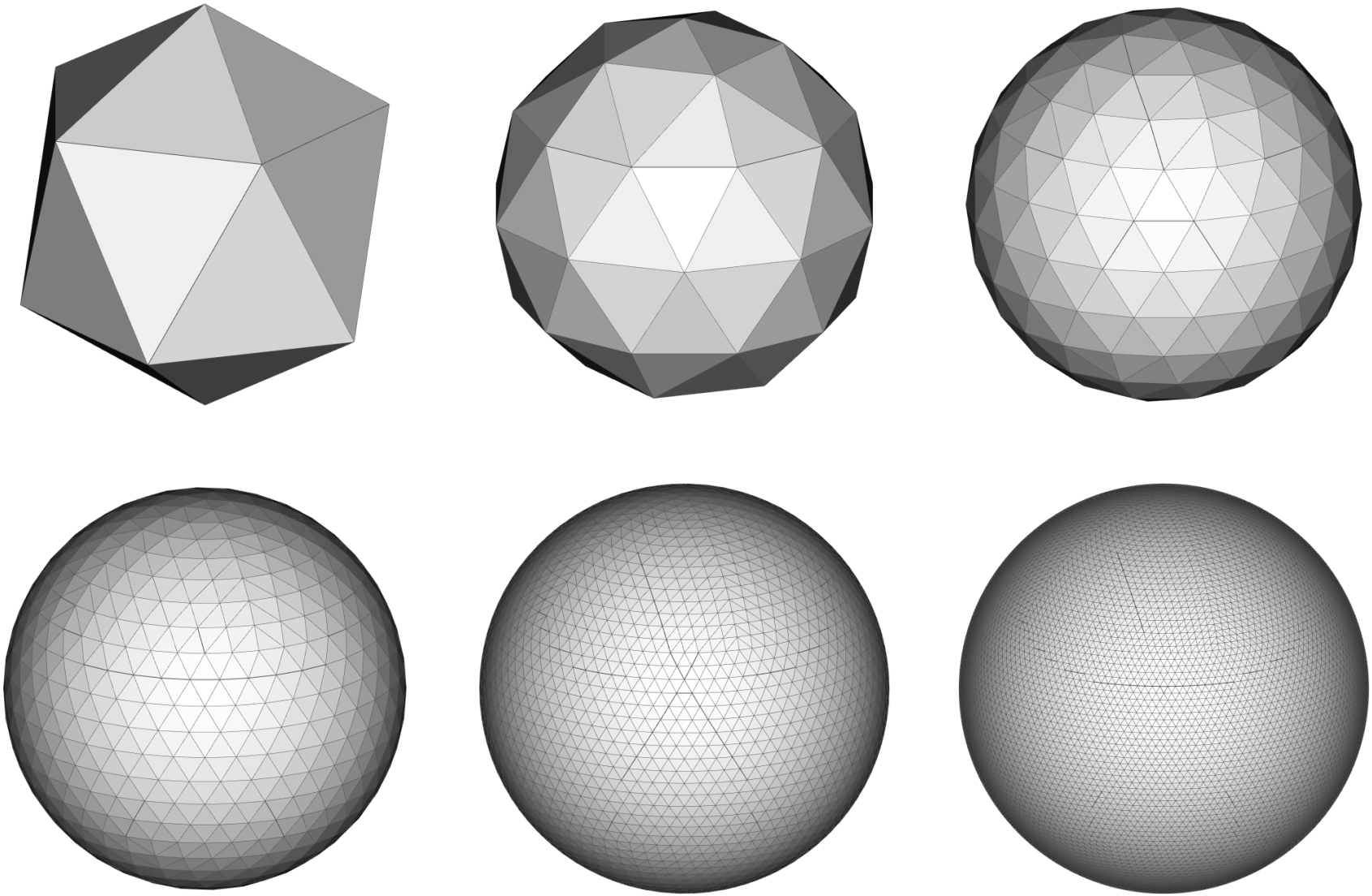
University of Reading

University of Bath

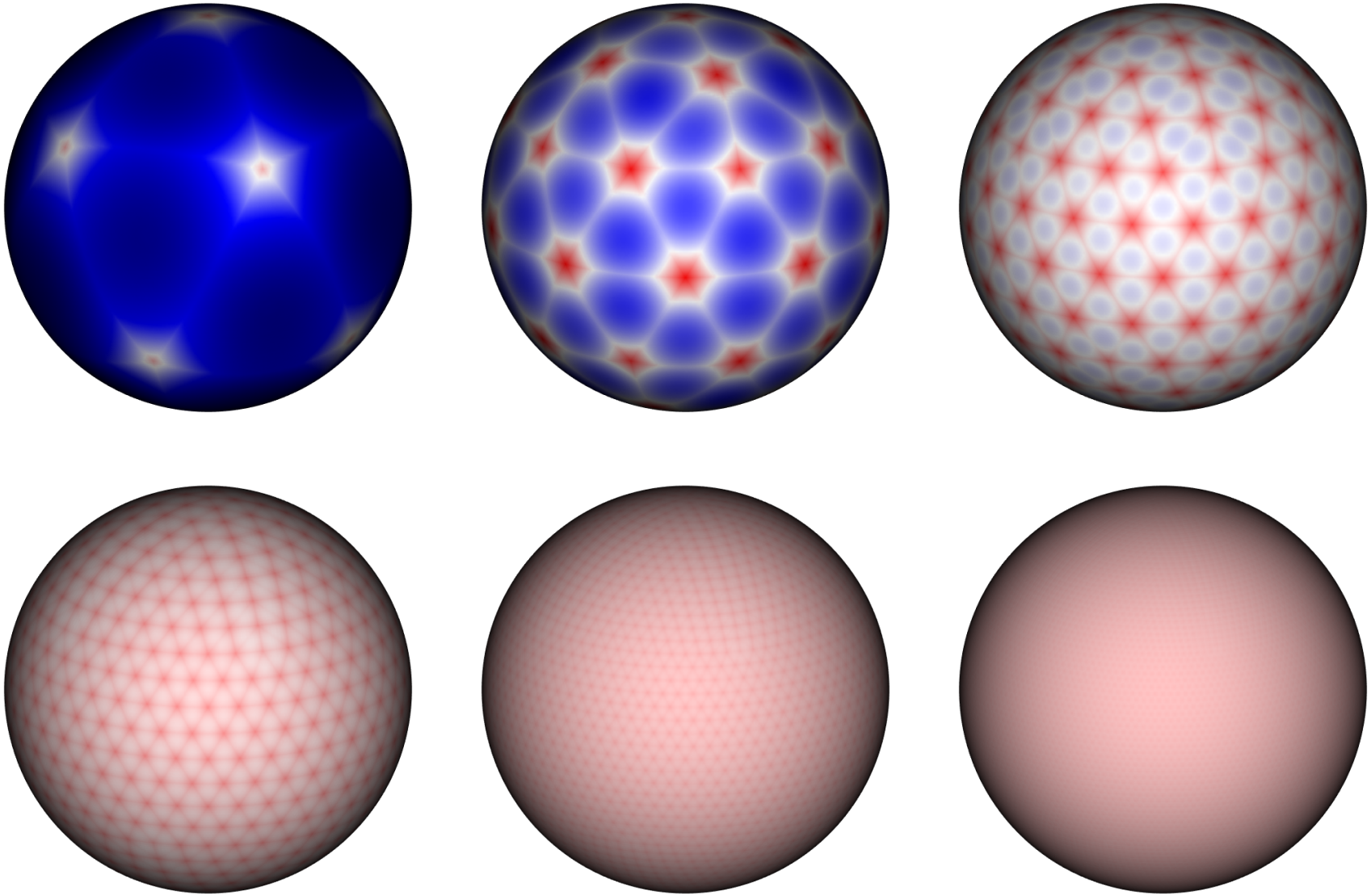
Universität Bern Oeschger Centre Climate Change Research

