Lecture 8: A connection between GFs and GMRFs

Gaussian Markov random fields

David Bolin
Chalmers University of Technology
February 17, 2015

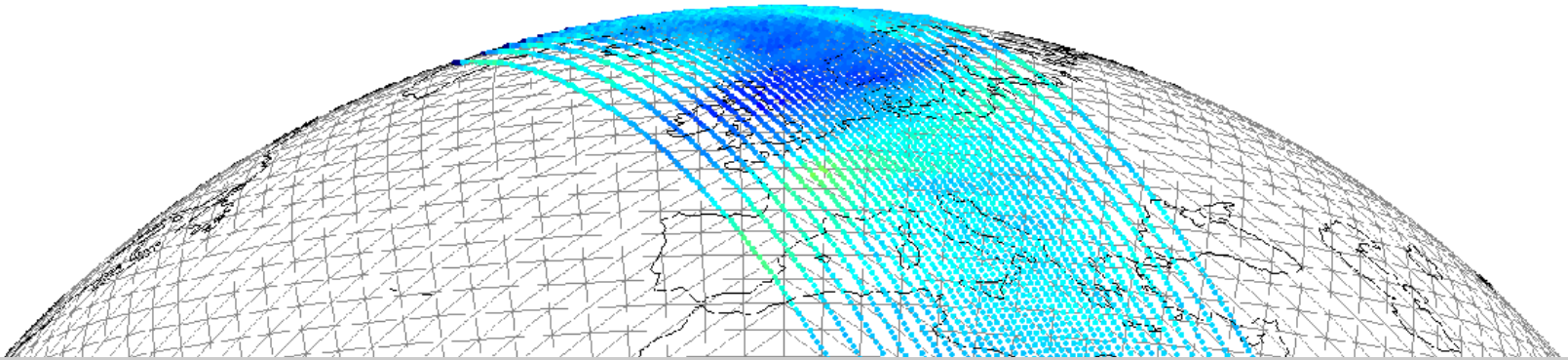
Modeling spatial data: What have we learned so far?

Modelling directly with covariance-based GRFs is too hard:

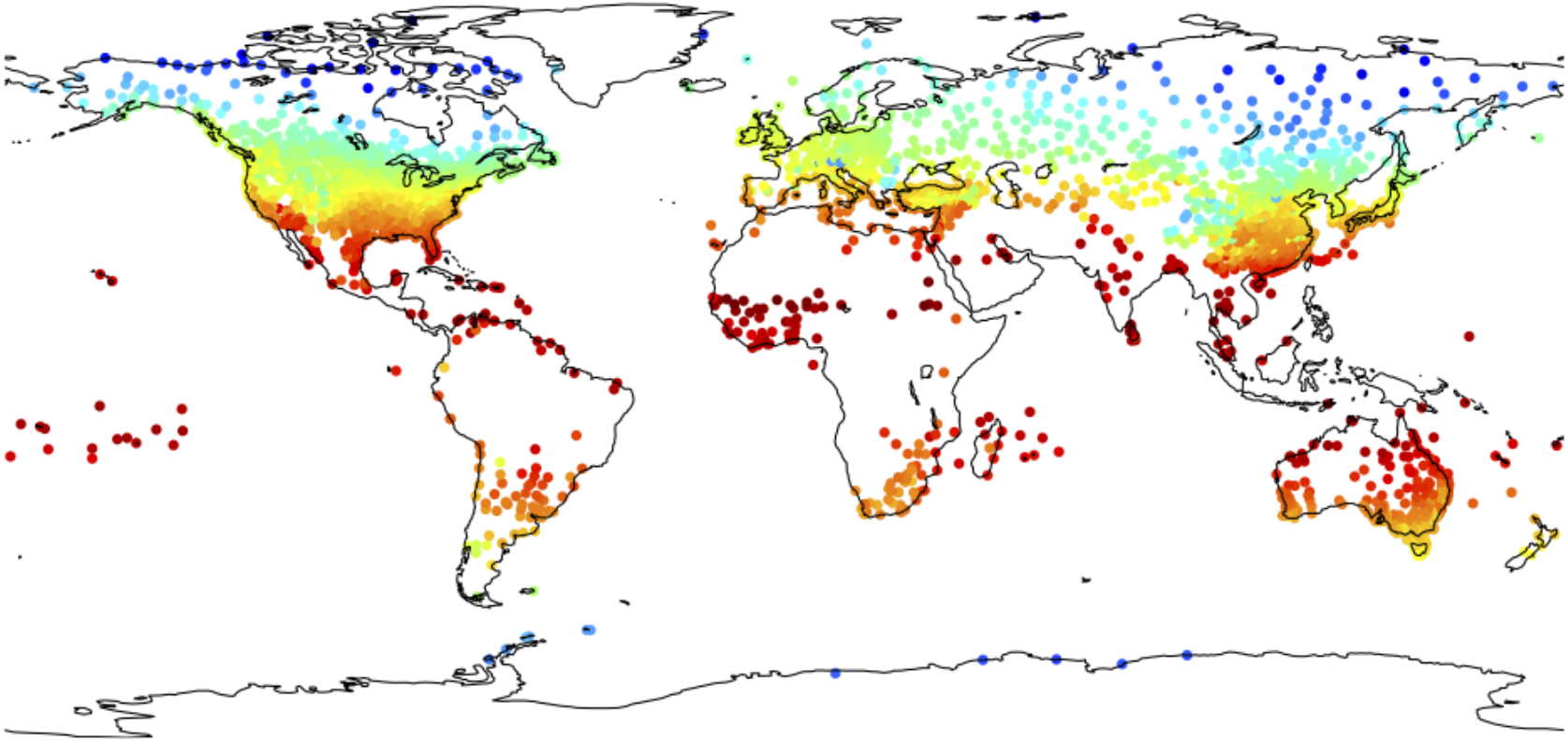
- Computationally, $\mathcal{O}(N^3)$ is simply too bad.
- They are difficult to extend: Non-stationary covariances, non-flat spaces, non-Gaussian models

Modelling with GMRFs has great computational properties but can be difficult when they're not on a lattice.

