

Lecture 9: SPDEs and GMRFs (part 2)

Gaussian Markov random fields

David Bolin
Chalmers University of Technology
February 23, 2015



The SPDE connection

Matérn fields, which have covariance function of the form

$$c(x, y) \propto (\kappa \|x - y\|)^\nu K_\nu(\kappa \|x - y\|),$$

are the stationary solutions to the SPDE

$$(\kappa^2 - \Delta)^{\frac{\nu+d/2}{2}} x(s) = W(s),$$

where

- $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$ is the Laplacian
- $W(s)$ is spatial white noise.
- The parameter ν controls the smoothness.
- The parameter κ controls the range.

We can use the SPDE to obtain a GMRF representation of Gaussian Matérn fields!

Step 1: The 'weak' solution

We want to construct a low-dimensional representation

$$x(\mathbf{s}) = \sum_{i=1}^m \varphi_i(\mathbf{s}) w_i$$

Recall that it only makes sense to talk about a solution to the SPDE in a stochastic weak sense. That is, we require that for every function $\psi(s)$ from some suitable space of test functions,

$$\langle \psi, (\kappa^2 - \Delta)x \rangle_{\Omega} \stackrel{D}{=} \langle \psi, W \rangle_{\Omega}$$

In theory, we typically take $\psi \in H_{\alpha}$.

In practice we construct a FEM approximation by choosing a space of test functions that is easier to handle, such as $\text{span}\{\varphi\}_{i=1}^m$

Step 2: Plug in the basis functions

Replace $x(s)$ with the basis function expansions and chose $\varphi(s)$ to be the set of basis functions

We get the system of linear equations

$$\left\langle \varphi_j, (\kappa^2 - \Delta) \left(\sum_i w_i \varphi_i \right) \right\rangle_{\Omega} \stackrel{D}{=} \langle \varphi_j, W \rangle_{\Omega}$$

for $j = 1, \dots, m$.

The LHS can be written as

$$\begin{aligned} \left\langle \varphi_j, (\kappa^2 - \Delta) \left(\sum_i w_i \varphi_i \right) \right\rangle_{\Omega} &= \sum_i \langle \varphi_j, (\kappa^2 - \Delta) w_i \varphi_i \rangle_{\Omega} \\ &= \sum_i (\kappa^2 \langle \varphi_j, \varphi_i \rangle_{\Omega} + \langle \nabla \varphi_j, \nabla \varphi_i \rangle_{\Omega}) w_i \end{aligned}$$

Step 2: The right-hand side

Remember that integrals of white noise are nice:

$$\int_{\Omega} \varphi(s) dW(s) \sim N(0, \int_{\Omega} \varphi^2(s) ds)$$

Also

$$C \left(\int_{\Omega} \varphi_i(s) dW(s), \int_{\Omega} \varphi_j(s) dW(s) \right) = \int_{\Omega} \varphi_i(s) \varphi_j(s) ds$$

Thus, the right-hand side is $N(0, \mathbf{C})$, where

$$\mathbf{C}_{ij} = \int_{\Omega} \varphi_i(s) \varphi_j(s) ds$$

Putting the pieces together

We have two matrices:

- $\mathbf{C}_{ii} = \int_{\Omega} \varphi_i(s) ds$ (the constant terms)
- $\mathbf{G}_{ij} = \int_{\Omega} \nabla \varphi_i(s) \cdot \nabla \varphi_j(s) ds$ (the Laplacian term)

The (scary) SPDE becomes the (normal) equation

$$(\kappa^2 \mathbf{C} + \mathbf{G}) \mathbf{w} \sim \mathcal{N}(0, \mathbf{C})$$

and therefore \mathbf{w} is a GMRF with precision matrix

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

... so what?

We have now constructed a FEM representation

$$x(\mathbf{s}) = \sum_{i=1}^m \varphi_i(\mathbf{s}) w_i$$

where $\{\varphi\}$ are our basis functions and $\mathbf{w} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$ is a GMRF.

Say now that we observe $x(s)$ at locations $\mathbf{s}_1, \dots, \mathbf{s}_N$ under Gaussian measurement noise: $y_i = x(\mathbf{s}_i) + \varepsilon_i$

Introduce the **observation matrix** \mathbf{A} with elements $A_{ij} = \varphi_j(s_i)$

We can then write the joint model as

$$\mathbf{y} \sim N(\mathbf{Aw}, \sigma^2 \mathbf{I})$$

$$\mathbf{w} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$$

Thus, we can now use our standard GMRF framework for problems on continuous spaces! **This is exactly what we need!!**

The procedure for $\alpha = 1$

$$(\kappa^2 - \Delta)^{\frac{1}{2}} x(s) = W(s),$$

This is more tricky, we have a fractional operator $\mathcal{L}^{\frac{1}{2}} = (\kappa^2 - \Delta)^{\frac{1}{2}}$!

We change our approximation technique and instead use

- $\text{span}\{\mathcal{L}^{\frac{1}{2}}\varphi\}_{i=1}^n$

as test functions in our FEM approximation.

This choice is usually called the least square method.

We get the system of linear equations

$$\left\langle \mathcal{L}^{\frac{1}{2}}\varphi_j, \mathcal{L}^{\frac{1}{2}} \left(\sum_i w_i \varphi_i \right) \right\rangle_{\Omega} \stackrel{D}{=} \left\langle \mathcal{L}^{\frac{1}{2}}\varphi_j, W \right\rangle_{\Omega}$$

Interlude: More Sobolev spaces!

The Sobolev space $H_1(\kappa)$

Let $\kappa \geq 0$ and define $H_1(\kappa)$ as the space of functions $\varphi \in H_1$, equipped with the inner product

$$\langle f, g \rangle_{H_1(\kappa)} = \kappa^2 \langle f, g \rangle + \langle \nabla f, \nabla g \rangle$$

- $H_1(\kappa)$ depends on Ω , so $H_1(\Omega, \kappa)$ is a more cumbersome but correct notation.
- Note that the inner product induces a norm for $\kappa > 0$ and a semi-norm for $\kappa = 0$.
- For $\kappa > 0$, the norms are equivalent, and $H_1(1) = H_1$.
- For $H_1(0)$ is a space of equivalence classes of functions that can be identified by functions with $\langle f, 1 \rangle = 0$
- (this looks familiar right?).

A technical lemma

Lemma

For $f, g \in H_1(\kappa)$, we have

$$\left\langle \mathcal{L}^{1/2}f, \mathcal{L}^{1/2}g \right\rangle = \langle f, g \rangle_{H_1(\kappa)}$$

whenever

- $\Omega = \mathbb{R}^d$
- Ω is closed or
- Ω is compact and $\langle f, \partial_n g \rangle = \langle \partial_n f, g \rangle = 0$.

As the stochastic Green's identity, this lemma naturally generalizes to random fields with L^2 -bounded gradients

The procedure for $\alpha = 1$

$$\left\langle \mathcal{L}^{\frac{1}{2}}\varphi_j, \mathcal{L}^{\frac{1}{2}}\left(\sum_i w_i \varphi_i\right) \right\rangle_{\Omega} \stackrel{D}{=} \left\langle \mathcal{L}^{\frac{1}{2}}\varphi_j, W \right\rangle_{\Omega}$$

For $\varphi_i \in H_1(\kappa)$, we use the lemma and obtain

$$\sum_i \langle \varphi_j, \mathcal{L}\varphi_i \rangle_{\Omega} w_i \stackrel{D}{=} \left\langle \mathcal{L}^{\frac{1}{2}}\varphi_j, W \right\rangle_{\Omega}$$

Thus, the LHS is again $(\mathbf{G} + \kappa^2 \mathbf{C})\mathbf{w}$.

Using the lemma when calculating covariances of elements in the RHS, we see that the RHS is $N(0, \mathbf{G} + \kappa^2 \mathbf{C})$.

Hence, we have transformed this SPDE to the equation

$$\mathbf{w} \sim N(0, (\mathbf{G} + \kappa^2 \mathbf{C})^{-1})$$

The de Wij connection

$$\mathbf{w} \sim N(0, (\mathbf{G} + \kappa^2 \mathbf{C})^{-1})$$

Let $\Omega = \mathbb{R}^2$ and take basis functions from a triangulation of a regular lattice.

With $\kappa = 0$, we see that \mathbf{w} is equivalent to the ICAR(1) model.

This shows that the ICAR model, seen as a Hilbert space representation with linear basis functions, converges to the de Wij process, which can be represented as the solution to the SPDE when $\alpha = 1$ and $\kappa = 0$.

The intrinsic properties can be found by looking at the null-space of the operator, which in case of the half-Laplacian consists of constant functions.