DMI

Triangulations for all corners of Earth

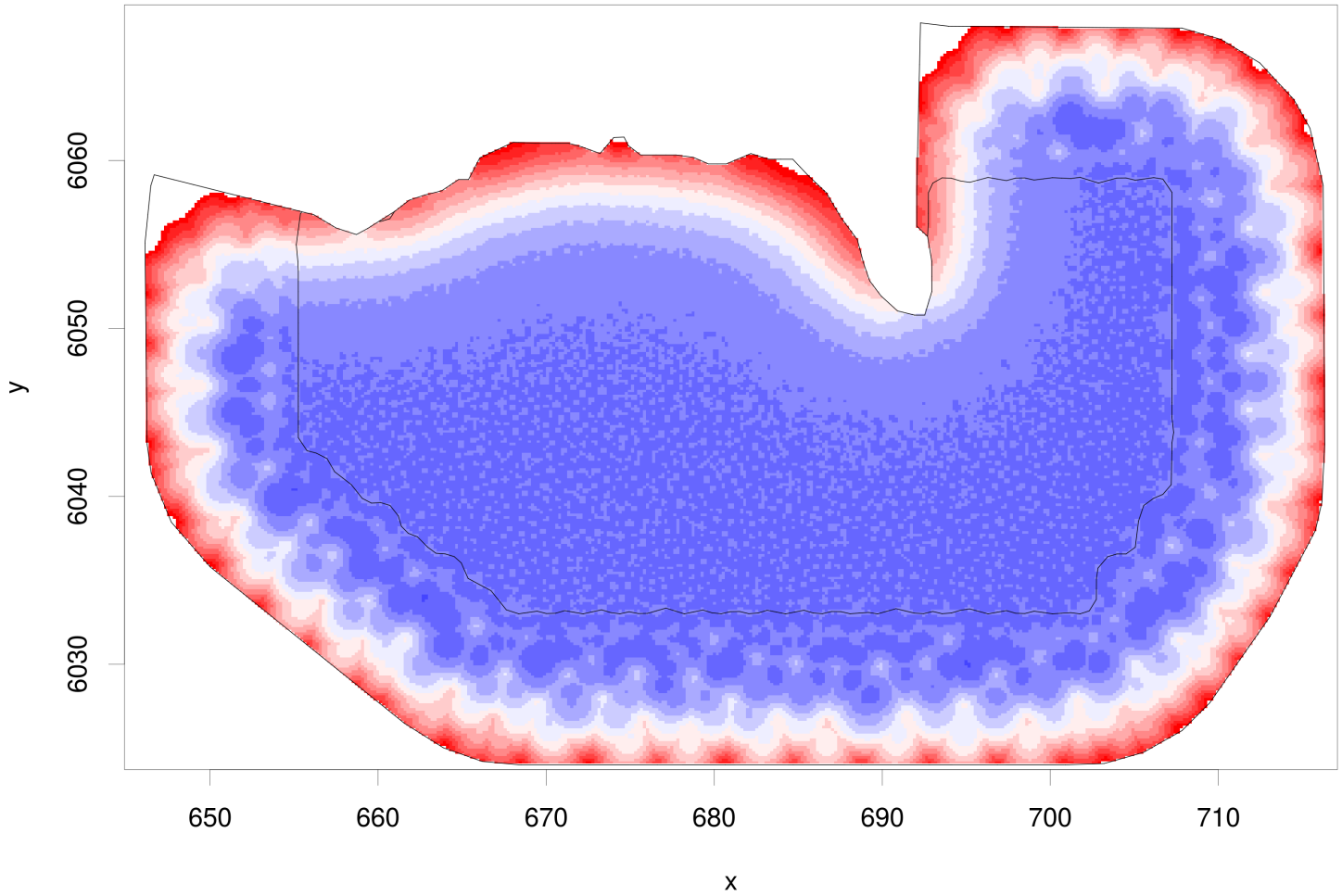


Triangulations for all corners of Earth



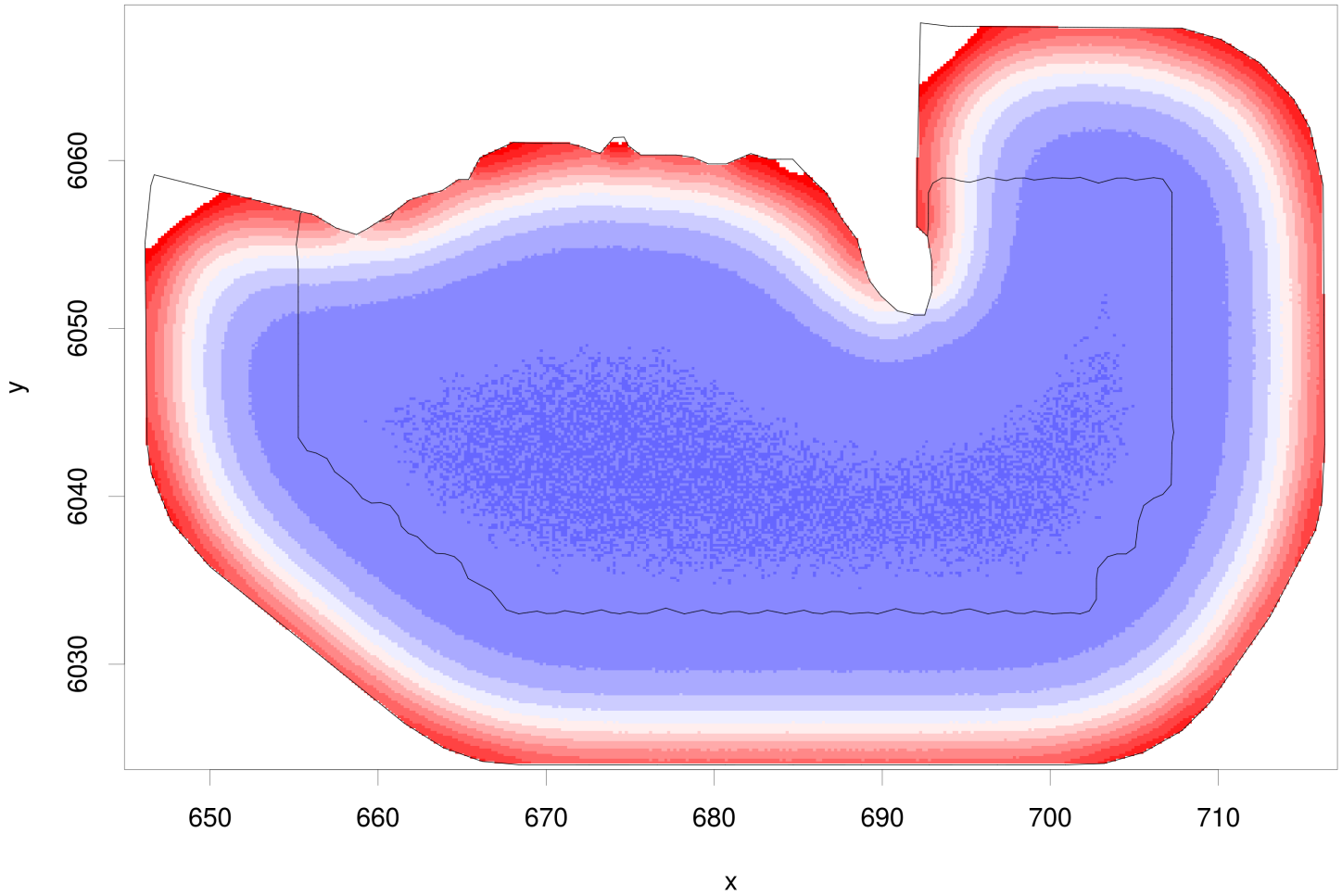
A coastline mesh

The boundary effect can be appropriate for coastlines



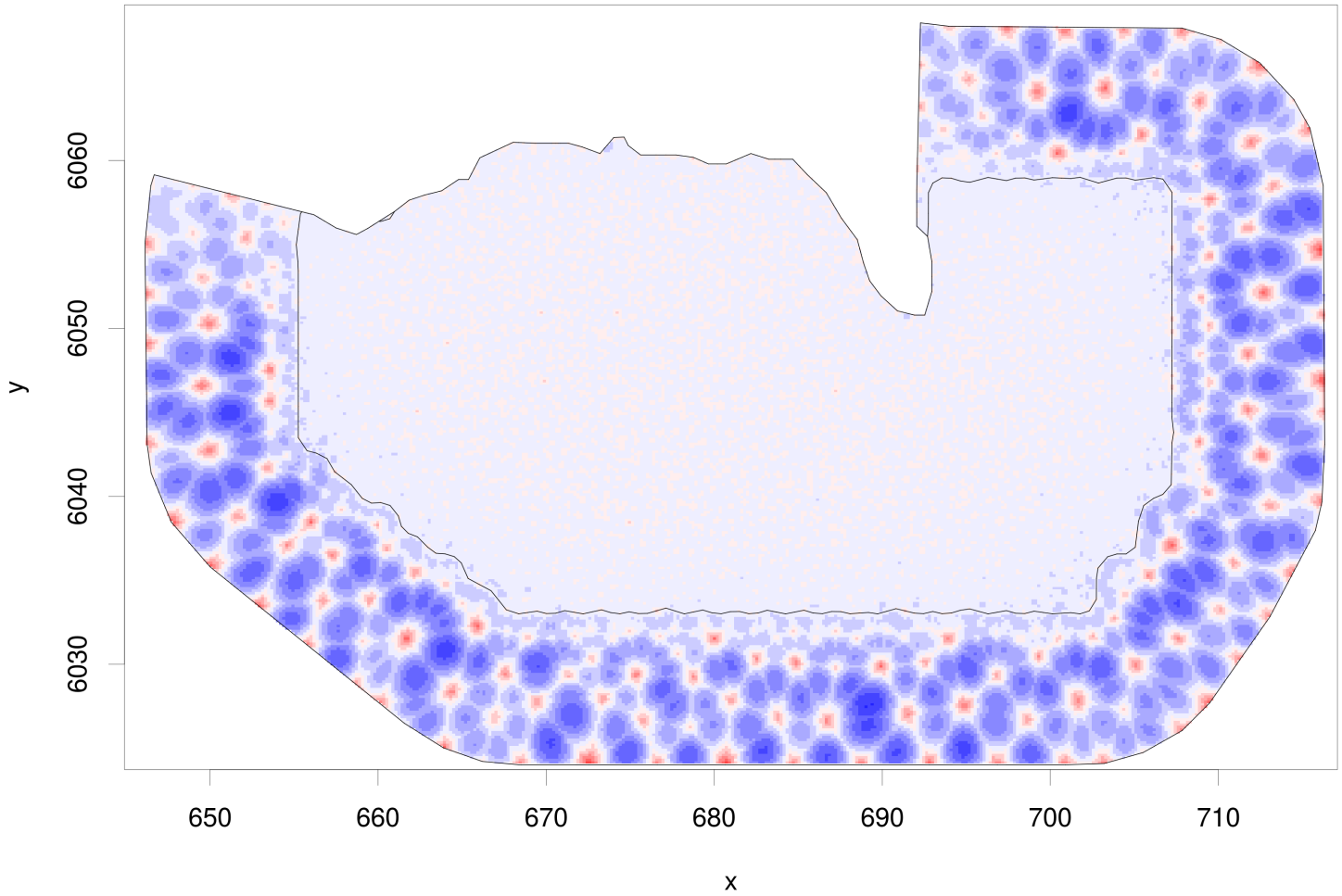
A coastline mesh

The boundary effect can be appropriate for coastlines



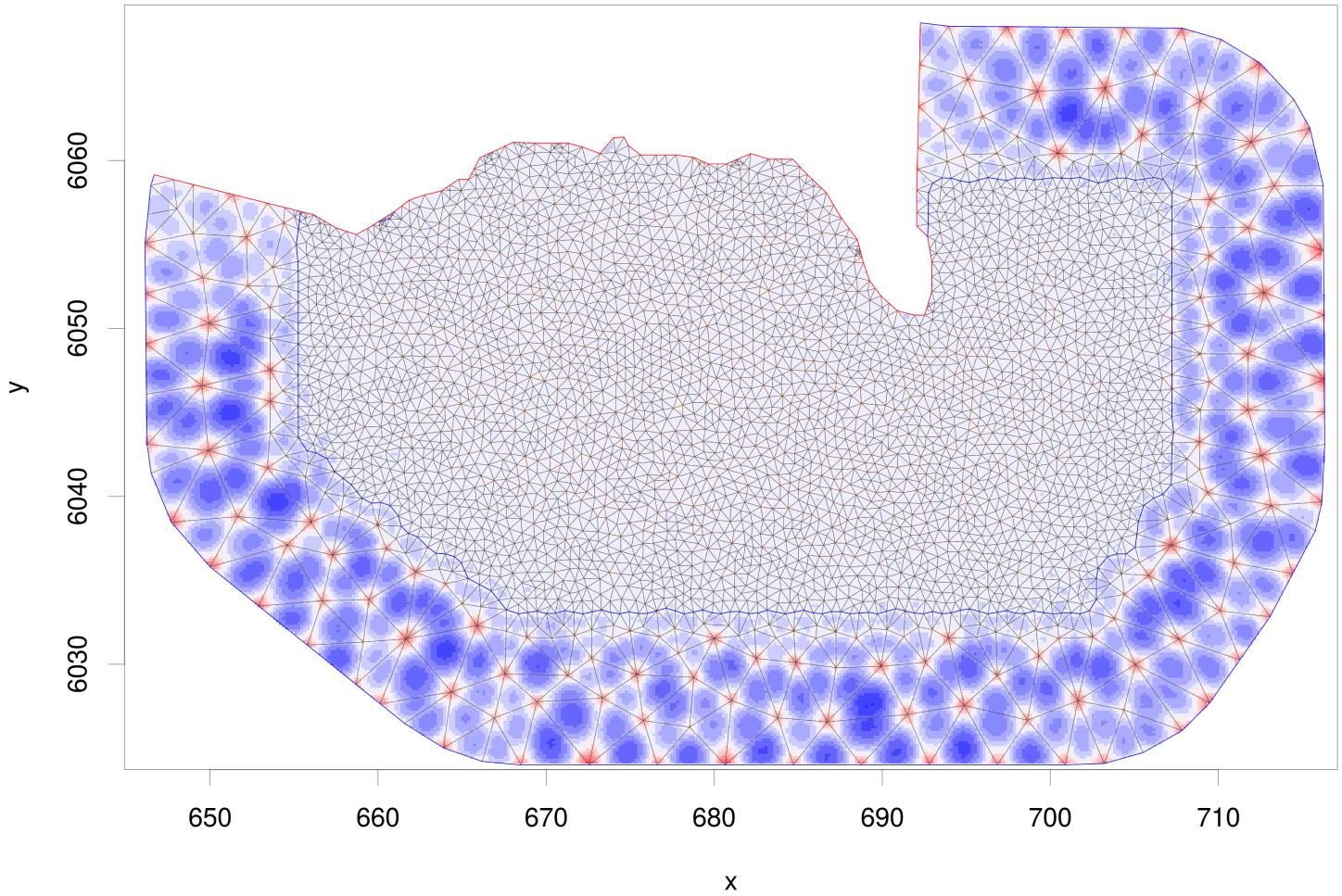
A coastline mesh

The boundary effect can be appropriate for coastlines

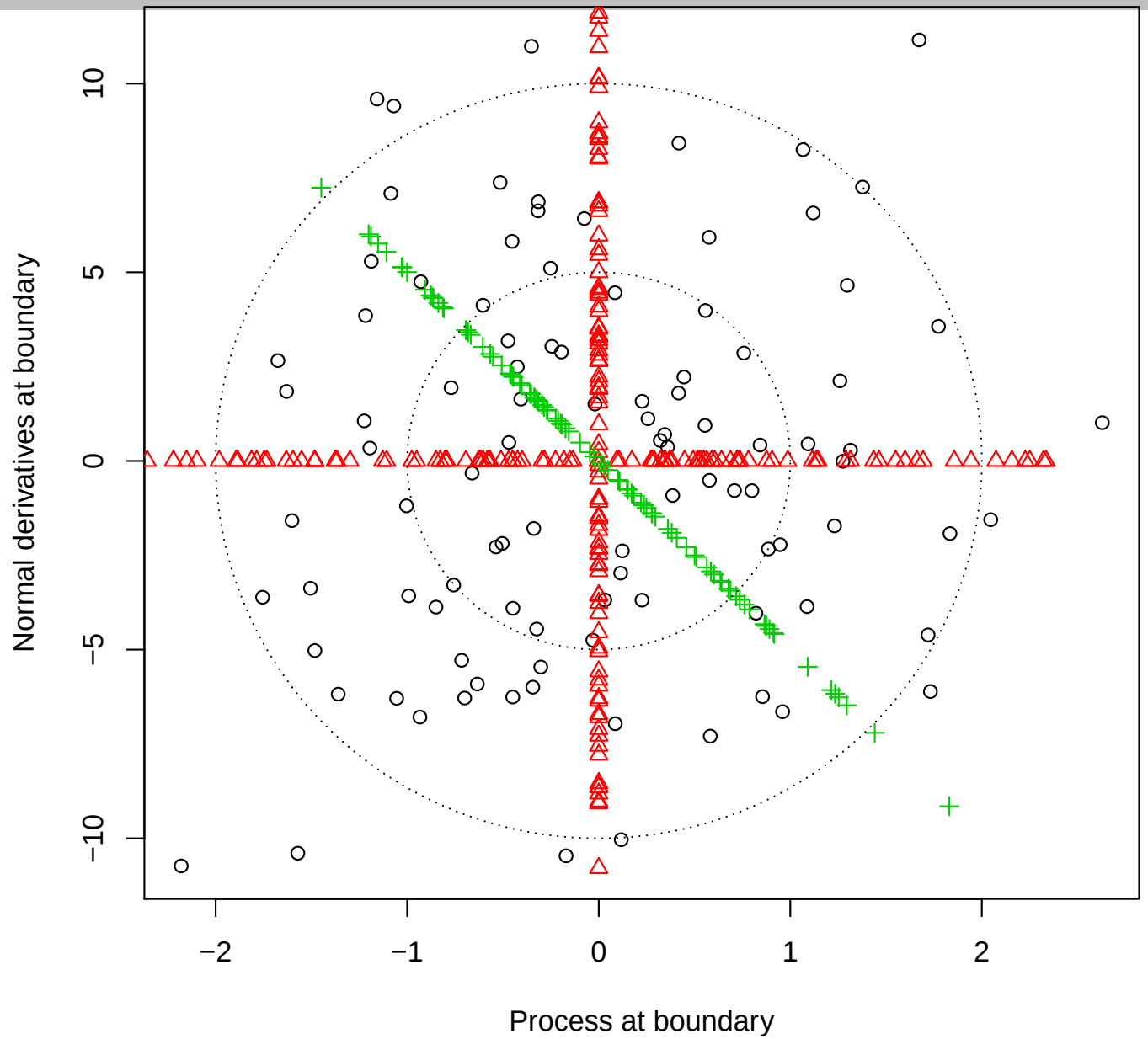


A coastline mesh

The boundary effect can be appropriate for coastlines