Especially for geo-statical data, it is important that we can use the model to interpolate data (i.e. kriging): this is difficult with GMRFs defined on graphs!



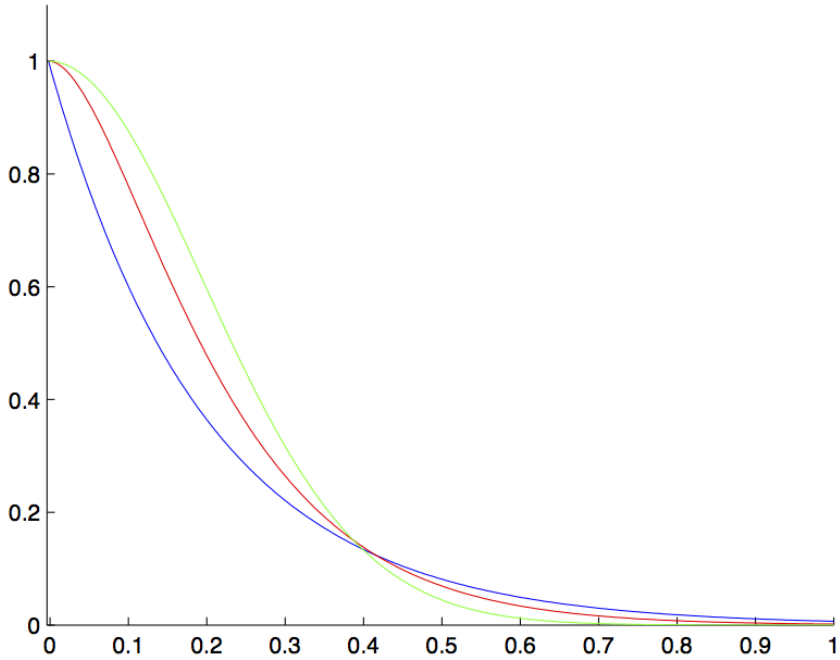
What do we want?



A method for constructing models that is

- flexible and easy to generalize
- computationally feasible
- applicable to both discrete and continuous domains
- easier to interpret than through precisions/covariances

A basic requirement



The most popular geostatistical models are the Gaussian Matérn fields.
Before we consider how easily we can generalize, a basic require will be that we can use the method to construct computationally efficient representations of these fields.

GMRF approximations to GRFs

A natural idea is to start with some GRF model and try to approximate it as a GRMF.

Say that we have a GRF $\mathbf{z}$ specified on a regular lattice using a covariance function $C(\mathbf{h})$.

Let $\mathbf{x}$ be a GMRF with some precision matrix $\mathbf{Q}(\boldsymbol{\theta})$.

If we let $\pi(\mathbf{x};\boldsymbol{\theta})$ denote the density of the GMRF and let $\pi(\mathbf{z})$ denote the density of the GRF we want to approximate.

Choose the parameters of the GMRF in some "optimal" way:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}_\infty^+}{\arg\min}\, \mathcal{D}(\pi(\mathbf{x};\boldsymbol{\theta}), \pi(\mathbf{z}))$$

where $\boldsymbol{\Theta}_\infty^+$ denotes the valid parameter space of the GMRF and $\mathcal{D}$ is some measure of discrepancy between the two densities.

GMRF approximations to GRFs

For example, let $\mathcal{D}$ be the KL-divergence and solve the minimization problem using a gradient descent method.

This often works surprisingly well, see the course book.

However:

- If we update the parameters in the GRF, we have to redo the GMRF fit.
- If we change the grid, we have to redo the GMRF fit.
- This is computationally expensive!
- One method is to tabulate values of $\boldsymbol{\theta}$ for common covariance functions, for different parameter values.

Thus, this is a possible solution, but it is not as flexible as we would like.

Reducing the dimension

Most of the methods aimed at reducing the "big N problem" in spatial statistics is based on some sort of low-dimensional approximation:

$$x(s) \approx \sum_{i=1}^{n} w_i \varphi_i(s),$$

where $\boldsymbol{w}$ is jointly Gaussian and $\varphi_i(s)$ are a set of known deterministic functions.

If $\boldsymbol{w} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then the covariance function of $x(\boldsymbol{s})$ is

$$c(\boldsymbol{s}_1, \boldsymbol{s}_2) = \boldsymbol{\Phi}(\boldsymbol{s}_1)^T \boldsymbol{\Sigma} \boldsymbol{\Phi}(\boldsymbol{s}_2),$$

where $\boldsymbol{\Phi}(\boldsymbol{s})$ is a vector with the φ_i functions evaluated at point $\boldsymbol{s}$.

Kernel representations

A prototypical low-dimensional approach is the kernel approach:

- It is popular!
- It is easy to analyse!

The idea is simple: Define a GRFs as a convolution

$$x(\mathbf{s}) = \int_{\mathbb{R}^2} k(\mathbf{s}, \mathbf{t}) \mathcal{B}(d\mathbf{t}),$$

where $\mathcal{B}(\mathbf{s})$ is a Brownian sheet, and $k(\mathbf{s}, \mathbf{t})$ is a deterministic "kernel" function.