The procedure higher $\alpha \in \mathbb{N}$

Note that

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} x(s) = W(s),$$

for odd values of α can be written as

$$(\kappa^2 - \Delta)^{\frac{\alpha-1}{2}} x(s) = z(s),$$

where

$$(\kappa^2 - \Delta)^{\frac{1}{2}} z(s) = W(s).$$

For even values of α , we can instead write

$$(\kappa^2 - \Delta)^{\frac{\alpha-2}{2}} x(s) = z(s),$$

where

$$(\kappa^2 - \Delta) z(s) = W(s),$$

The result for general $\alpha \in \mathbb{N}$

We can now combine

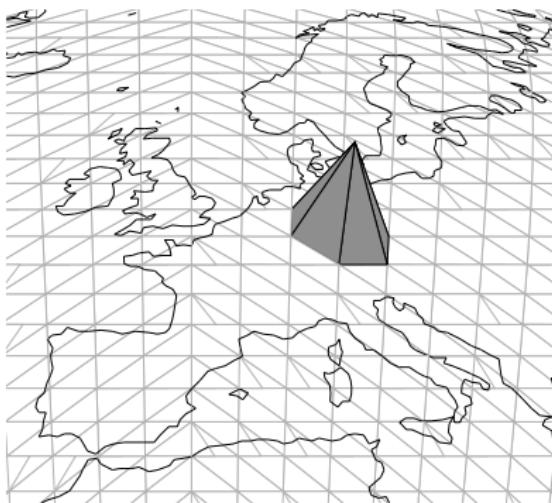
- ① The recursive formulation of the equation, and
 - ② The results for the two fundamental cases $\alpha = 1$ and $\alpha = 2$,
- to get the following general result

Matérn precisions

Let $\mathbf{K} = \mathbf{G} + \kappa^2 \mathbf{C}$, then the finite element representation of the Matérn SPDE with smoothness $\alpha = \nu + d/2$ has precision matrix \mathbf{Q}_α , where:

- $\mathbf{Q}_1 = \mathbf{K}$
- $\mathbf{Q}_2 = \mathbf{K}\mathbf{C}^{-1}\mathbf{K}$
- $\mathbf{Q}_\alpha = \mathbf{K}\mathbf{C}^{-1}\mathbf{Q}_{\alpha-2}\mathbf{C}^{-1}\mathbf{K}$ for $\alpha = 3, 4, \dots$

Choosing the basis functions

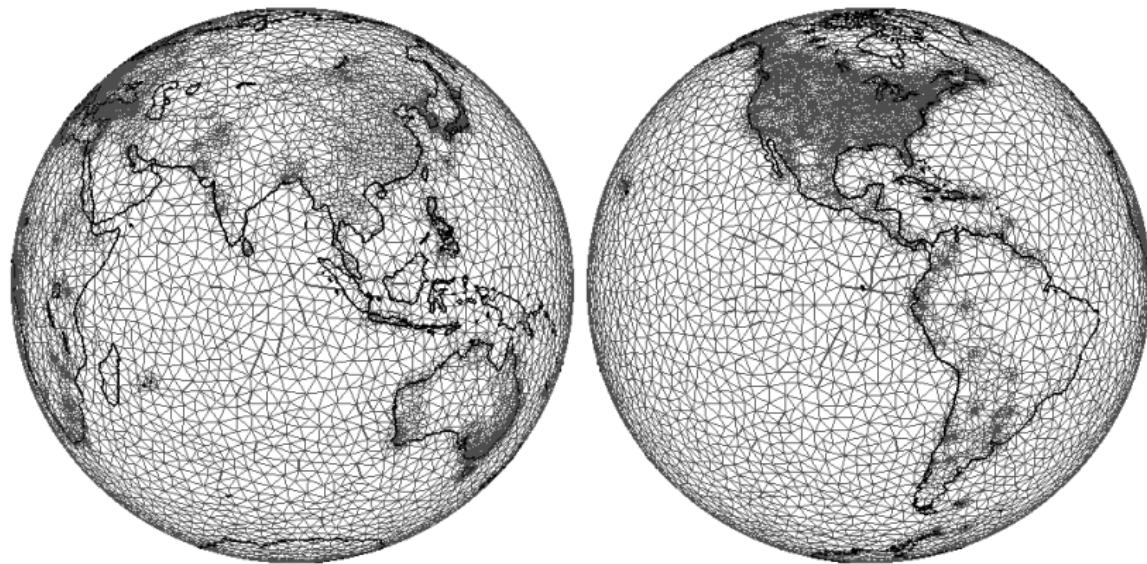


- Standard choice: piecewise linear functions.
- Given by triangulating the domain of interest.
- With this choice, the matrices \mathbf{G} and \mathbf{C} are very easy to compute.

Choosing the basis functions



Choosing the basis functions



Note that we can reduce computational cost by having coarser triangulations in areas where we do not have data.

Weak convergence

It might be of interest to specify the approximation properties of the procedure.

For completeness, we now briefly look at some asymptotic results.

Weak convergence

A sequence of $L^2(\Omega)$ -bounded generalized GF $\{x_n\}$ is said to converge weakly to an $L^2(\Omega)$ -bounded generalized GF x , if for any $f, g \in L^2(\Omega)$,

$$\mathbb{E}(\langle f, x_n \rangle_{\Omega}) \rightarrow \mathbb{E}(\langle f, x \rangle_{\Omega})$$

$$\text{Cov}(\langle f, x_n \rangle_{\Omega}, \langle g, x_n \rangle_{\Omega}) \rightarrow \text{Cov}(\langle f, x \rangle_{\Omega}, \langle g, x \rangle_{\Omega})$$

as $n \rightarrow \infty$. Denote this $x_n \xrightarrow{D(L^2(\Omega))} x$.

FEM convergence

Let $\Phi_n = \{\varphi_1, \dots, \varphi_n\}$ denote a set of piecewise linear basis functions $\varphi_i \in H_1(\kappa)$, induced by some triangulation of the domain, and let $H_1^n(\kappa) = \text{span}\Phi_n$.

Furthermore, let x_n be the corresponding FEM approximation of the SPDE, which we call a $H_1^n(\kappa)$ approximation.

We now want to show that by increasing the number of basis functions, and letting the maximal edge of the triangles go to zero, we get convergence.

Dense subspace sequences

We say that a sequence of subspaces $\{H_1^n\}$ is dense in H_1 if for every $f \in H_1$, there is a sequence $\{f_n\}$, $f_n \in H_1^n$, such that $\lim_{n \rightarrow \infty} \|f - f_n\|_{H_1(\kappa)} = 0$

FEM convergence

FEM convergence

Let x be a weak solution to the Gaussian Matérn SPDE with Neumann boundary conditions and let x_n be a weak $H_1^n(\kappa)$ approximation, we then have

$$\begin{aligned}x_n &\xrightarrow{D(L^2(\Omega))} x \\ \mathcal{L}^{\alpha/2} x_n &\xrightarrow{D(L^2(\Omega))} \mathcal{L}^{\alpha/2} x\end{aligned}$$

if the sequence $\{H_1^n(\kappa)\}$ is dense in $H_1(\kappa)$ and either

- ① $\alpha = 2$ and x_n is the Galerkin solution, or
- ② $\alpha = 1$ and x_n is the least squares solution.

Having this, we can also easily show convergence of the iterated FEM approximations for higher order α .

The rate of convergence

Using standard results from finite element analysis, we can also obtain a bound for the error for a specific set of basis functions.

For example, with $\alpha = 2$ we have

$$\sup_{f \in H_1 : \|f\|_{H_1} \leq 1} E(\langle f, x_n - x \rangle_{H_1}^2) \leq ch^2$$

here

- x_n is the H_1^n approximation of the solution
- x is the true solution
- h is the diameter of the largest circle that can be inscribed in a triangle in the triangulation.
- c is some constant.

We have a GMRF connection, almost...

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

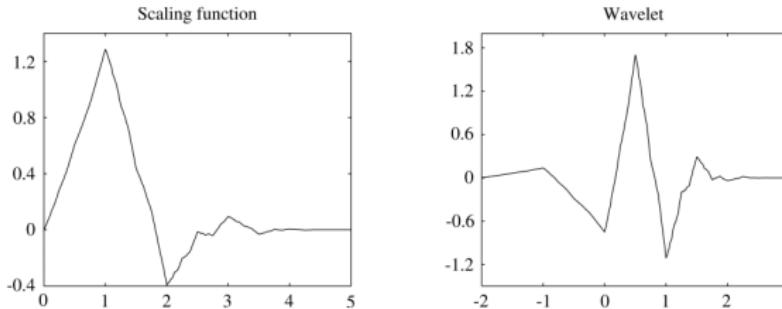
Note that \mathbf{C}^{-1} in general is a dense matrix!

Thus, the procedure does not give us a GMRF representation for the standard piecewise linear basis functions.