All deterministic boundary conditions are ‘inappropriate’



Stationary stochastic boundary adjustment (current work)

Recall the Matérn generating SPDE

$$(\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s})$$

RKHS inner product/precision operator for Matérn fields on $\mathbb{R}^d$:

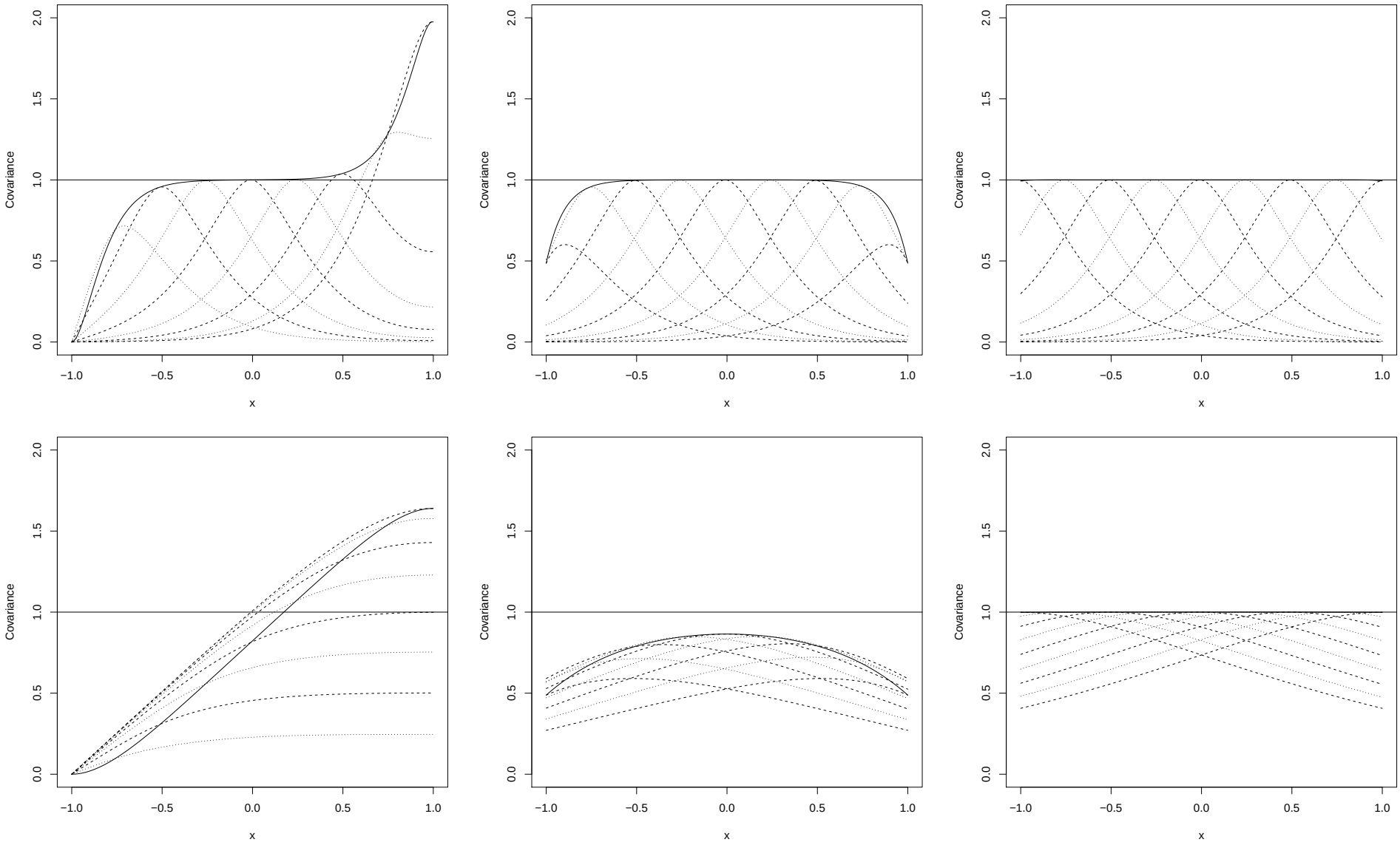
$$\langle f, \mathcal{Q}_{\mathbb{R}^d} g \rangle_D = \sum_{k=0}^{\alpha} \binom{\alpha}{k} \kappa^{2\alpha-2k} \langle \nabla^k f, \nabla^k g \rangle_D$$

Boundary adjusted precision operator on a compact subdomain, where $\mathcal{P}$ is a conditional expectation operator:

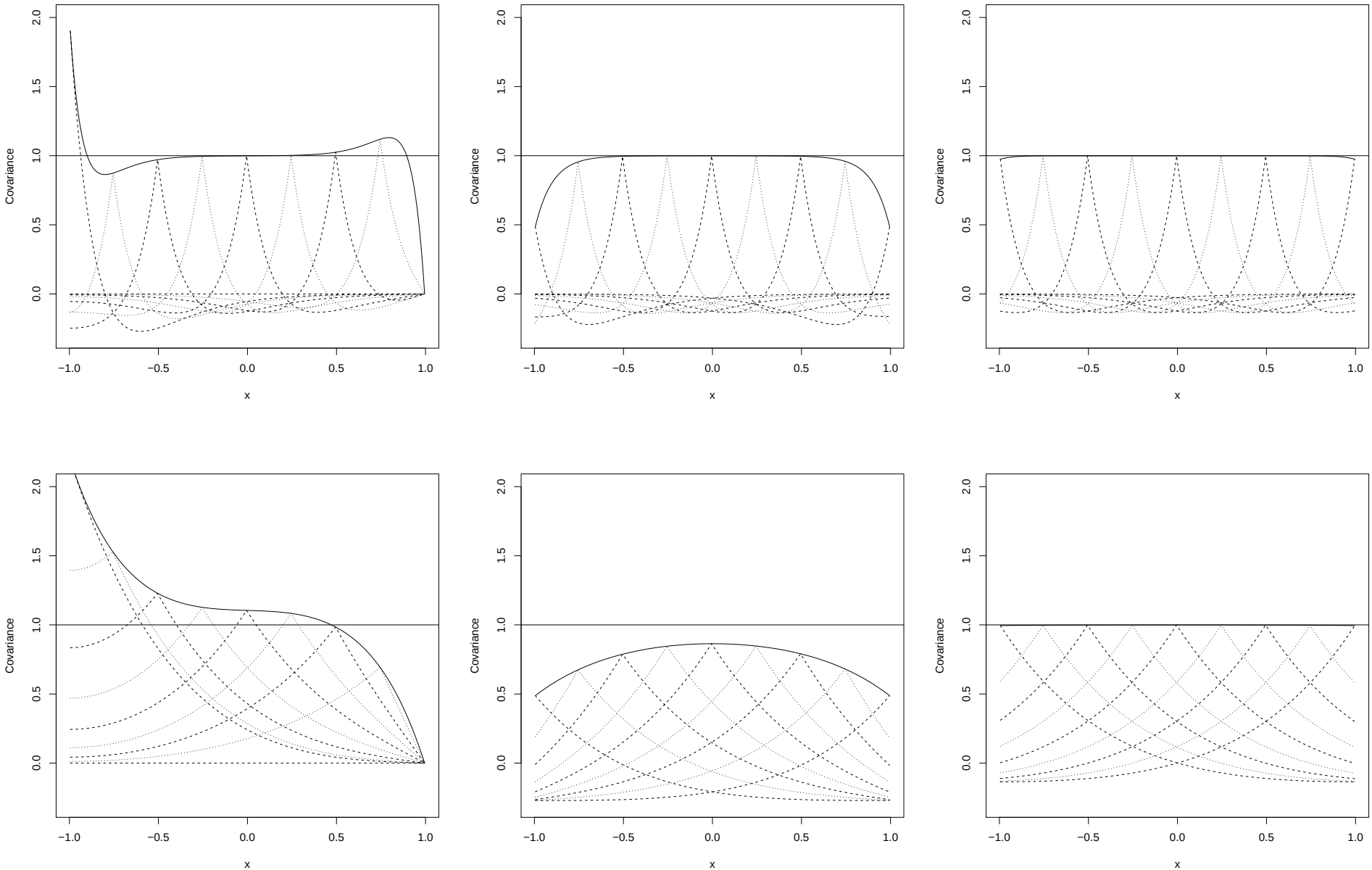
$$\begin{aligned} \langle f, \mathcal{Q}_D g \rangle_D &= \langle f, \mathcal{Q}_{\mathbb{R}^2} g \rangle_D - \langle \mathcal{P} f, \mathcal{Q}_{\mathbb{R}^2} \mathcal{P} g \rangle_D + \langle f, \mathcal{Q}_{\partial D} g \rangle_{\partial D} \\ &= \langle f - \mathcal{P} f, \mathcal{Q}_{\mathbb{R}^2} (g - \mathcal{P} g) \rangle_D + \langle f, \mathcal{Q}_{\partial D} g \rangle_{\partial D} \,, \end{aligned}$$

where the boundary precision operator may involve normal derivatives of f and g.

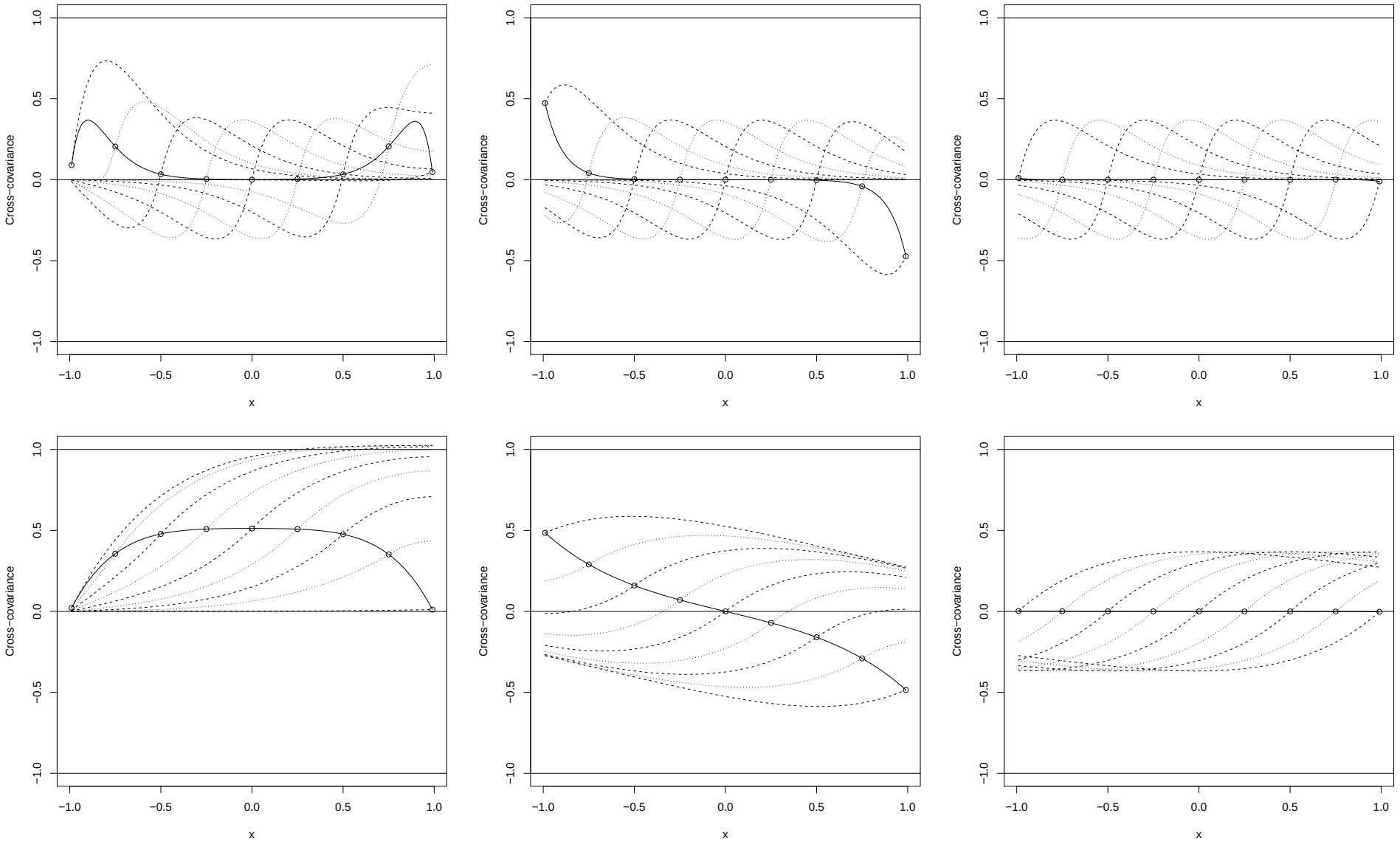
Covariances (D&N, Robin, Stoch) for $\kappa = 5$ and 1



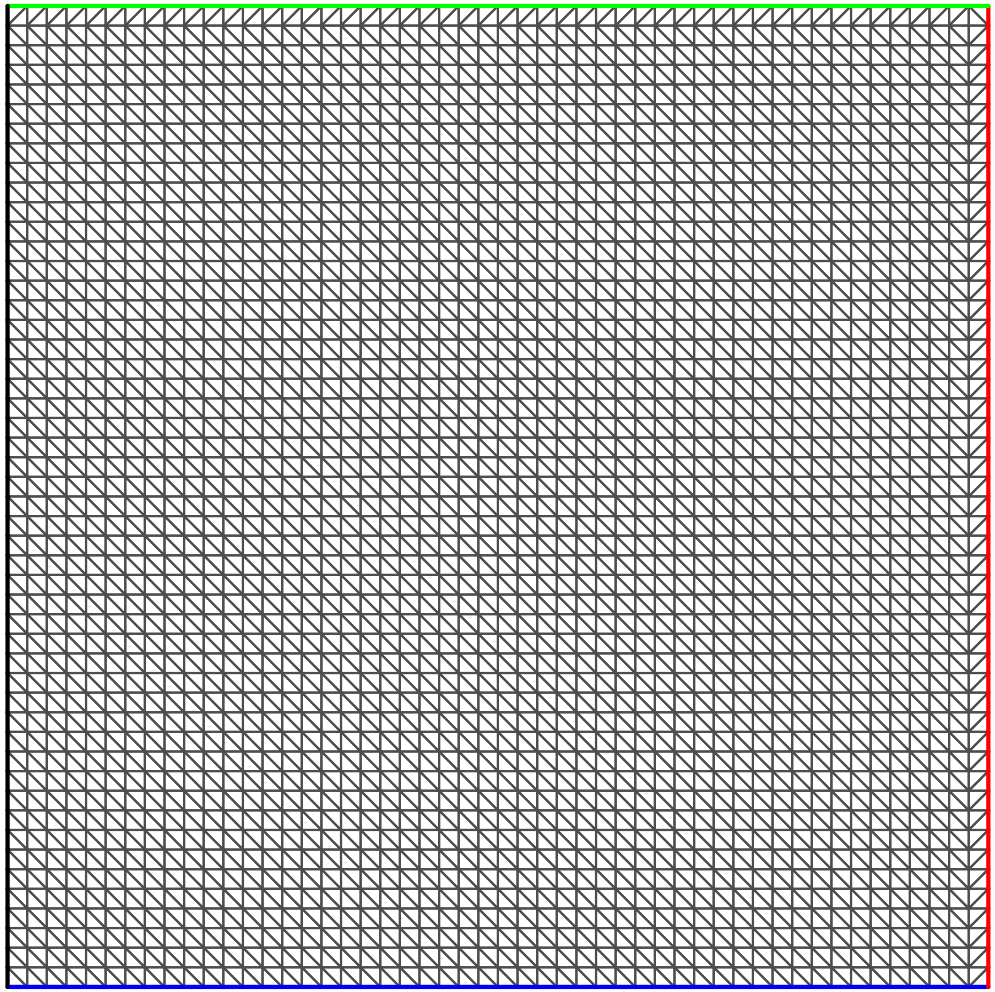
Derivative covariances (D&N, Robin, Stoch) for $\kappa = 5$ and 1