The covariance function of $x(\mathbf{s})$ is given by

$$C(\mathbf{s}, \mathbf{t}) = \int k(\mathbf{s}, \mathbf{u}) k(\mathbf{t}, \mathbf{u}) d\mathbf{u}$$

Kernel representations

If the kernel is stationary: $k(\mathbf{s}, \mathbf{t}) = k(\mathbf{s})$, we can relate the kernel and the covariance function using the Fourier transform $\mathcal{F}$:

$$(2\pi)^d |\mathcal{F}(k)|^2 = \mathcal{F}(C) = S$$

where S denotes the spectrum.

Thus, if for example C is a Matérn covariance function

$$C(\mathbf{h}, \kappa, \nu, \phi),$$

the corresponding symmetric non-negative kernel is

$$C(\mathbf{h}, \kappa, \nu/2 - d/4, \sqrt{\phi}).$$

(Show this using the spectrum of the Matérn covariance function)

Kernel representations

It is often suggested that we model $k(\cdot, \cdot)$ directly. This is an easy way to obtain non-stationary covariance models.

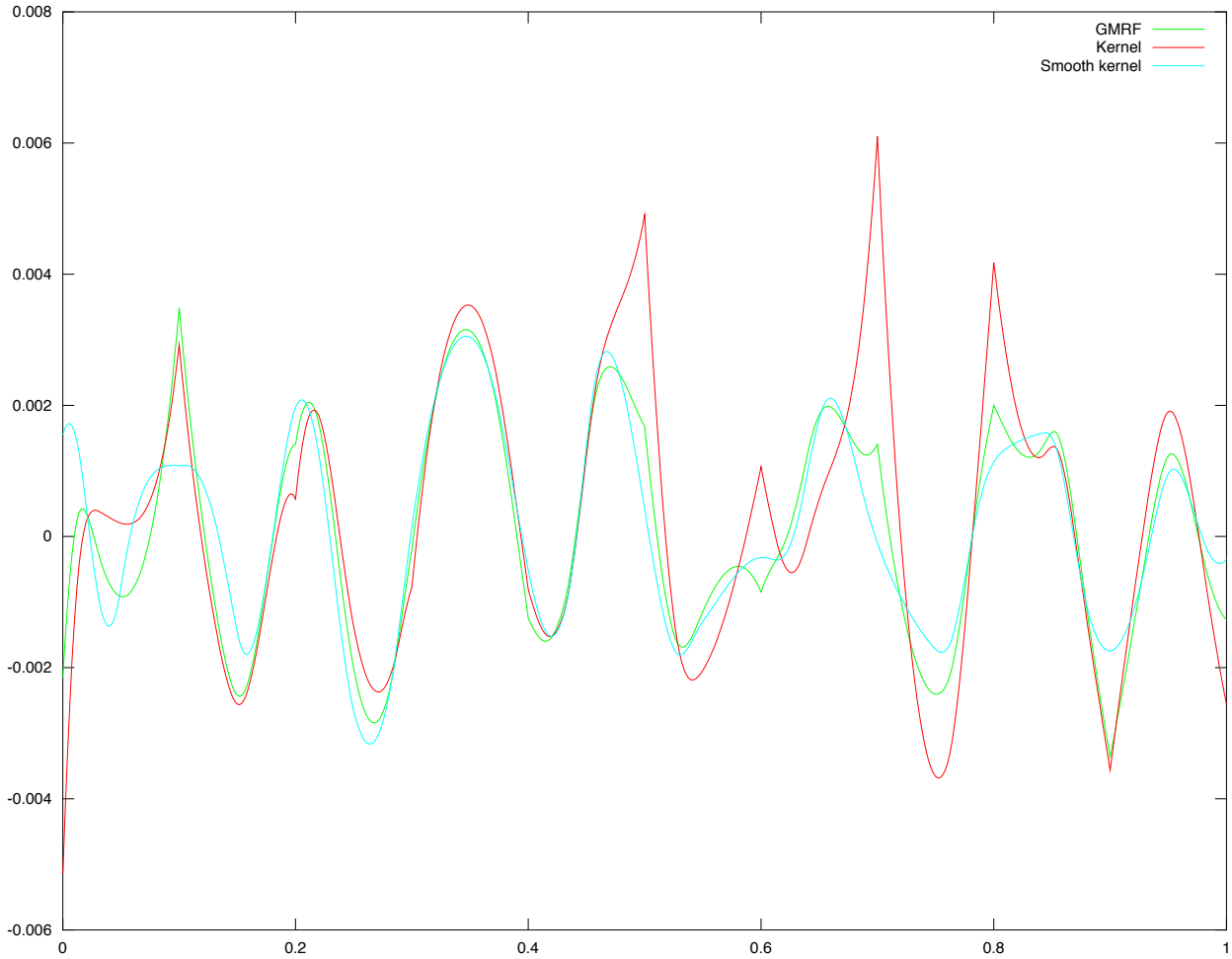
To "solve" the big N problem, a common strategy is to approximate the integral in the definition by a sum (Higdon, '98)

$$x(s) \approx \sum_{i=1}^{n} k(x, t_i)\xi_i,$$

where ξ_i are i.i.d. normals.

This does not work well. (Bolin and Lindgren '10, Simpson, Lindgren, Rue, '10)

So what happens?