We can do two things:

- ① Change to a set of basis functions for which \mathbf{C}^{-1} is sparse.
- ② Use the piecewise linear basis functions but approximate \mathbf{C}^{-1} with a sparse matrix

Choice 1: Use basis functions so that \mathbf{C}^{-1} is sparse.



- Use a family of orthonormal basis functions.
- Need compact support of these functions
- We also want functions for which the basis function matrices are easy to evaluate.
- A good choice is therefore a wavelet basis such a Daubechies wavelets
- See Bolin and Lindgren (2013) for details.

Choice 2: Markov approximations

- The simplest choice is to use a diagonal approximation.
- Do this by replacing \mathbf{C} with a diagonal matrix $\tilde{\mathbf{C}}$ with elements

$$\tilde{C}_{ii} = \int_{\Omega} \varphi(s) ds = \sum_j C_{ij}$$

- This is sometimes referred to as “mass lumping”.
- The convergence rate of the FEM approximation is preserved.
- For a fixed triangulation, the covariance error is bounded by ch^2 , where c is a constant and h is the diameter of the largest circle that can be inscribed in a triangle of the triangulation.

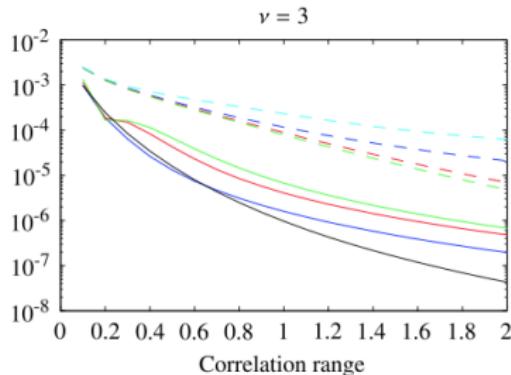
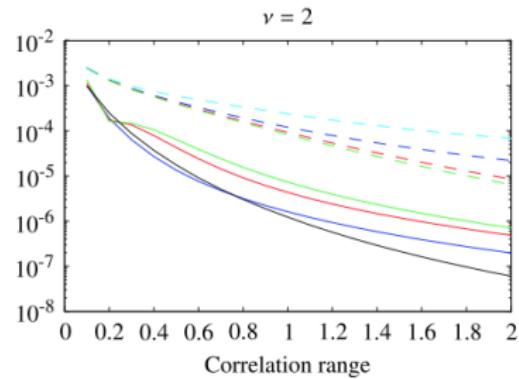
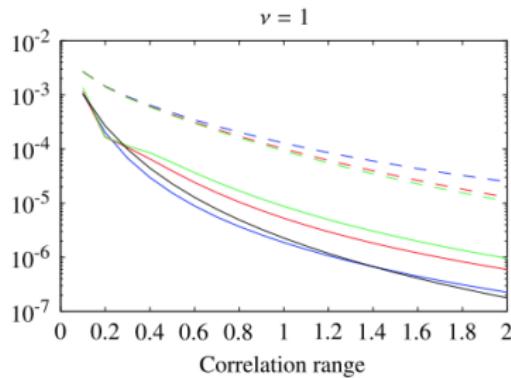
So which option is best?

- Using Daubechies wavelets, we obtain a GMRF representation without any further approximations.
- The disadvantage is that the wavelets have larger support than the piecewise linear basis functions.
- Thus, we get a GMRF with more non-zero elements in the precision matrix.
- For a piecewise linear basis, we can therefore use more basis functions for the same computational cost (approximately a factor of six to be precise).

We do a simple comparison of the errors for two choices

- Compare the L^2 -norms of the covariance errors for different covariance ranges
- The number of basis functions chosen to give similar computational cost for kriging prediction.

Results



- - - 1st order spline
- - - 2nd order spline
- - - 3rd order spline
- - - Process convolution
- - Approx 1st order spline
- - Approx 2nd order spline
- - Approx 3rd order spline
- Daubechies 3

The posterior Markov property (I)

$$x(\mathbf{s}) = \sum_{i=1}^m \varphi_i(\mathbf{s}) w_i$$

If we have basis functions with compact support, the SPDE representation gives us $\mathbf{w} \sim N(0, \mathbf{Q}^{-1})$ for a sparse precision matrix.

However, some other popular basis representation methods give us $\mathbf{w} \sim N(0, \mathbf{D}^{-1})$ for a diagonal matrix \mathbf{D} !

- the convolution kernel methods
- Karhunen-Loève expansions (if we actually could compute them for any interesting models...)
- Spectral representations

What's the catch with these methods?

The posterior Markov property (I)

If we only simulate “from the prior” they can work great.

However, for parameter estimation or simulation conditionally on data, it is not enough that \mathbf{Q} is sparse, we also need that the posterior precision matrix is sparse:

$$\mathbf{Q}_{X|Y} = \mathbf{Q} + \mathbf{A}^T \mathbf{Q}_\varepsilon \mathbf{A}$$

where $A_{ij} = \varphi_i(\mathbf{s}_j)$.

- For our FEM basis, $\mathbf{Q}_{X|Y}$ is as sparse as \mathbf{Q} , since all observations are “local” with respect to the basis.
- For the fourier basis, \mathbf{A} and hence $\mathbf{Q}_{X|Y}$ are *completely dense*!
- Thus, computations typically scale as $O(m^{3/2})$ for our FEM method, whereas we get $O(m^3)$ for the other methods.
- If we use m basis functions for the SPDE method, we can only use \sqrt{m} for the other methods!

So how good is this really?

We now have asymptotic properties of our FEM representation, but how does this perform in practise?

We will compare the FEM representations of a Matérn field on \mathbb{R}^2 using

- Piecewise linear basis functions (first order B-splines)
 - Daubechies wavelets (DB3 to be precise)
- ... with two other popular methods for “solving” the Big N problem:

- The kernel method
- Covariance tapering

Tapering is based on introducing sparsity in the *covariance matrix* by replacing the covariance function $r(\mathbf{h})$ with a tapered version $r_{tap}(\mathbf{h}) = r_\theta(\mathbf{h})r(\mathbf{h})$, where $r_\theta(\mathbf{h})$ has compact support.

Setup for comparison

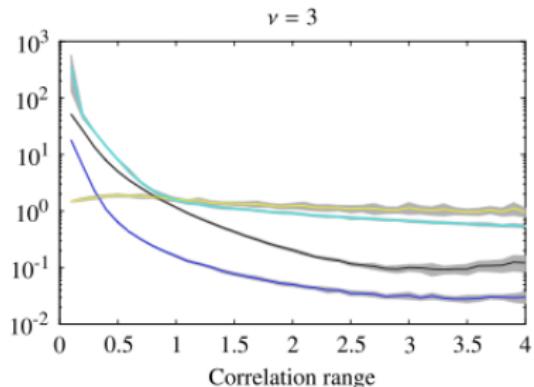
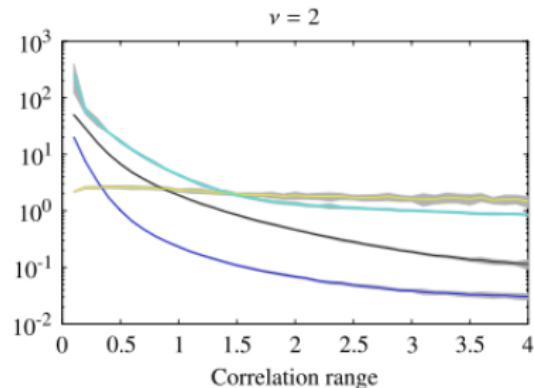
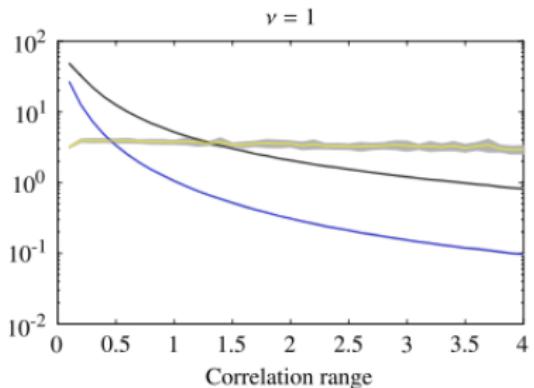
Sample a Gaussian Matérn field $x(\mathbf{s})$ at 5000 locations chosen uniformly distributed on $[0, 5] \times [0, 5]$, under Gaussian measurement noise.

Compute the kriging prediction of $x(\mathbf{s})$ to all locations in a 70×70 regular lattice in the domain.

Compute the approximate predictions using the FEM representations, the kernel approach, and the tapering approximation, and compare the results to the optimal prediction.

The number of basis functions and the tapering range are chosen to give equivalent computational costs for the predictions.