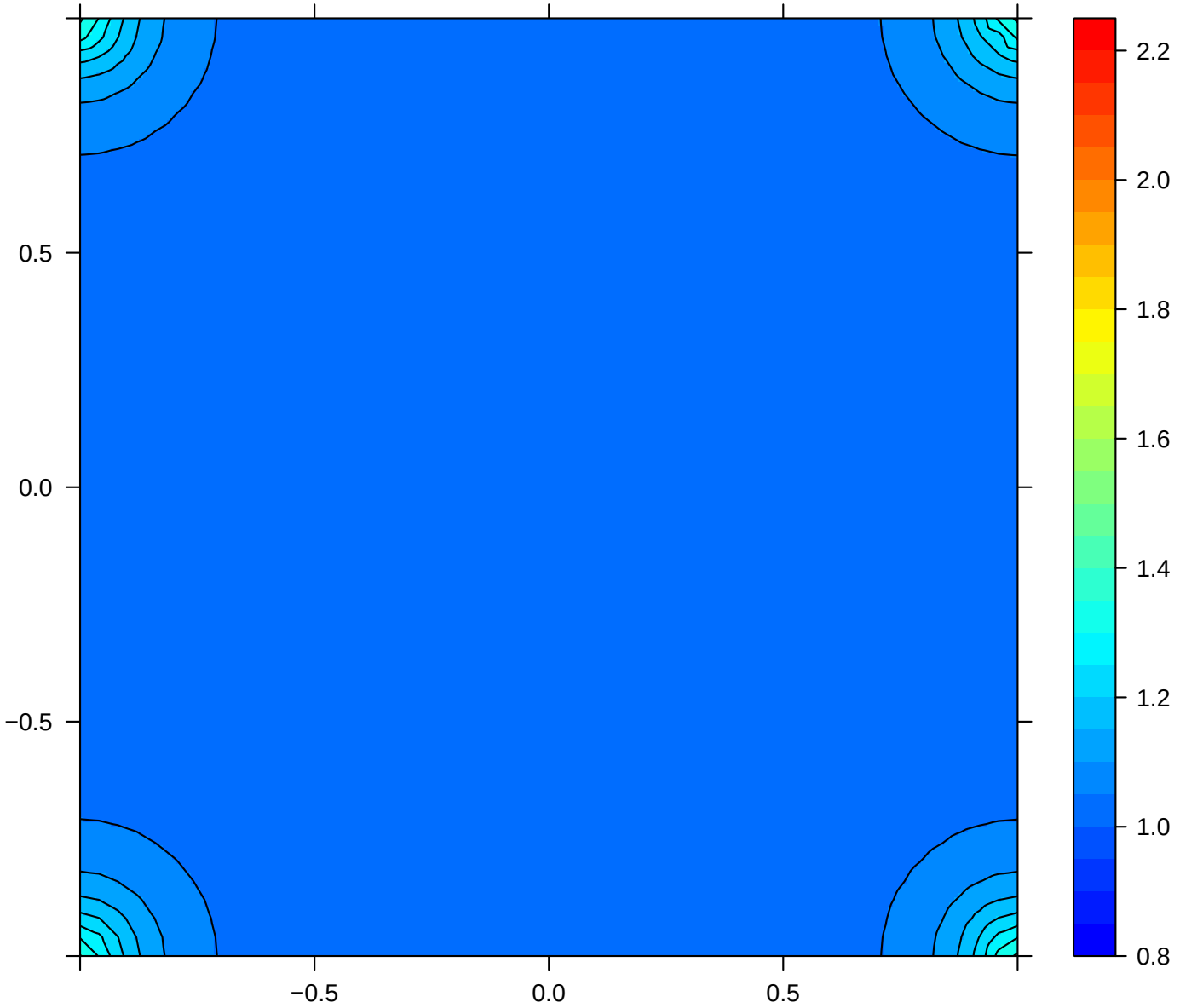
Process-derivative cross-covariances (D&N, Robin, Stoch)



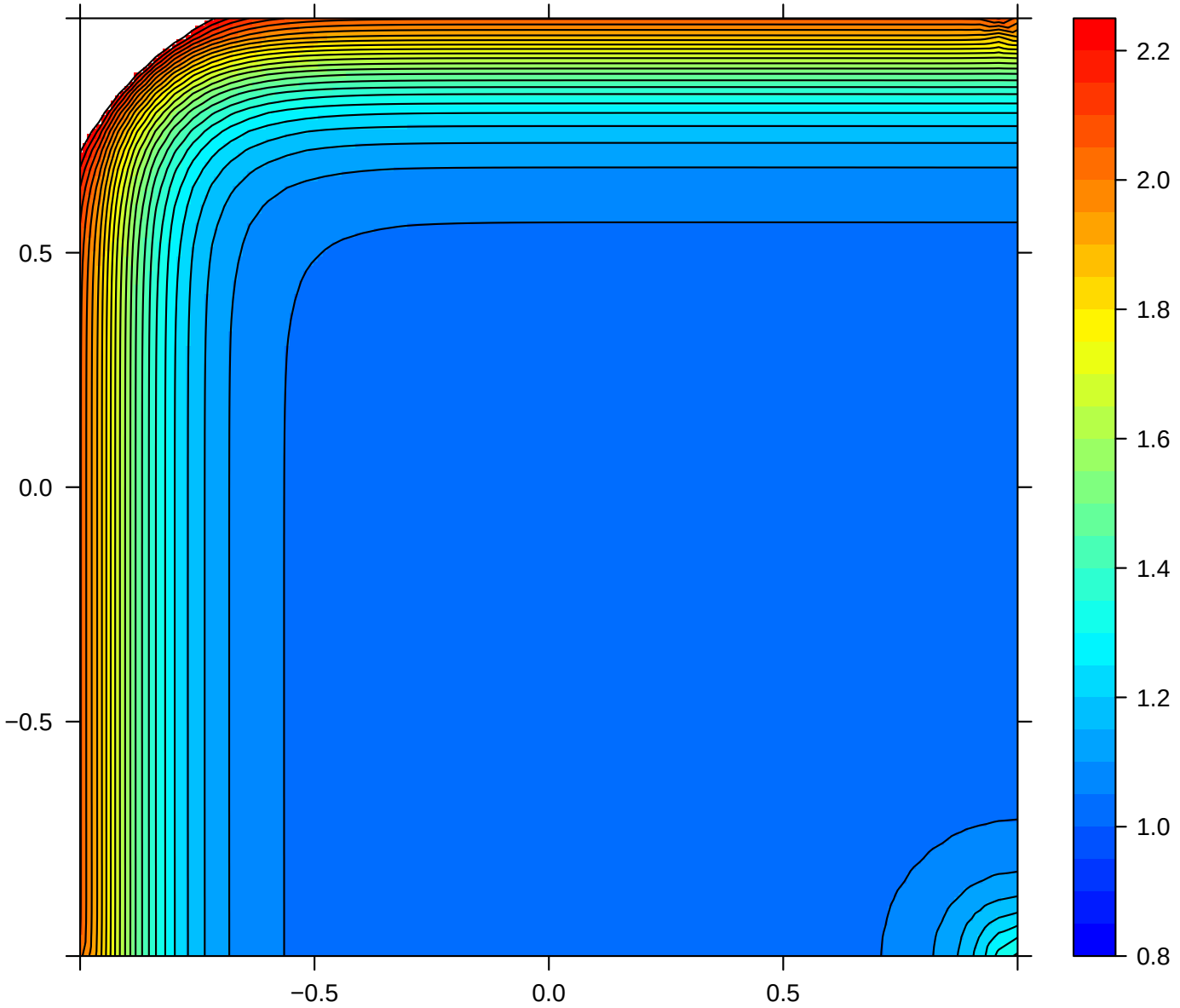
Square domain, basis triangulation



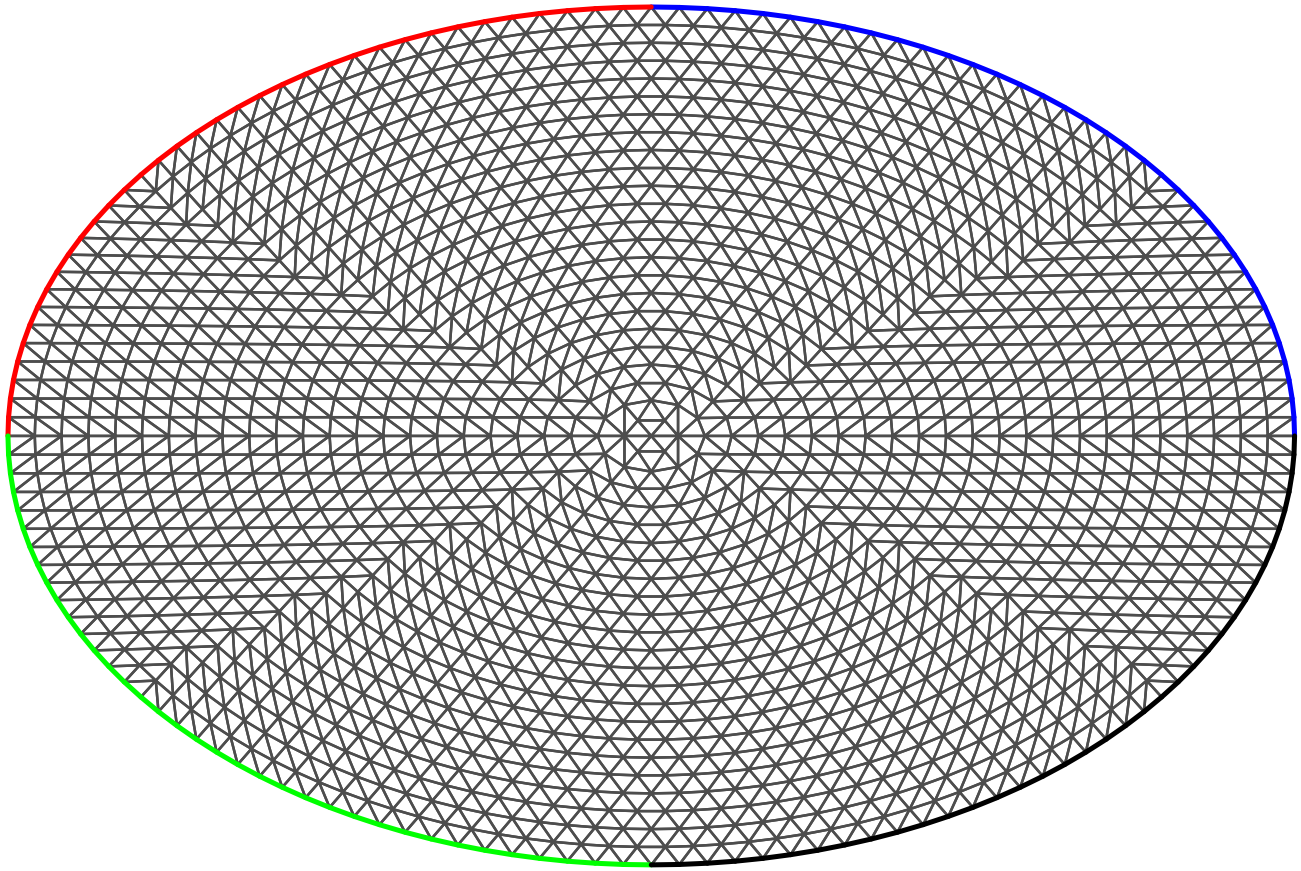
Square domain, stochastic boundary (variances)