Approximation properties

Appropriate question

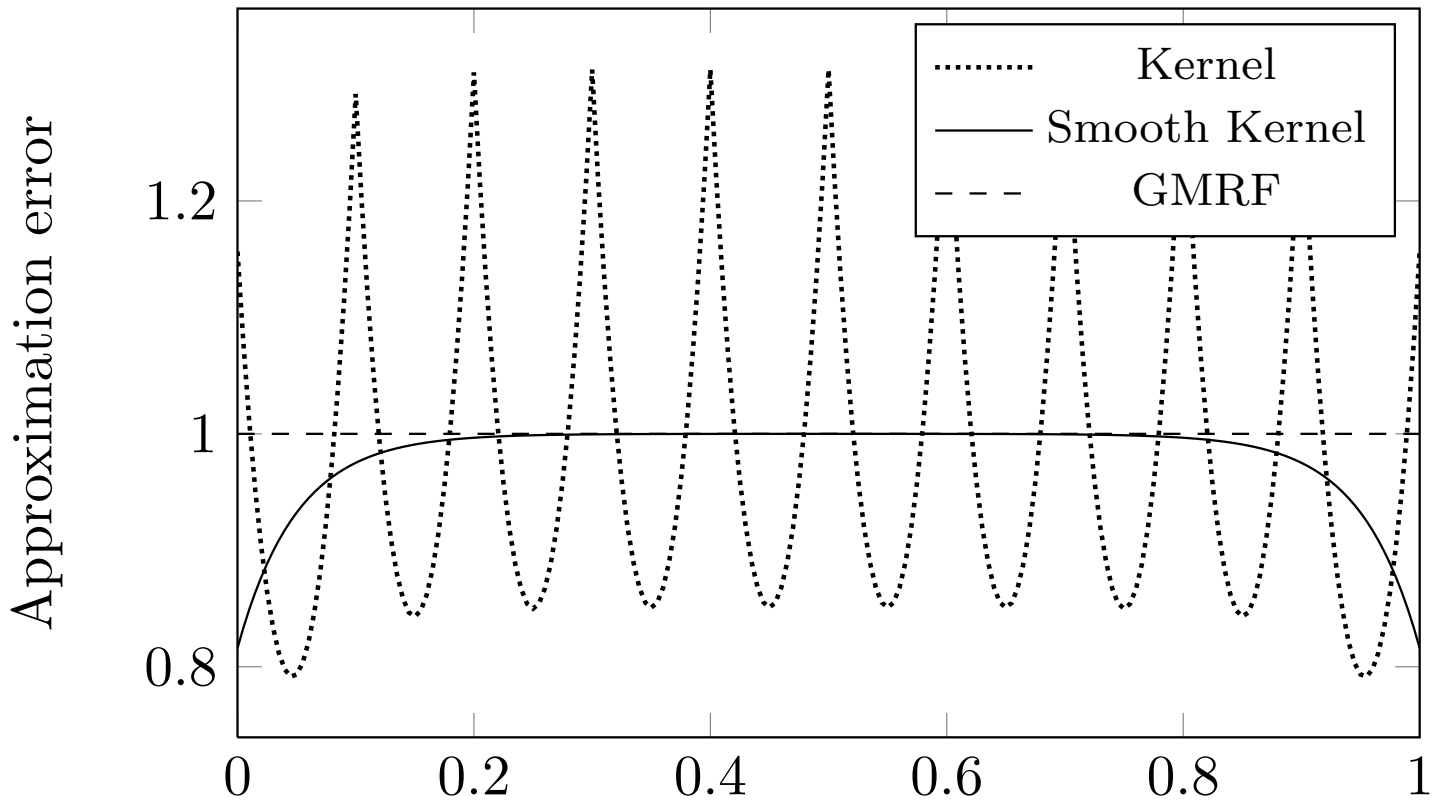
How well can realisations of $x(\cdot)$ be approximated by functions of the form $\sum_{i=1}^{n} w_i \varphi_i(s)$?

Approximation methods are often motivated through asymptotic results. But this is *not* an asymptotic question! n never goes to infinity.

- Equivalent question: How well do functions in $\mathrm{span}\{\varphi_i\}_{i=1}^{n}$ approximate functions of a given smoothness?
- Related question: How stable is the approximation procedure? (is $\|P_n f(\cdot)\| \leq C \|f\|$)?

Without considering these questions, you *cannot* know how a method will work!

Best Kernel approximation to a constant

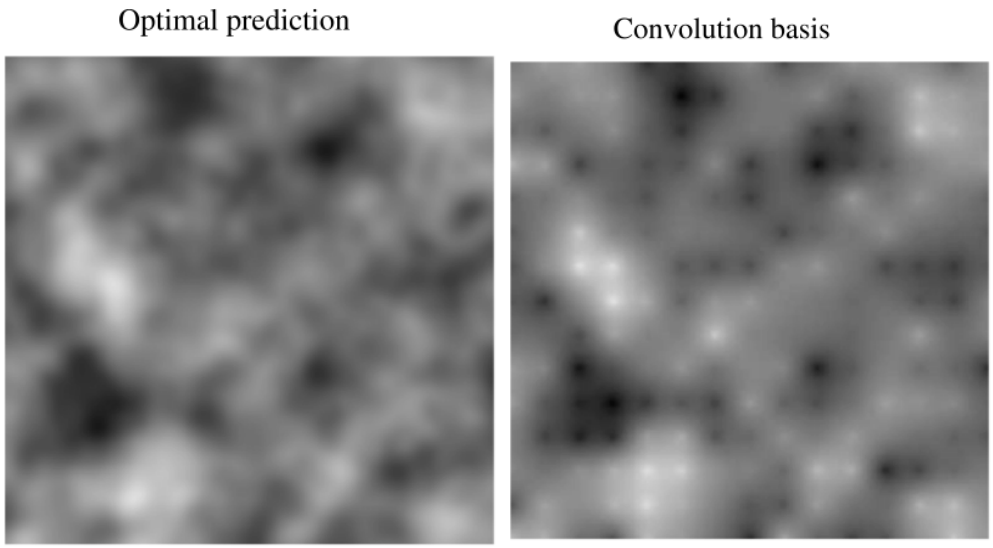


Why did kernel methods perform badly?