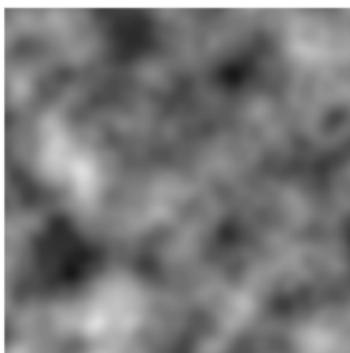
Results



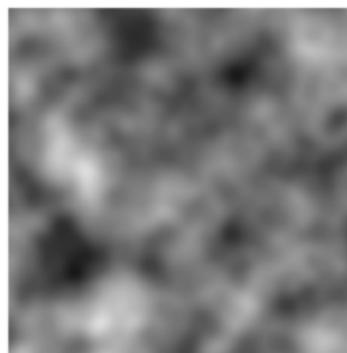
— S1
— DB3
— Process convolution
— Tapering

Example: approximations for fixed computational cost

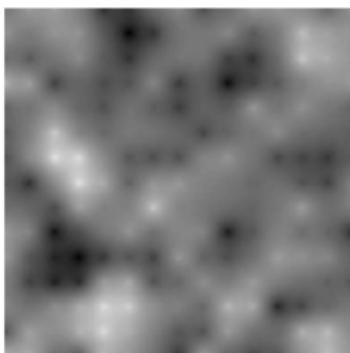
Optimal prediction



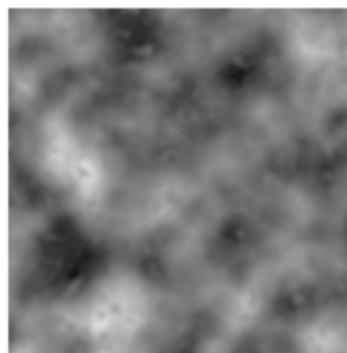
SI basis



Convolution basis



Tapered covariance



Extensions

Key lesson (Simpson, Lindgren, Rue, 2011)

In order to make spatial statistics computationally feasible, we need to forget about the covariance function

The SPDE method allows us to do this...

Also, now when we have let go of covariance-based modelling, the really nice advantages occurs: Simple extensions to

- Models on manifolds
- Non-stationary models
- More general SPDE models
- Non-Gaussian models
- Multivariate models
- +++

Manifolds/curved-spaces

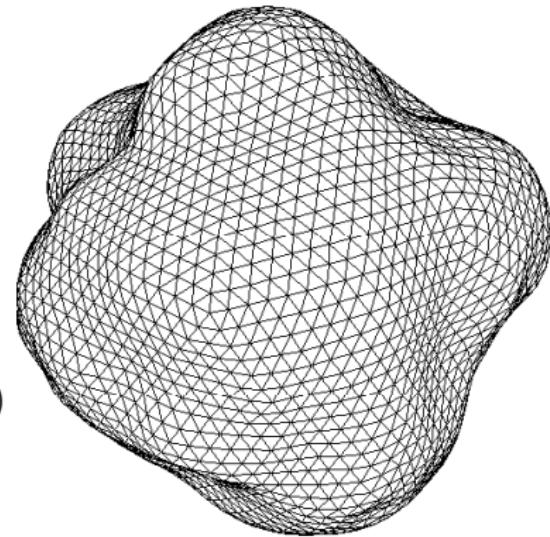
What if our data is not collected on a flat surface?

Define Matérn fields using

$$(\kappa^2 - \Delta)^{\alpha/2} \mathbf{x}(s) = \mathbf{W}(s)$$

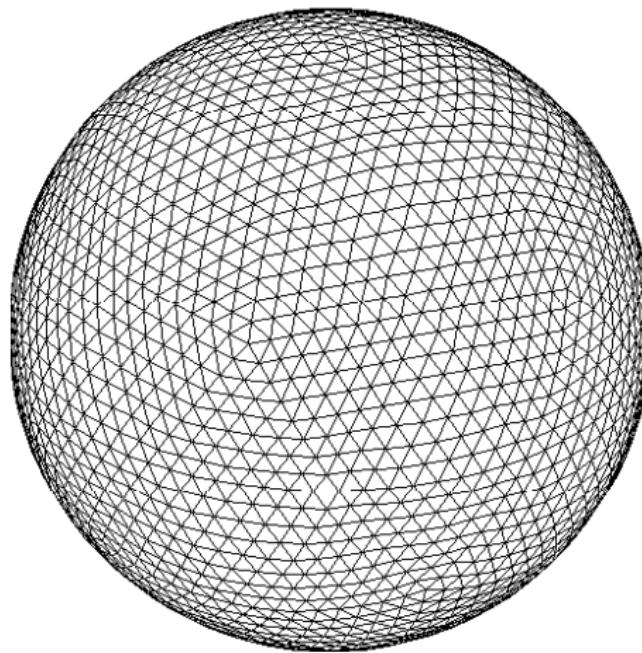
on the manifold \mathcal{S} , driven by Gaussian “white noise” on \mathcal{S}

$$\text{Cov}(\mathbf{W}(A_i), \mathbf{W}(A_j)) = \int_{A_i \cap A_j} d\mathcal{S}(s)$$

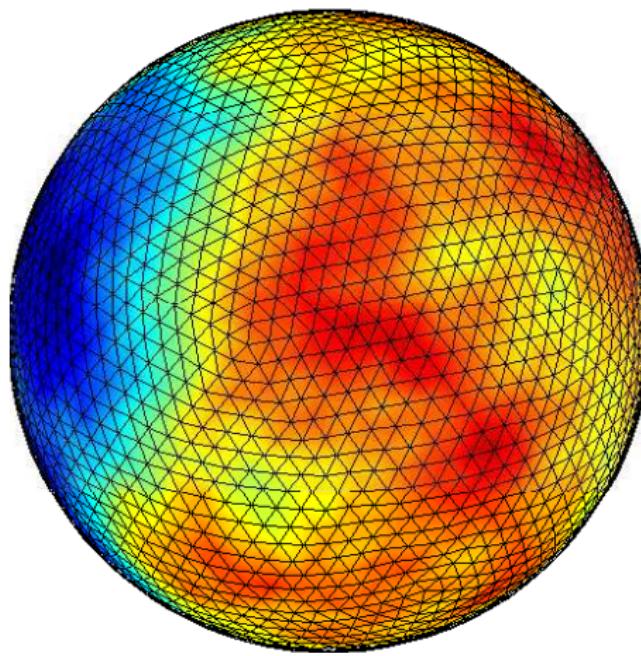


Everything stays the same

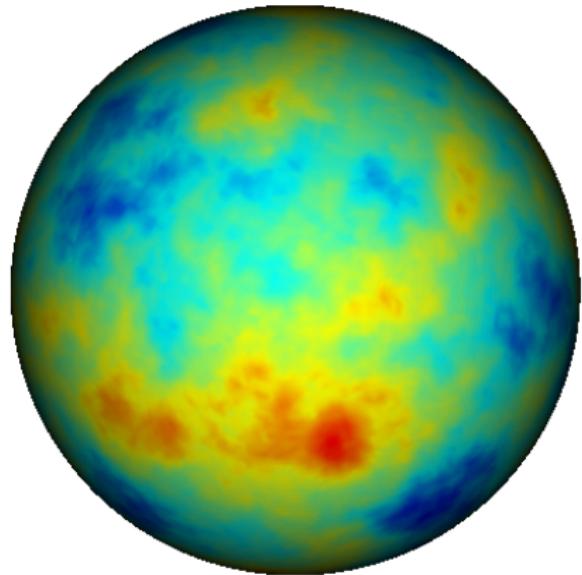
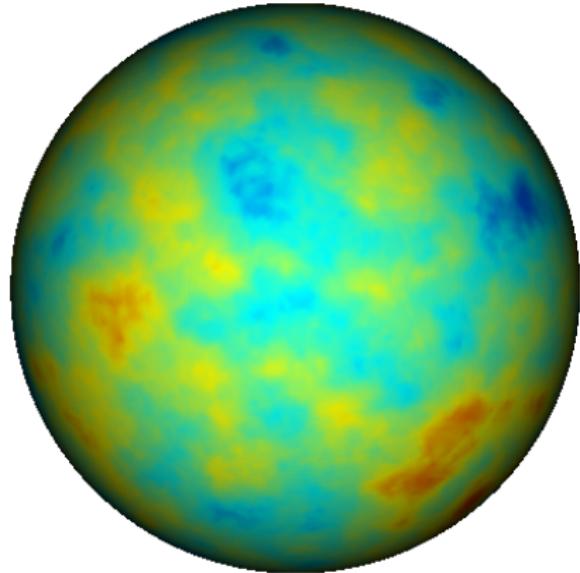
Example: Sphere