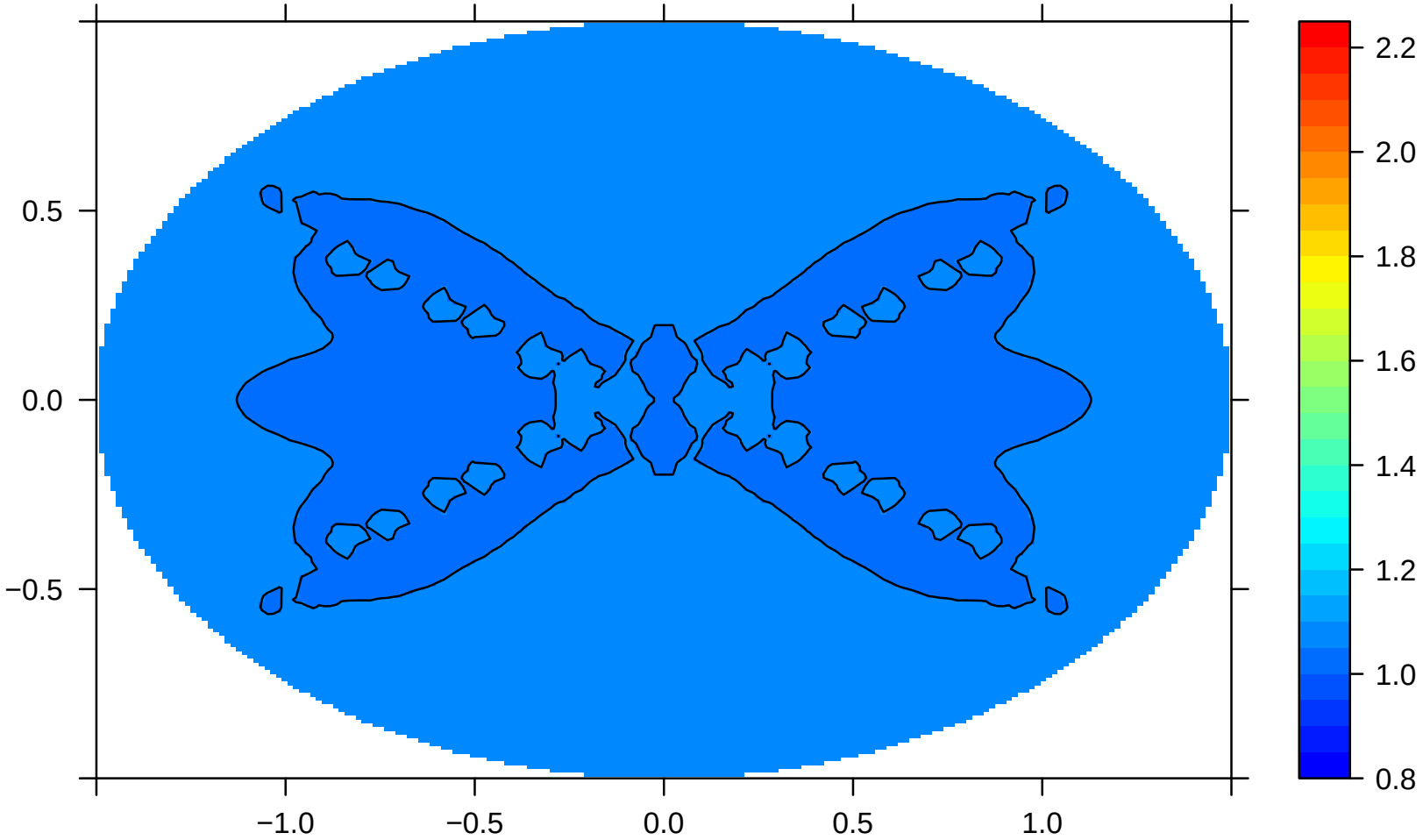
Square domain, mixed boundary (variances)



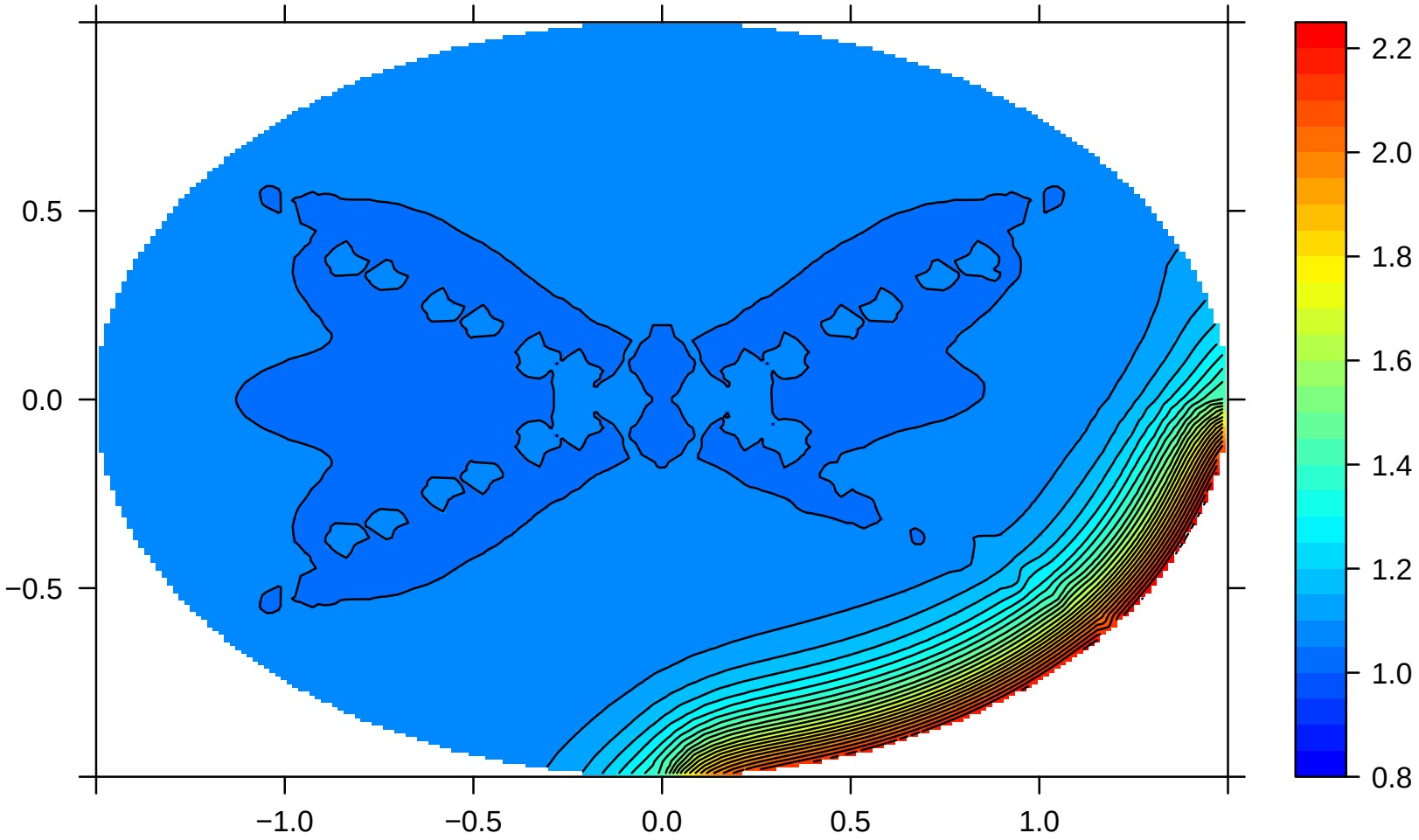
Elliptical domain, basis triangulation



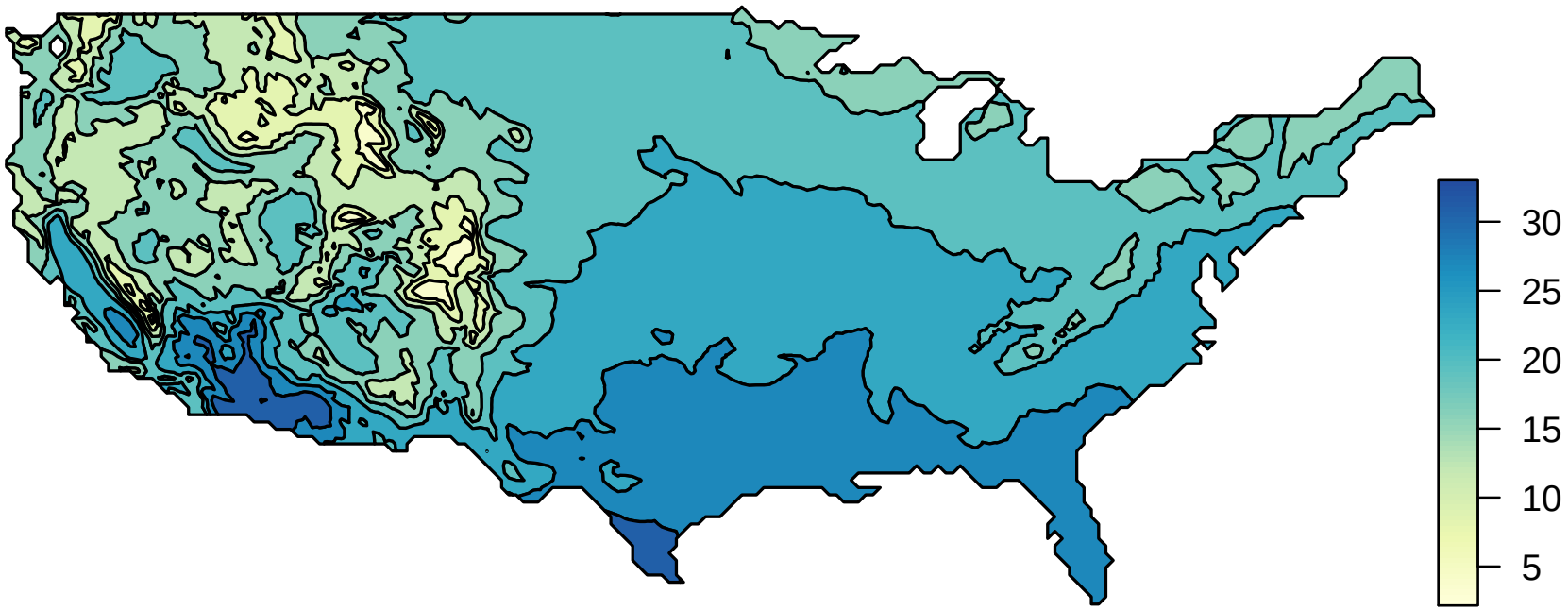
Elliptical domain, stochastic boundary (variances)



Elliptical domain, mixed boundary (variances)



Contour map for estimated US summer mean temperature



Can we trust the apparent details af the level crossings?
How many levels can we sensibly use?
Can we put a number on the statistical quality of the contour map?
Fundamental question:
What *is* the statistical interpretation of a contour map?