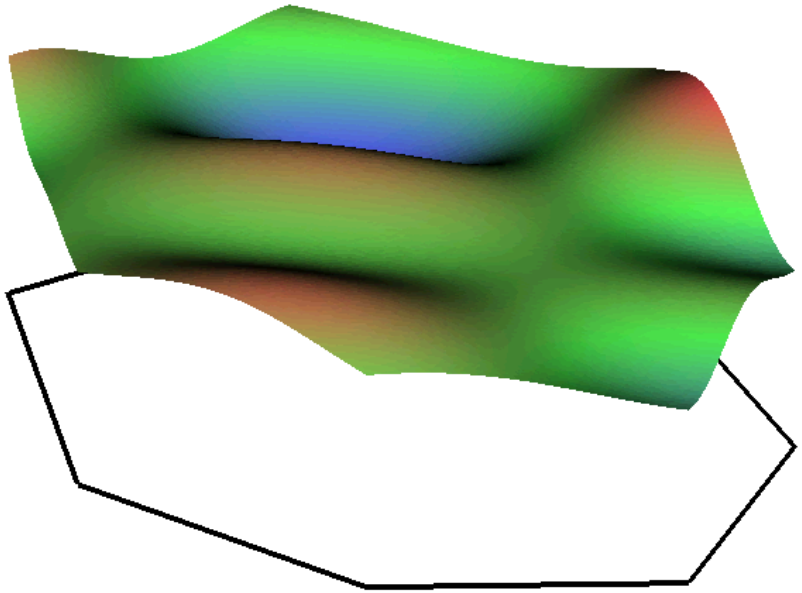
Kernel methods performed badly because

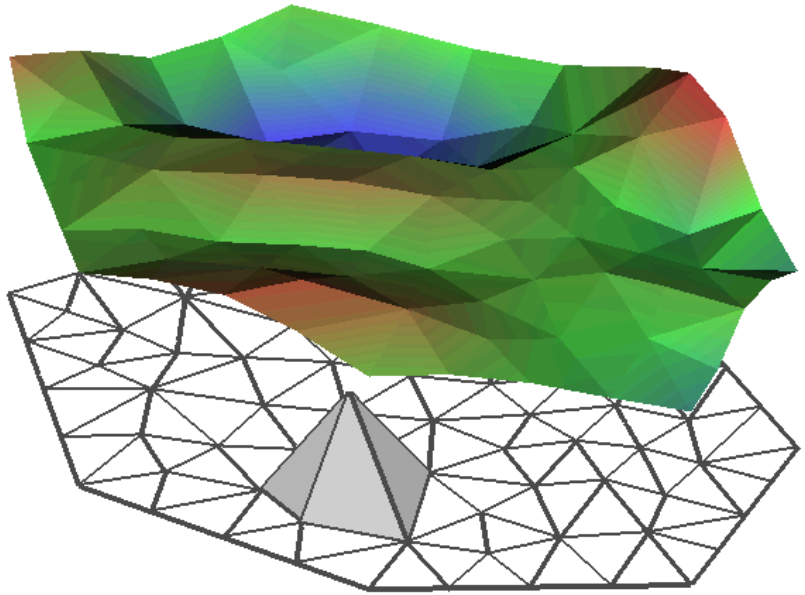
- there weren't enough points.
- the range was smaller than the grid spacing.
- the basis functions depend on the parameter being inferred!

This is a common problem and leads to "spotty" spatial predictions and bad uncertainty estimates.