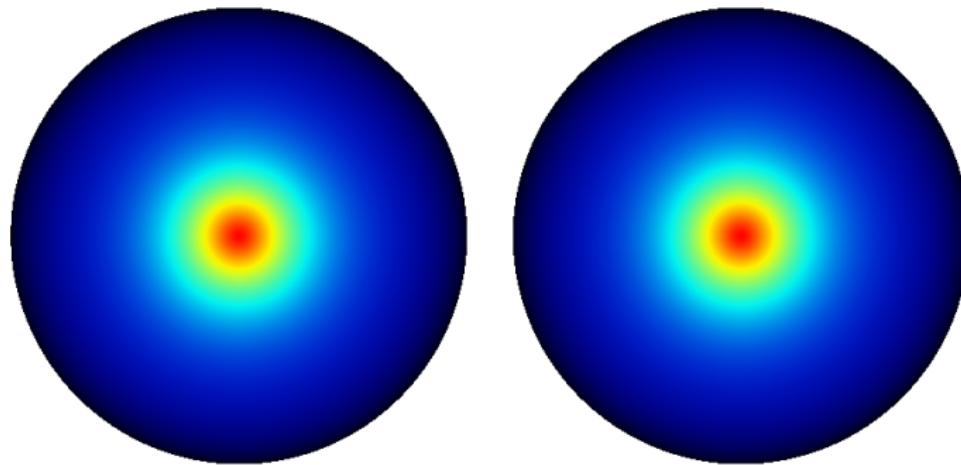
Example: Sphere



Example: Sphere



Example: Covariances on sphere



Oscillating Matérn

Take two independent white noise processes $\mathcal{W}_1(\mathbf{u})$ and $\mathcal{W}_2(\mathbf{u})$ and define $\mathcal{W}(\mathbf{u}) = \mathcal{W}_1(\mathbf{u}) + i\mathcal{W}_2(\mathbf{u})$.

Replace κ^2 with $\kappa^2 e^{i\pi\theta}$, where $0 \leq \theta < 1$.

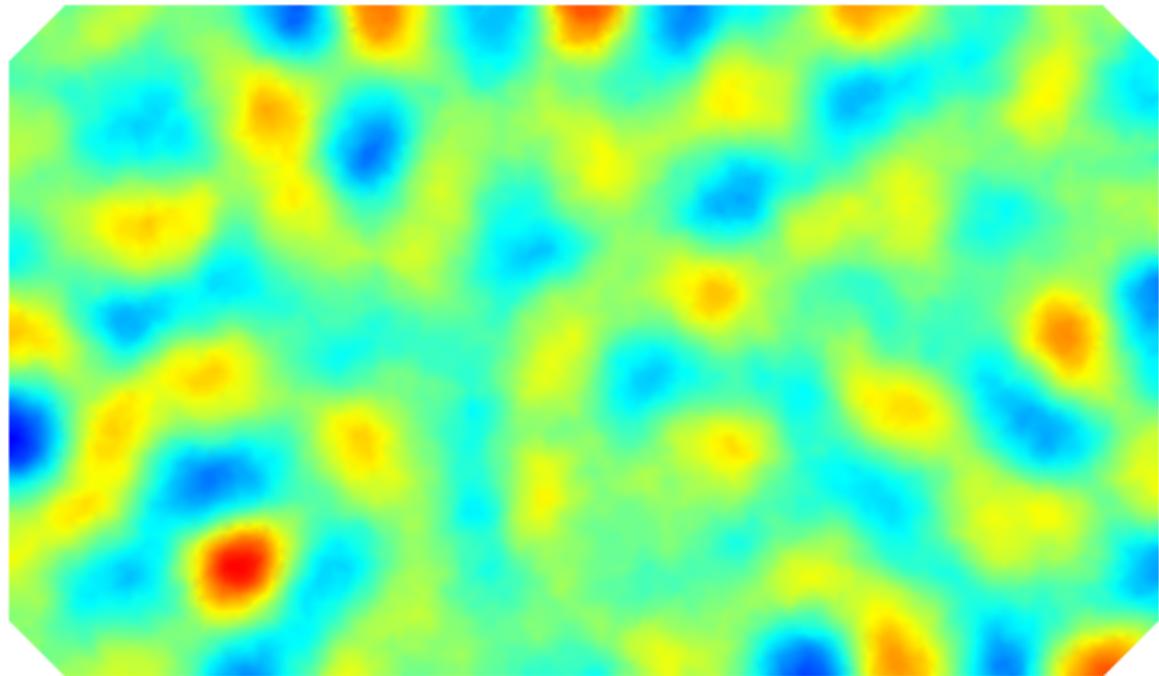
$$(\kappa^2 e^{i\pi\theta} - \Delta)(\tau x(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \Omega$$

The real and imaginary parts of the solution x are now independent random fields, with spectral densities

$$S(\mathbf{k}) = \frac{1}{(2\pi)^d} \frac{1}{\|k\|^4 + 2\cos(\pi\theta)\kappa^2\|k\|^2 + \kappa^4}$$

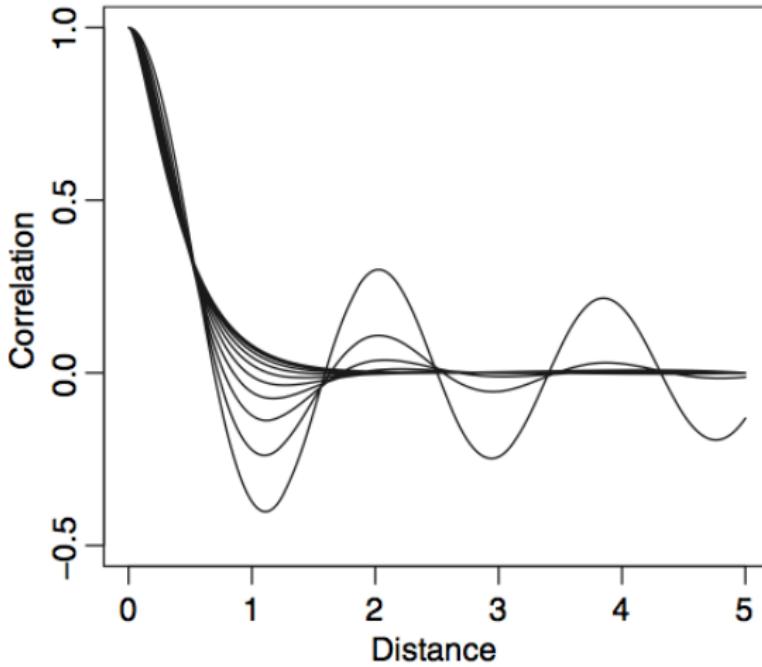
Resulting FEM precision:

$$\mathbf{Q} = \kappa^4 \mathbf{C} + 2\kappa^2 \cos(\pi\theta) \mathbf{G} + \mathbf{G} \mathbf{C}^{-1} \mathbf{G}$$