Spatial latent Gaussian models

Consider a spatial linear model

$$\begin{aligned}
\boldsymbol{\beta} &\sim \mathsf{N}(\mathbf{0}, \boldsymbol{I}\sigma_\beta^2), \\
\xi(\boldsymbol{s}) &\sim \text{Gaussian random field}, \\
x(\boldsymbol{s}) &= \boldsymbol{z}(\boldsymbol{s})\boldsymbol{\beta} + \xi(\boldsymbol{s}), \\
(y_i | x(\cdot)) &\sim \mathsf{N}(x(\boldsymbol{s}_i), \sigma_e^2),
\end{aligned}$$

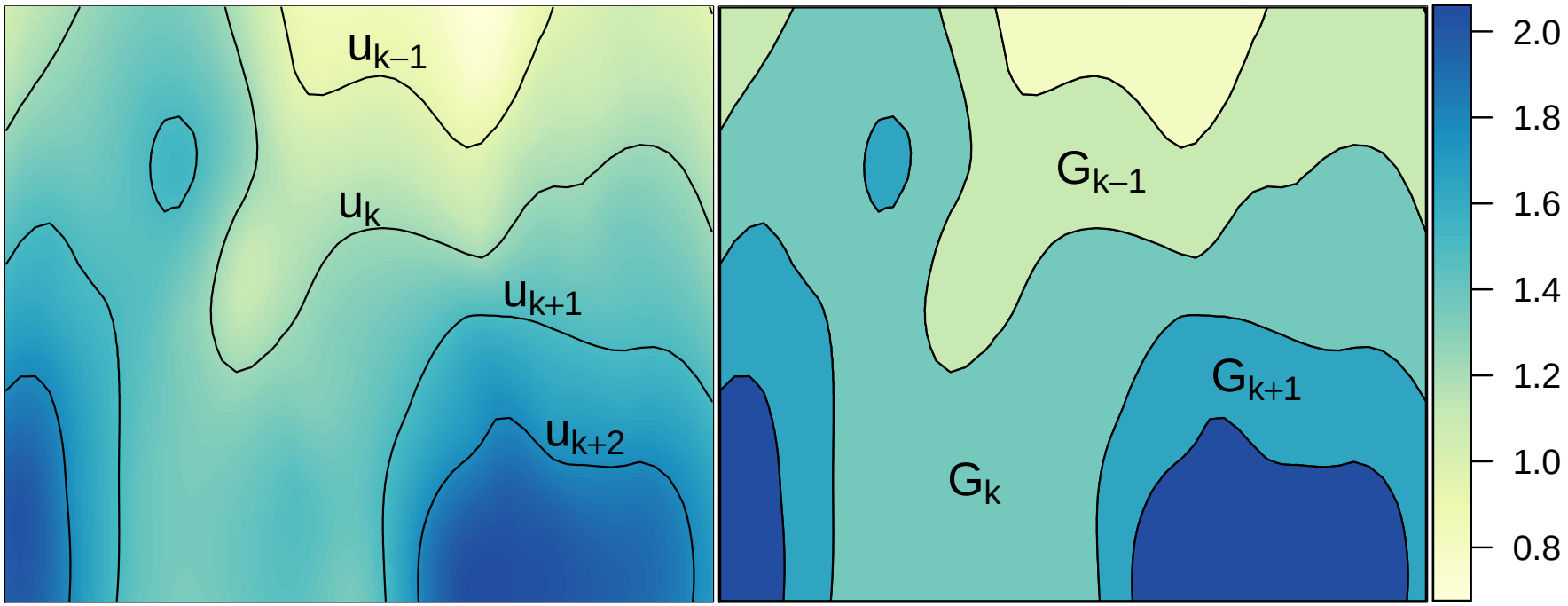
where $\boldsymbol{z}(\cdot)$ are spatially indexed explanatory variables, and y_i are conditionally independent observations.

- A contour curve for a level u crossing is typically calculated as the level u crossing of $\widehat{x}(\boldsymbol{s}) = \mathsf{E}[x(\boldsymbol{s})|\boldsymbol{y}]$.
- In practice, we want to interpret it as being informative about the potential level crossings of the unknown random field $x(\boldsymbol{s})$ itself.

Level sets and contour maps

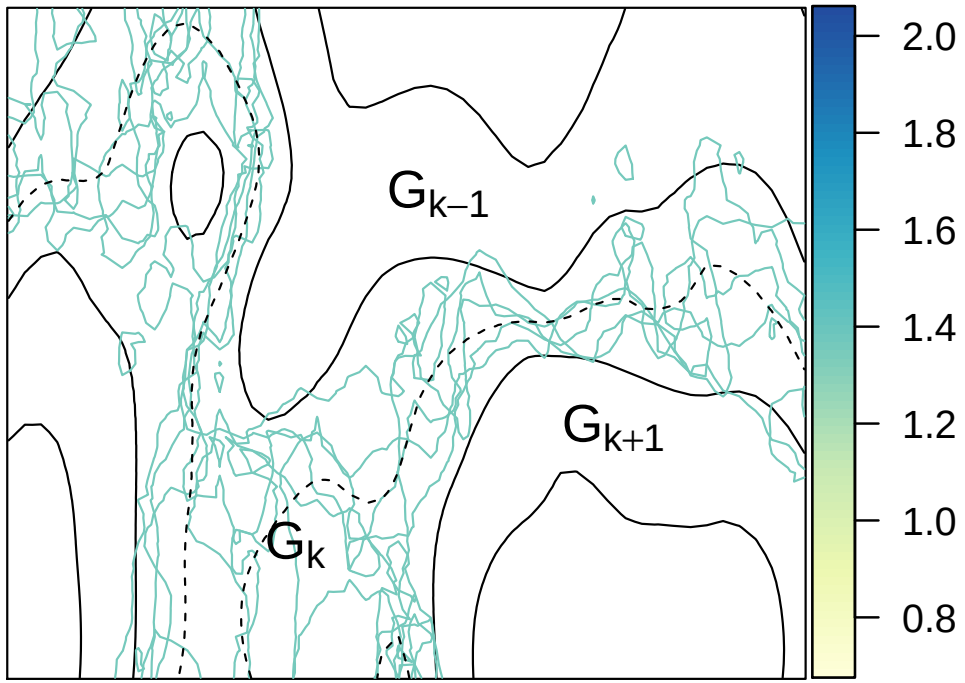
Given a function $f(\boldsymbol{s})$, $\boldsymbol{s} \in \Omega$, and levels $u_1 < u_2 < \cdots < u_K$, the *level sets* are $G_k(f) = \{\boldsymbol{s};\ u_k < f(\boldsymbol{s}) < u_{k+1}\}$.

The contour map uses colours from the midpoint between the levels.



The *contour map* for $f(\cdot)$ is the collection $C_{\boldsymbol{u}}(f) = \{G_k(f),\ k = 1, \ldots, K\}$

Probabilistic interpretation of contour maps for random fields



$C_{\boldsymbol{u}}(\widehat{x})$ and five random contour curves for $x(\cdot)$, for $u_k^e = \frac{u_k + u_{k+1}}{2}$. If

$$1 - \alpha \leq P_2 = \mathsf{P}\left(\cap_k \{x(\boldsymbol{s}) < u_k^e \text{ where } \widehat{x}(\boldsymbol{s}) < u_k\} \cap \{x(\boldsymbol{s}) > u_k^e \text{ where } \widehat{x}(\boldsymbol{s}) > u_{k+1}\} \mid \boldsymbol{y}\right),$$

then $C_{\boldsymbol{u},\alpha}(x) = C_{\boldsymbol{u}}(\widehat{x})$ are joint $1-\alpha$ credible regions for the $\{u_k^e, k = 1, \ldots, K\}$ crossings.

Integrating for probabilities

Assuming that $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ is, or can be approximated as, Gaussian, there are several ways to calculate the excursion probabilities, one of which is

Numerical integration

Numerically approximate the excursion probability by approximating the posterior probability as

$$\begin{aligned}\mathsf{P}(\boldsymbol{a}<\boldsymbol{x}<\boldsymbol{b}|\boldsymbol{y}) &= \mathsf{E}[\mathsf{P}(\boldsymbol{a}<\boldsymbol{x}<\boldsymbol{b}|\boldsymbol{\theta},\boldsymbol{y})|,\boldsymbol{y}] \\ &\approx \sum_{j=1}^{m} w_i \mathsf{P}(\boldsymbol{a}<\boldsymbol{x}<\boldsymbol{b}|\boldsymbol{\theta}_j,\boldsymbol{y}),\end{aligned}$$

where the configurations $\{\boldsymbol{\theta}_j, j=1,\ldots,m\}$ are taken from INLA and the weights w_j are chosen proportional to $\pi(\boldsymbol{\theta}_i|\boldsymbol{y})$.

High dimensional sequential integration

- A GMRF can be viewed as a sequential process defined backwards in the indices of $\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$.
- Let $\boldsymbol{L}$ be the Cholesky factor in $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^\top$. Then

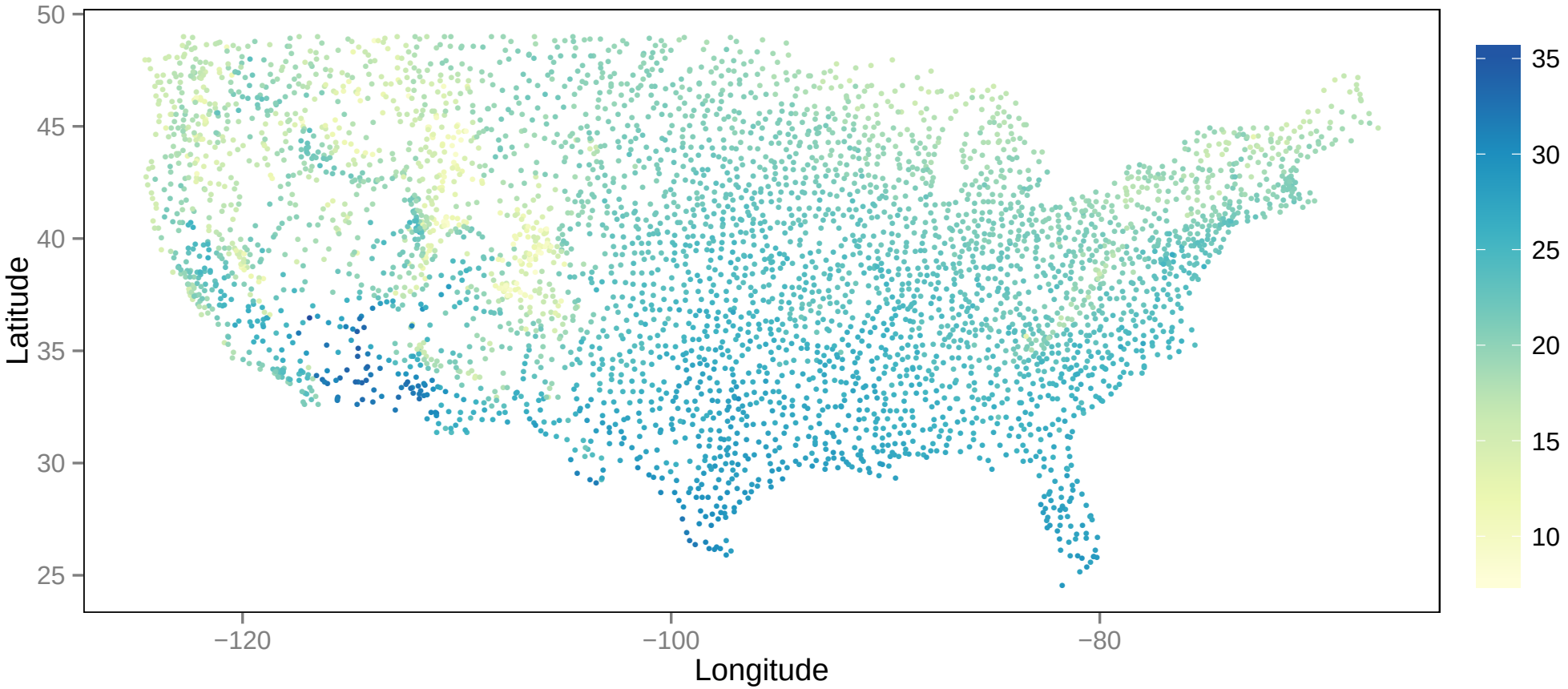
$$x_i | x_{i+1}, \ldots, x_n \sim \mathsf{N}\left(\mu_i - \frac{1}{L_{ii}} \sum_{j=i+1}^{n} L_{ji}(x_j - \mu_j), L_{ii}^{-2}\right)$$