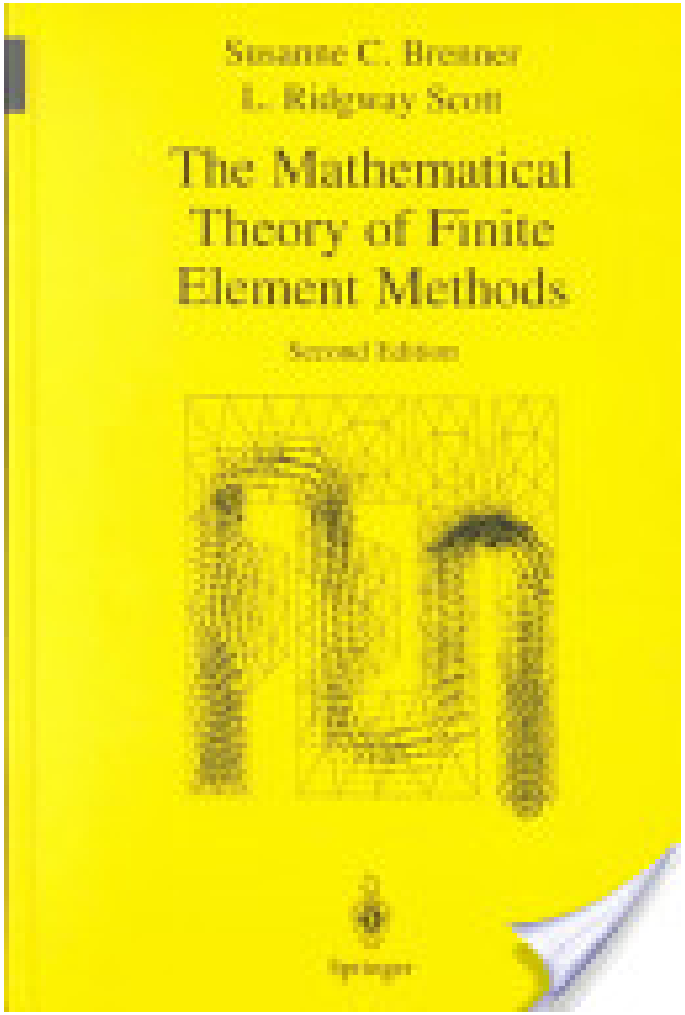


Piecewise linear approximations of surfaces





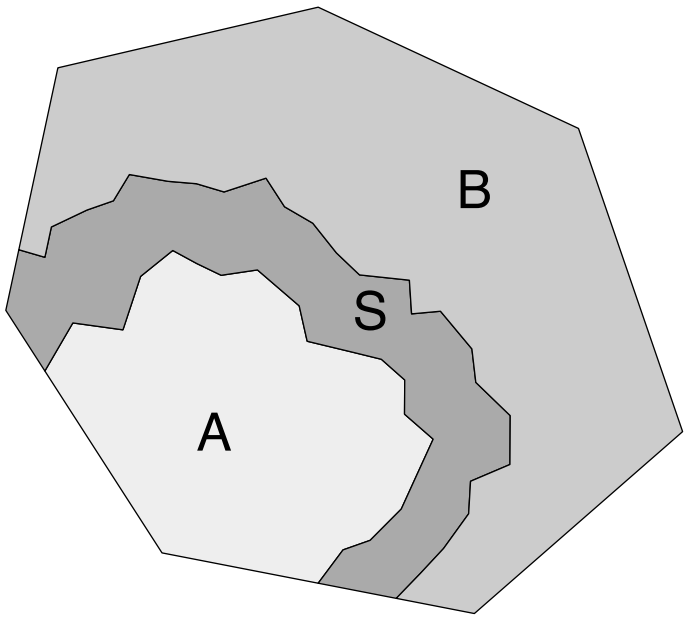
Known approximation properties



How can we use these functions?

There is no obvious way to use piecewise linear functions...

The secret is in the Markov property



General Result (Rosanov 1977)

The power spectrum of a stationary Markovian Gaussian random field has the form $R(\boldsymbol{k}) = 1/p(\boldsymbol{k})$, where $p(\boldsymbol{k})$ is a positive, symmetric polynomial.

Can we salvage something from this?

Sometimes it's useful to be an engineer!

An engineering calculation

Let L be a differential operator. Then the solution to

$$Lx(s) = W(\cdot)$$

is a Gaussian random field and it has the Markov property.

- "Prove" it using Fourier transforms.
- The derivatives (local) produce the Markov property (local)
- Now we're solving (partial) differential equations!

What does this remind us of?

- Recall the SAR(1) model

$$4x_i - (x_n + x_s + x_e + x_w) \sim N(0, \sigma^2).$$

- Also remember that

$$-\frac{d^2x}{ds^2} \approx \frac{-x(s+h) + 2x(s) - x(s-h)}{h^2}$$

- So if we scale our SAR(1) model and let the lattice spacing $h \to 0$, we get

$$-\Delta x(s) \equiv -\left(\frac{d^2x}{ds_1^2} + \frac{d^2x}{ds_2^2}\right) \stackrel{d}{=} W(s)$$

In the context of GMRFs

In this context, this was first noted by Whittle in the 50s (!!) who noted that 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.

Practical interpretation of the parameters

We have

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

where $\alpha = \nu + d/2$ is an integer.

- κ^2 is a range parameter. The approximate range is

$$\texttt{range} \approx \frac{\sqrt{8\nu}}{\kappa}.$$

- The variance of the model is

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

So which models do we get?

So, according to the Whittle characterisation of the Matérn covariance functions, we get a Markovian random field when $\alpha = \nu + d/2$ is an integer.

When d is odd, Matérn models with $\nu \in 1/2\mathbb{N}$.

- This include the Thin Plate Spline model ($\nu = 1$)
- And the exponential covariance ($\nu = 1/2$).

When d is even, we get Matérn models with $\nu \in \mathbb{N}$

- This include the Thin Plate Spline model ($\nu = 1$)
- But not the exponential covariance ($\nu = 1/2$).
- or as Besag called it, the "spatially unappealing exponential correlation" function.

A more proper derivation

Before we go into how we can use these thing, it may be instructive to try to understand how we define these concepts mathematically.

Do not panic!

This will require some prior knowledge of analysis, so the purpose here is mostly to give a flavour of how we deal with these things.

We will start with a more general construction, which also works for Lévy fields (details in Bolin, 2014), and then move to generalised Gaussian fields (details in Lindgren et al, 2011).

The goal is to clarify in what way the solution to

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X = \dot{M} \tag{1}$$

exists, and quantify some properties of the solution.

The noise $\dot{M}$

Let M be an independently scattered L_2-valued random measure with $\mathsf{E}(|M(\,\mathrm{d}\mathbf{x})|^2) = C\,\mathrm{d}\mathbf{x}$ for some constant $C < \infty$

Example: A measure of this type is the classical Brownian sheet:

- A Gaussian process on $\mathbb{R}^2$ with covariance function $C(\mathbf{s}, \mathbf{t}) = \min(s_1, t_1)\min(s_2, t_2)$.
- In terms of white noise, a measure $\mathbf{t} \to \dot{M}([0, t_1] \times [0, t_2])$.