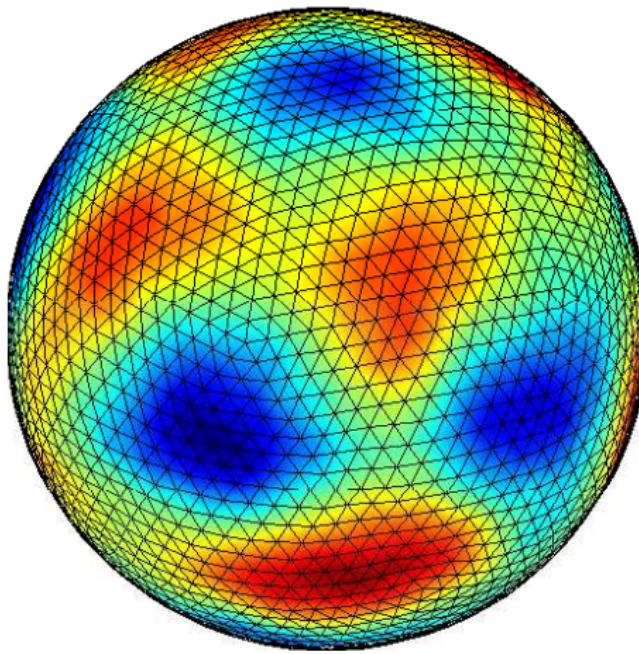
Simulation on \mathbb{R}^2 

Induced covariance function for $\Omega = \mathbb{R}^2$

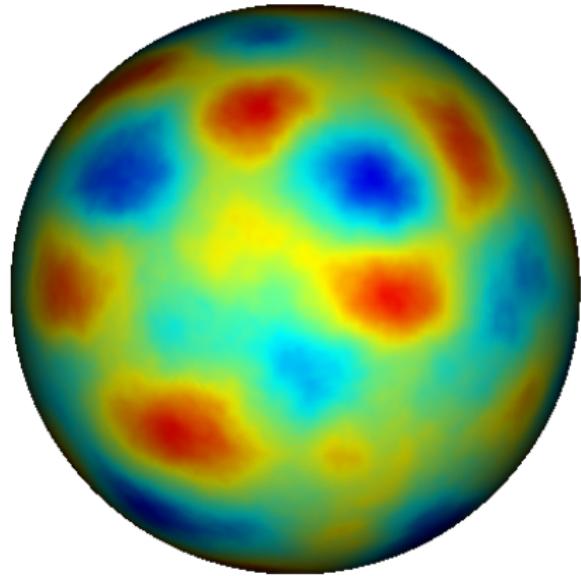
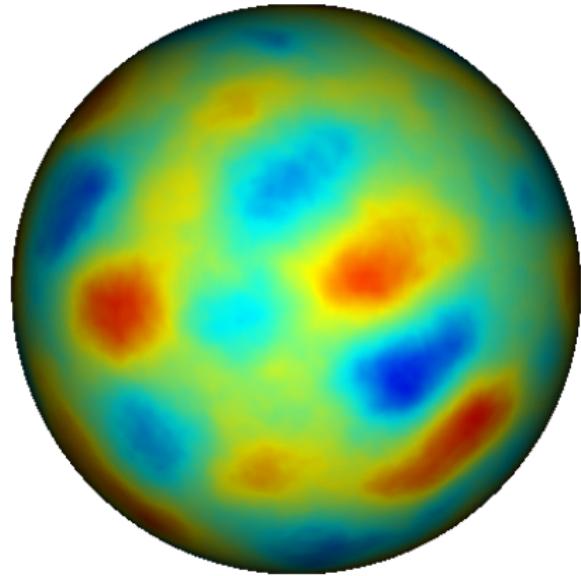
$$C(\mathbf{h}) = \frac{1}{4\pi \sin(\pi\theta)\kappa^2 i} \left[K_0 \left(\kappa \mathbf{h} \exp \left(-\frac{i\pi\theta}{2} \right) \right) - K_0 \left(\kappa \mathbf{h} \exp \left(\frac{i\pi\theta}{2} \right) \right) \right]$$



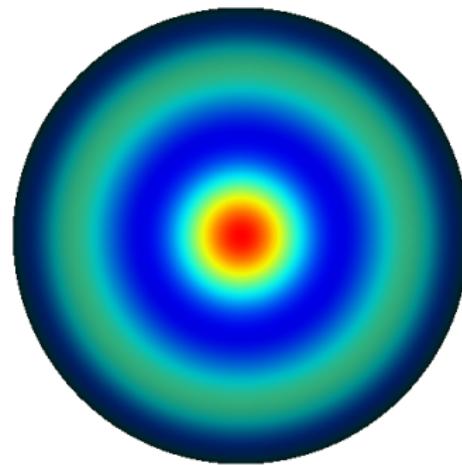
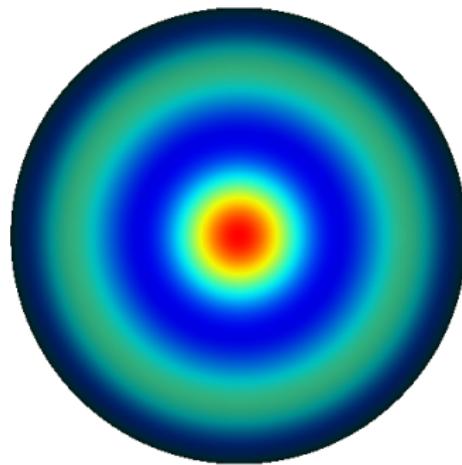
Example: $\kappa = -30 + 4i$, $\alpha = 2$



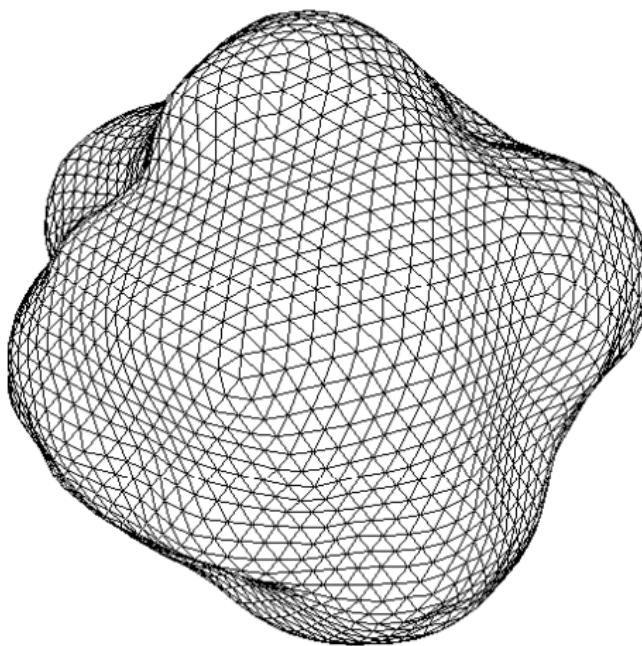
Example: Sphere



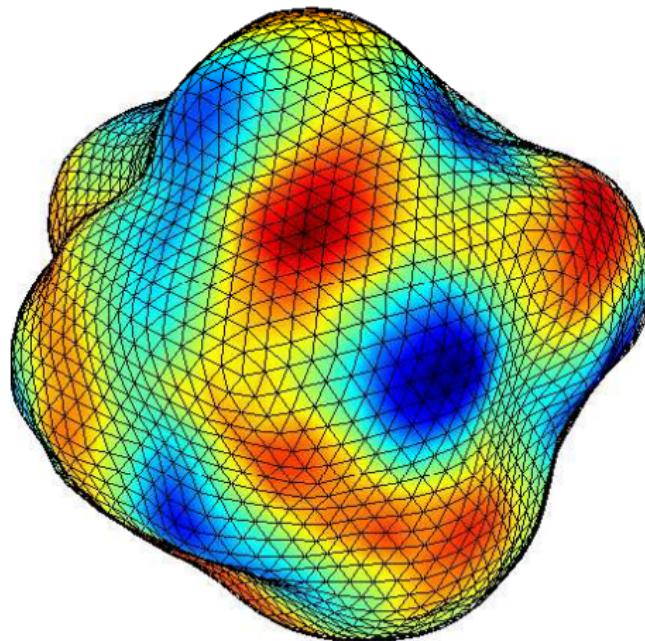
Example: Covariances on sphere



Example: Potato



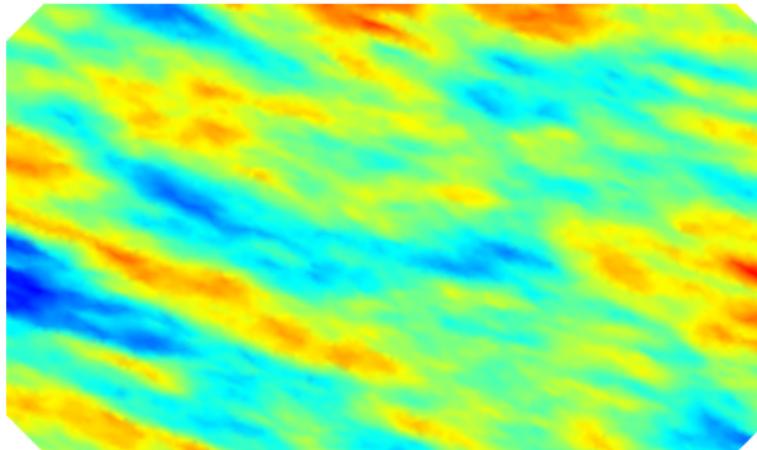
Example: Potato



Anisotropic Matérn

Another extension we could do is to introduce an anisotropic Laplacian:

$$(\kappa^2 - \nabla \cdot \mathbf{M} \nabla)^{\alpha/2}(\tau x(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^d$$



Nested SPDE models

We can obtain many flexible oscillating covariance models by considering nested SPDEs:

$$\left(\prod_{i=1}^{n_1} (\kappa_i^2 - \Delta)^{\frac{\alpha_i}{2}} \right) X(s) = \left(\prod_{i=1}^{n_2} (b_i + \mathbf{B}_i^\top \nabla) \right) \mathcal{W}(s)$$

Since the two operators are commutative, this is equivalent to

$$\begin{aligned} & \left(\prod_{i=1}^{n_1} (\kappa_i^2 - \Delta)^{\frac{\alpha_i}{2}} \right) X_0(s) = \mathcal{W}(s), \\ & X(s) = \left(\prod_{i=1}^{n_2} (b_i + \mathbf{B}_i^\top \nabla) \right) X_0(s). \end{aligned}$$

Differentiations in several directions \mathbf{B}_i .