- Denote the integral of the last $d - i$ components as I_i,

$$I_i = \int_{a_i}^{b_i} \pi(x_i | x_{i+1:d}) \cdots \int_{a_{d-1}}^{b_{d-1}} \pi(x_{d-1} | x_d) \int_{a_d}^{b_d} \pi(x_d) \, \mathrm{d}x,$$

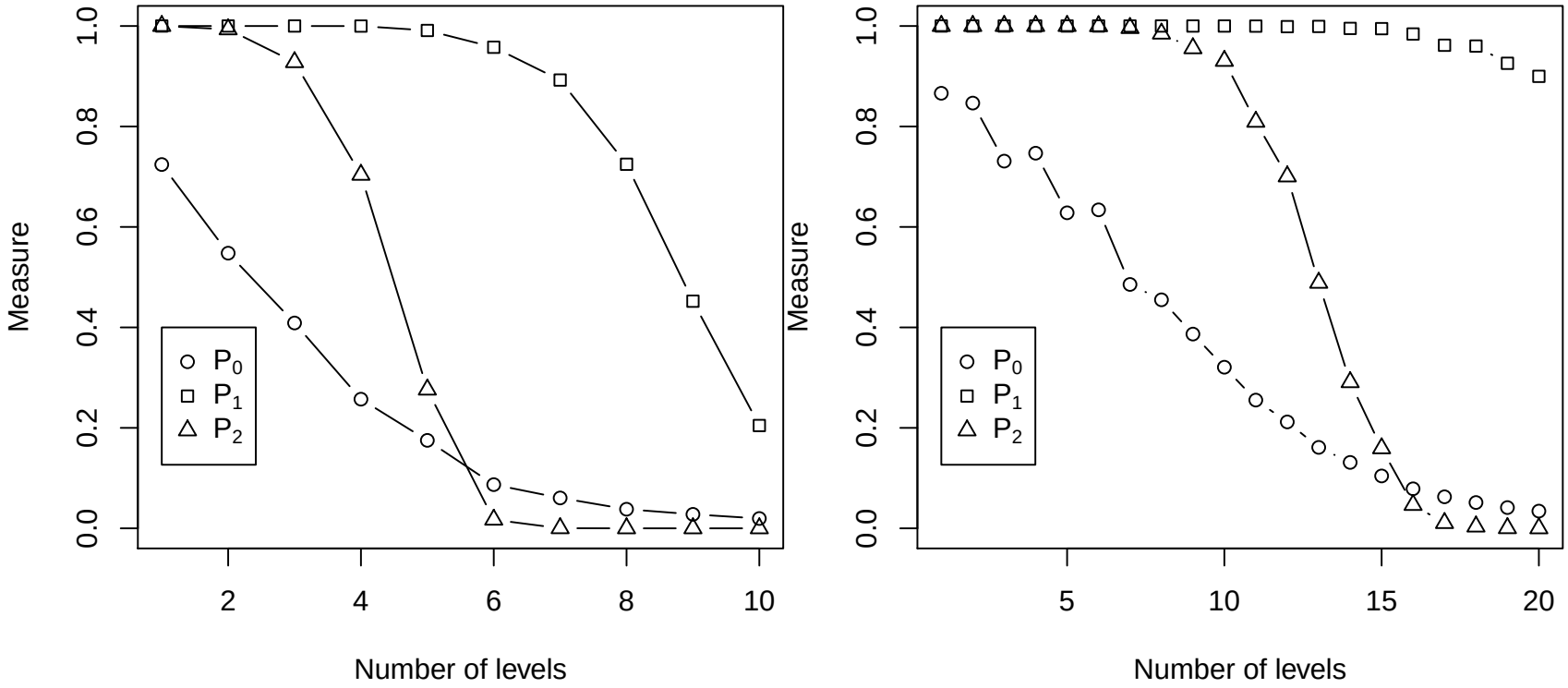
- $x_i | x_{i+1:d}$ only depends on the elements in $x_{\mathcal{N}_i \cap \{i+1:d\}}$, where $\mathcal{N}_i$ are the non-zero element indices for row i of $\boldsymbol{Q}$, plus Cholesky infill.
- The integrals are approxmated with sequential importance sampling.

Mean summer temperature measurements for 1997



We investigate a model without spatial explanatory variables, and a model using elevation as predictor for temperature.

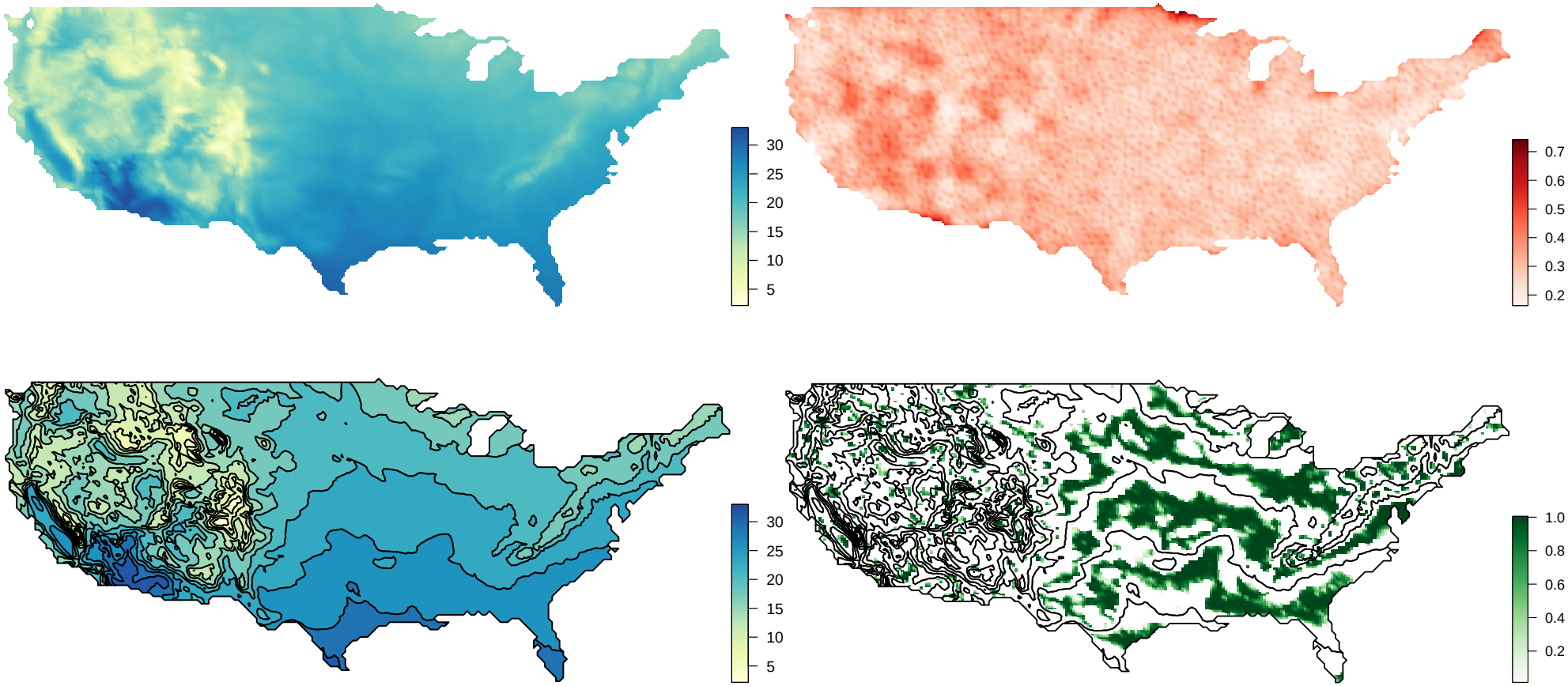
Contour map quality for different K and different models



We pick the largest K such that $P_2 \geq 1 - \alpha = 90\%$.

The spatial predictions are more certain in the model using elevation, allowing more contour levels to be used.

Posterior mean, s.d., contour map, and F_u, for $K = 10$



Contour map quality measure: $P_2 = 0.94$

Summary

- Random spatial variability can be represented in many different ways
- Some representations lead to efficient computational methods
- SPDEs lead to sparse precision matrices
- Direct approximate calculation of posterior distributions can be both faster and more accurate than MCMC simulation, for a given computational budget.
- Current work deals with scaling the high order operator sparse matrix algebra to large problems with 10^9–10^{11} unknowns.

Not discussed today:

- Spatio-temporal multiscale random field structure inspired by physics lead to large block matrices with high order operators
- Links between preconditioners and Metropolis-Hastings proposals

References

- Lindgren, F., Rue, H. and Lindström, J. (2011): An explicit link between Gaussian fields and Gaussian Markov eandom fields: the stochastic partial differential equation approach (with discussion); *JRSS Series B*, 73(4):423–498
- David Bolin and Finn Lindgren (2015): Excursion and contour uncertainty regions for latent Gaussian models, *JRRS Series B*, 77(1):85–106
- David Bolin and Finn Lindgren (2016): Quantifying the uncertainty of contour maps, in review. `http://arxiv.org/abs/1507.01778`
- David Bolin and Finn Lindgren (2013–2016): R CRAN package `excursions`

  ```
  contourmap(mu=expectation, Q=precision)
  contourmap.inla(result.inla)
  continuous(..., geometry)
  ```