Example: Stochastic generalized asymmetric Laplace noise:

- Take a Borel set B and define a measure through the characteristic function

$$\varphi_{\Lambda(B)}(u) = e^{i\gamma m(B)u}\left(1 - i\mu u + \frac{\sigma^2}{2}u^2\right)^{-m(B)},$$

where the measure m is the Lebesgue measure.

The differential operator

$$\mathcal{T} = (\kappa^2 - \Delta)^{\frac{\alpha}{2}}$$

is defined using the Fourier transform through

$$\mathcal{F}(\mathcal{T} f) = \mathcal{P}\hat{f},$$

where $\hat{f}$ is the Fourier transform of the function f,

$$(\mathcal{P}\hat{f})(\mathbf{k}) = (\kappa^2 + \mathbf{k}^\top \mathbf{k})^{\frac{\alpha}{2}} \hat{f}(\mathbf{k}),$$

The operator is well-defined for example for $f \in L_p(\mathbb{R}^d)$ for $1 \leq p \leq \infty$.

The definition applies also when f is a distribution (generalized function) or, more specifically, a tempered distribution.

The equation

The equation is viewed as an equation for two random (tempered) distributions so it has to be interpreted in the weak sense:

$$\mathcal{T}X(\varphi) = \dot{M}(\varphi), \tag{2}$$

where φ is in some appropriate space of test functions

$\dot{M}(\varphi)$ is defined as the linear functional $\dot{M}(\varphi) = \int \varphi(\mathbf{s}) M(\,\mathrm{d}\mathbf{s})$.

We can rewrite the equation in a more explicit fashion as

$$X(\mathcal{T}\varphi, \omega) = \int \varphi(\mathbf{s}) M(\,\mathrm{d}\mathbf{s}, \omega). \tag{3}$$

The equation should hold for ω in a certain full probability set $\mathcal{S}_0 \in \mathcal{S}$ (i.e. $\mathsf{P}(\mathcal{S}_0) = 1$) and universally for each φ.

Classical Sobolev spaces

Let E be the Schwartz space of rapidly decreasing functions on $\mathbb{R}^d$.

Take $u \in E'$ (the dual of E, the space of tempered distributions)

Define the Fourier transform of u as $\hat{u}(\varphi) = u(\hat{\varphi})$, where $\hat{\varphi}$ is the usual Fourier transform on $\mathbb{R}^d$ of $\varphi \in E$.

Define a norm on E by

$$\|u\|_n = \int_{\mathbb{R}^d} (1 + |\mathbf{k}|^2)^n |\hat{u}(\mathbf{k})|^2 \, \mathrm{d}\mathbf{k}$$

and let H_n be the completion of E in this norm.

For $n \in \mathbb{N}$, H_n is identical to the classical Sobolev space of L_2 functions with all partial derivatives of order n or less in L_2.

H_{-n} is the dual space of H_n and contains distributions.

Regularizations

The right hand side of (3) may in principle not be defined on a full probability set uniformly for all φ.

However, the random linear functional $\varphi \to \dot{M}(\varphi)$ is continuous in probability on H_n for any $n \geq 0$, which follows from

$$\mathsf{E}(|\dot{M}(\varphi)|^2) = C \int \varphi(\mathbf{s})^2 \, \mathrm{d}\mathbf{s} = C\|\varphi\|_0^2,$$

This is enough to show that there exists a version of $\dot{M}$ which is almost surely in H_{-n} for $n > d/2$ (see Walsh, 1984)

Thus, there exists a version of $\dot{M}$ so that $\varphi \to \dot{M}(\varphi)$ is a random distribution.

We always assume that we have such a regularized version of $\dot{M}$.

The solution

We say that $X(\cdot, \omega)$ is an H_n-solution of (1) if

1. for a.e. ω, $X(\cdot, \omega)$ is an element of H_{-n}, and
2. (3) holds for every $\varphi \in H_n$.

In other words, we aim at finding a random functional X that

1. almost surely is a distribution
2. satisfies (1) as a continuous functional on H_n.

Theorem

Assume that M is an independently scattered L_2-valued random measure with $\mathrm{E}(|M(\,\mathrm{d}\mathbf{x})|^2) = C\,\mathrm{d}\mathbf{x}$. Then for $\kappa > 0$, $\alpha > 0$, there exists a random functional $X : H_n \times \mathcal{S} \to \mathbb{R}$ such that for a certain set $\mathcal{S}_0$, $P(\mathcal{S}_0) = 1$ and for all $\omega \in \mathcal{S}_0$ and all $\varphi \in H_n$

$$X(\varphi, \omega) = \int G^\alpha \varphi(\mathbf{x}) M(\,\mathrm{d}\mathbf{x}, \omega),$$

where $G^\alpha \varphi(\mathbf{x}) = \int G_\alpha(\mathbf{s}, \mathbf{x})\varphi(\mathbf{s})\,\mathrm{d}\mathbf{s}$ and G_α is given by

$$G_\alpha(\mathbf{s}, \mathbf{t}) = \frac{2^{1-\frac{\alpha-d}{2}}}{(4\pi)^{\frac{d}{2}}\Gamma(\frac{\alpha}{2})\kappa^{\alpha-d}} (\kappa\|\mathbf{s}-\mathbf{t}\|)^{\frac{\alpha-d}{2}} K_{\frac{\alpha-d}{2}}(\kappa\|\mathbf{s}-\mathbf{t}\|).$$

This is the unique H_n-solution to

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X = \dot{M}$$

if $n > d/2$, and moreover we have $X \in H_m$ almost surely for $m < \alpha - d/2$.

Embeddings and smoothness

The Sobolev embedding theorem

H_n can be embedded in the Hölder space $C_k^r(\mathbb{R}^d)$ where $n - (r + k) = d/2$ and $r \in (0, 1)$.