If the solution to this system, X , should be at least as “well behaved” as white noise, one must have $\sum_{i=1}^{n_1} \alpha_i \geq n_2$.

Properties in \mathbb{R}^d

- The spectral density for X is given by

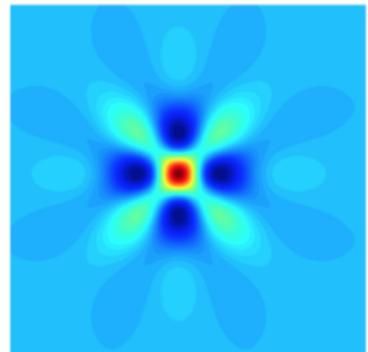
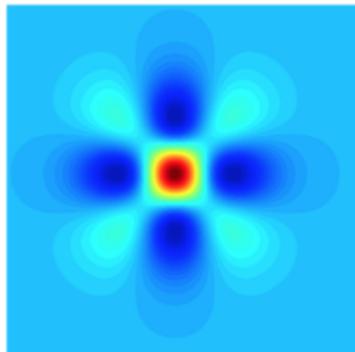
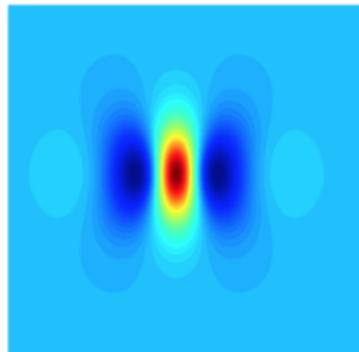
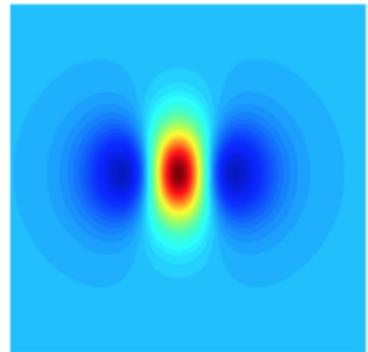
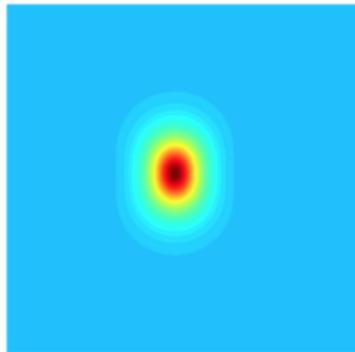
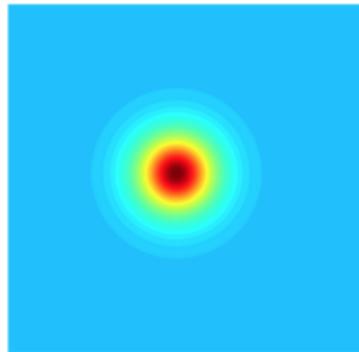
$$S(\mathbf{k}) = \frac{\phi^2}{(2\pi)^d} \frac{\prod_{j=1}^{n_2} (b_j^2 + \mathbf{k}^\top \mathbf{B}_j \mathbf{B}_j^\top \mathbf{k})}{\prod_{j=1}^{n_1} (\kappa_j^2 + \|\mathbf{k}\|^2)^{\alpha_j}}$$

- X has almost surely continuous sample functions if
 $2 \sum_{i=1}^{n_1} \alpha_i - 2n_2 > d.$
- The m :th order directional derivative of X has almost surely continuous sample functions given that
 $2 \sum_{i=1}^{n_1} \alpha_i - 2n_2 - d > m.$
- We have analytic expressions for the induced covariance function, e.g. for the simplest nested model:

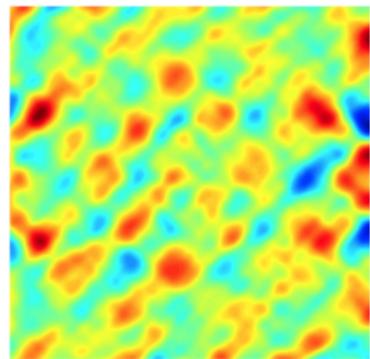
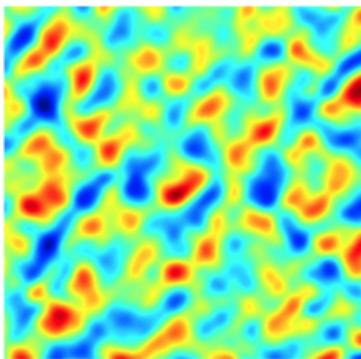
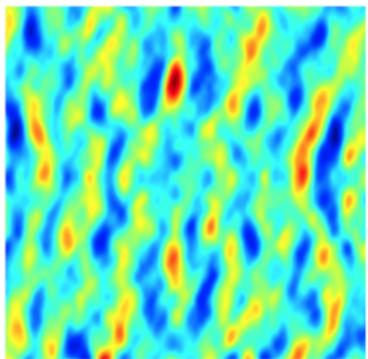
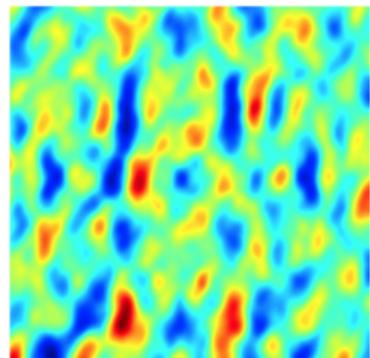
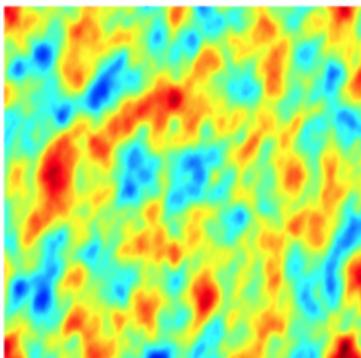
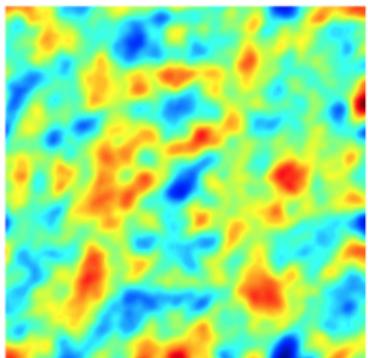
$$C(\mathbf{h}) = b C_\kappa^\nu(\mathbf{h}) + \frac{\mathbf{B}^\top \mathbf{B}}{2\nu} C_\kappa^{\nu-1}(\mathbf{h}) - \frac{\mathbf{h}^\top \mathbf{B} \mathbf{B}^\top \mathbf{h}}{4\nu(\nu-1)} C_\kappa^{\nu-2}(\mathbf{h}).$$

where C_κ^ν is the Matérn correlation function

Induced covariance functions



Corresponding realizations



Non-stationary Models

Perhaps the most surprising and useful extension

We can introduce non-stationary models by allowing the parameters to vary spatially:

$$(\kappa(\mathbf{u})^2 - \Delta)^{\alpha/2}(\tau(\mathbf{u})x(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^d$$

Of particular importance is the case when we let the parameters be slowly varying using, for example, low-dimensional representations

$$\log \kappa(s) = \sum_{i=1}^k B_i^\kappa \theta_i^\kappa, \quad \text{and} \quad \log \tau(s) = \sum_{i=1}^k B_i^\tau \theta_i^\tau$$

The local interpretation of the parameters are highly desirable from a modeling perspective

We can now define non-stationary random fields on manifolds

Non-stationary Matérn

- The Markov properties are preserved
- The actual form of the resulting covariance function is unknown (and we don't care about it)
- No increased computational cost if we can assume that the parameters are constant within the support of the basis functions:

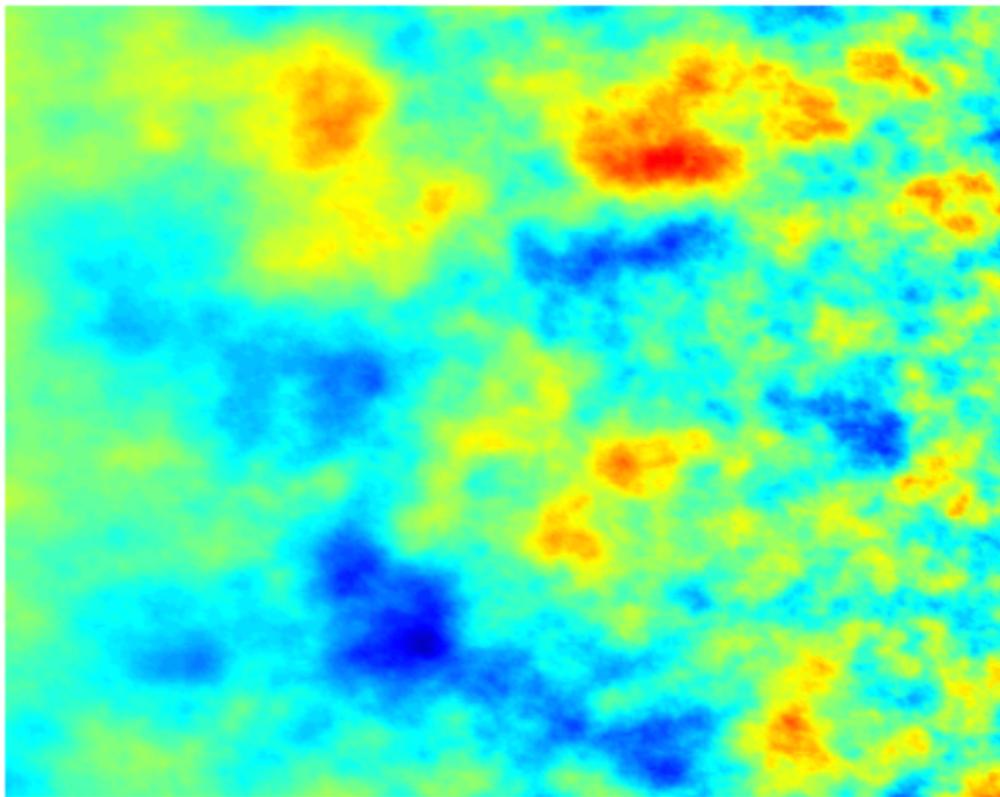
$$\langle \varphi_i, \kappa^2 \varphi_j \rangle = \int \varphi_i(\mathbf{s}) \varphi_j(\mathbf{s}) \kappa^2(\mathbf{s}) d\mathbf{s} \approx C_{ij} \kappa^2(\mathbf{s}^*)$$

- $\mathcal{O}(1)$ increased cost for locally planar parameters.
- Resulting precision matrix:

$$\mathbf{Q} = \text{diag}(\boldsymbol{\tau})(\mathbf{G} + \text{diag}(\kappa)\mathbf{C})^\top \mathbf{C}^{-1} (\mathbf{G} + \text{diag}(\kappa)\mathbf{C}) \text{diag}(\boldsymbol{\tau})$$

where $\kappa_i = \kappa^2(\mathbf{s}^*)$ and $\tau_i = \tau(\mathbf{s}^*)$

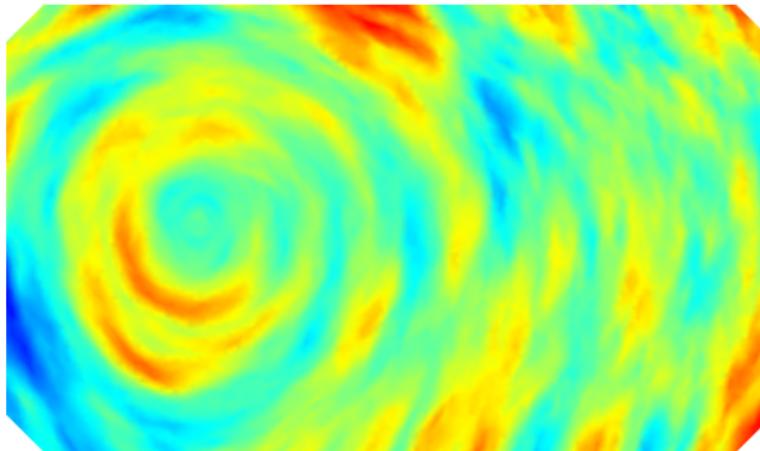
Example



Further examples (I)

We can now combine these different extensions in various ways, an anisotropic non-stationary field is given by

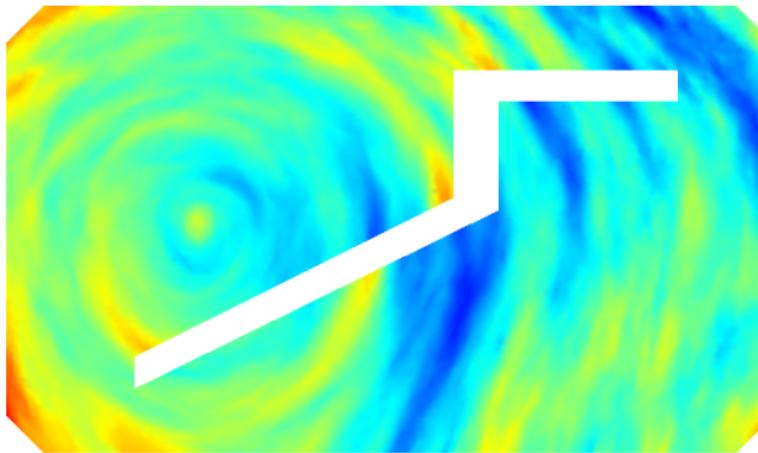
$$(\kappa_{\mathbf{u}}^2 + \nabla \cdot \mathbf{m}_{\mathbf{u}} - \nabla \cdot \mathbf{M}_{\mathbf{u}} \nabla)^{\alpha/2} (\tau_{\mathbf{u}} x(\mathbf{u})) = \mathcal{W}(\mathbf{u})$$



Further examples (II)

We can now combine these different extensions in various ways, an anisotropic non-stationary field is given by

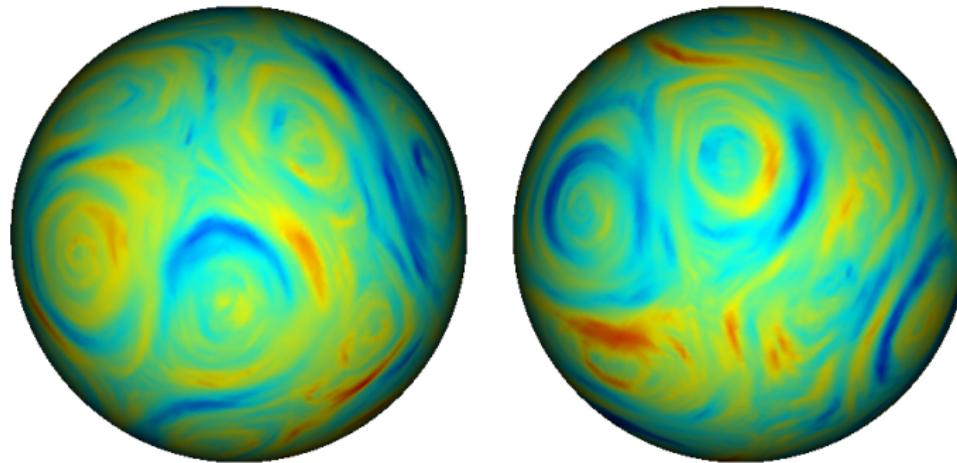
$$(\kappa_{\mathbf{u}}^2 + \nabla \cdot \mathbf{m}_{\mathbf{u}} - \nabla \cdot \mathbf{M}_{\mathbf{u}} \nabla)^{\alpha/2} (\tau_{\mathbf{u}} x(\mathbf{u})) = \mathcal{W}(\mathbf{u})$$



Further examples (III)

And finally, just because we can: An anisotropic non-stationary non-separable spatio-temporal model on a manifold...

$$\left(\frac{\partial}{\partial t} + \kappa_{\mathbf{u},t}^2 + \nabla \cdot \mathbf{m}_{\mathbf{u},t} - \nabla \cdot \mathbf{M}_{\mathbf{u},t} \nabla \right) (\tau_{\mathbf{u},t} x(\mathbf{u}, t)) = \mathsf{E}(\mathbf{u}, t), \quad (\mathbf{u}, t) \in \Omega \times \mathbb{R}$$



Tomorrow

We will start with Project 2.

The project will be about the SPDE approach and how we can combine this with the INLA approach to statistical inference.

Part 1: Implement the non-stationary Matérn SPDE model in one dimension, and then use it to analyse simple time series data.

Part 2: Do a full spatial analysis with the models we have discussed, using the INLA package. Here you can analyse

- ① a spatial dataset from your own research, or
- ② a dataset of US temperatures.

Next lecture

The plan for next time is

- to show a few applications
- present one or more related topics.
- give some comments about the first project
- perhaps give a practical intro to R-INLA (for project 2)

If anything on the following list sounds particularly interesting, let me know and I can make sure to spend time on that:

- ① Multivariate SPDE models
- ② Spatio-temporal SPDE models
- ③ Non-Gaussian SPDE models
- ④ Excursions and computations of joint probabilities for GMRFs
- ⑤ Log-Gaussian Cox processes and GMRFs