- The solution X can be identified with a random function if $\alpha > d/2$ since $X \in H_m$ almost surely for $m < \alpha - d/2$.
- If $\nu > d/2$, one has that $X \in C_k^r(\mathbb{R}^d)$ almost surely, where k is the integer part of $\nu - d/2$ and $r = \nu - d/2 - k$.
- Only $\nu > 0$ is required for continuity in the Gaussian case.
- Thus, general Matérn fields are less smooth than the Gaussian Matérn fields, for the same smoothness parameter ν.

Let's focus on Gaussian fields for now...

... But to make things a bit more interesting, relax the assumption of $\mathbf{s} \in \mathbb{R}^d$, and instead let $\mathbf{s} \in \Omega$.

To not make things too interesting, let Ω be a "nice" manifold: It should be a metric d-manifold (behaves locally as $\mathbb{R}^d$) with a well-behaved boundary.

To simplify notation, define the inner product

$$\langle f, g \rangle_\Omega = \int_\Omega f(s)g(s)ds$$

Gaussian field

A random function $x : \Omega \to \mathbb{R}$ is a Gaussian field if $\{x(\mathbf{u}_k), k = 1, \ldots, n\}$ are jointly Gaussian random vectors for every finite set of points $\{\mathbf{u}_k, k = 1, \ldots, n\}$.

White noise on manifolds

The "standard" definition of white noise via Brownian sheets is not applicable on general manifolds.

Generalized Gaussian fields

A random $L^2(\Omega)$ (generalized) function $x : \Omega \to \mathbb{R}$ is a generalized Gaussian field if $\{\langle x, \varphi \rangle, k = 1, \ldots, n\}$ are jointly Gaussian for any set of test functions $\{\varphi_i \in L^2(\Omega), i = 1, \ldots, n\}$

Gaussian white noise

Gaussian white noise, W, is an $L^2(\Omega)$-bounded generalized GF such that, for any set of test functions $\{\varphi_i \in L^2(\Omega), i = 1, \ldots, n\}$, $\{\langle x, \varphi \rangle, k = 1, \ldots, n\}$ are jointly Gaussian with mean zero and covariance $\mathrm{Cov}(\langle W, \varphi_i \rangle, \langle W, \varphi_j \rangle) = \langle \varphi_i, \varphi_j \rangle$.

Boundary conditions (I)

If we are on a bounded manifold, we typically use Neumann boundary conditions for the SPDE:

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X(\mathbf{u}) = W(\mathbf{u}), \quad \mathbf{u} \in \Omega$$
$$\partial_d (\kappa^2 - \Delta)^j X = 0, \quad \mathbf{u} \in \partial\Omega, \quad , j = 0, 1, \ldots, \lfloor (\alpha - 1)/2 \rfloor$$

The solution to this boundary value problem will *not* have a Matérn covariance function, for example for $\Omega = [0, L]$, we have that the solution has covariance

$$\mathrm{Cov}(x(s), x(u)) = \sum_{k=-\infty}^{\infty} (r_M(u, v - 2kL) + r_M(u, 2kL - v))$$

where r_M is the Matérn covariance on $\mathbb{R}$

The theorem generalizes naturally to rectangles in $\mathbb{R}^d$.

Boundary conditions (II)

In practice, if the effective range, ρ, of the field is small compared with the length of the interval, only the three main terms need to be included for a very close approximation:

$$\mathrm{Cov}(x(s), x(u)) = r_M(u, v) + r_M(u, -v) + r_M(u, 2L - v)$$

The resulting covariance function is nearly indistinguishable from the Matérn covariance function at distances greater than 2ρ from the boundary.

Approximating the SPDE

So we now have an SPDE representation of Gaussian Matérn fields. Can we use this to construct a useful practical method?

Let's simplify things: set $\nu + d/2 = 2$, the SPDE becomes

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

which only involves second derivatives, which are nice.

We are looking for a piecewise linear random field

$$x_n(s) = \sum_{i=1}^{n} w_i \varphi_i(s)$$

for piecewise linear functions $\varphi_i(s)$ that *best* approximates the solution to $(\kappa^2 - \Delta)x(s) = W(s)$.

Step 1: The 'weak' solution

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 $\psi \in H_\alpha$.

However, in practice we construct a finite element approximation by choosing the space of test functions as something that is easier to handle, such as

- span$\{\varphi\}_{i=1}^n$

This particular choice is called the Galerkin method.

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

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

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

This is good:

1. LHS has things we can compute
2. RHS has integrals of white noise.

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 $\mathsf{N}(0, \mathbf{C})$, where

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

Step 2: The left-hand side

The stochastic Green's first identity

If $\nabla f \in L^2(\Omega)$ and Δx is $L^2(\Omega)$ bounded, then with probability 1,

$$\langle f, -\Delta x \rangle_\Omega = \langle \nabla f, \nabla x \rangle_\Omega - \langle f, \partial_n x \rangle_{\partial\Omega}$$

If ∇x is $L^2(\Omega)$ bounded and $\Delta f \in L^2(\Omega)$, then with probability 1,

$$\langle x, -\Delta f \rangle_\Omega = \langle \nabla x, \nabla f \rangle_\Omega - \langle x, \partial_n f \rangle_{\partial\Omega}$$

Using this, we can write the LHS as

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

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

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

This looks familiar!

... 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 $\mathbf{w} \sim \mathsf{N}(\mathbf{0}, \mathbf{Q}^{-1})$ is a GMRF.
Say now that we observe $x(s)$ at locations $\mathbf{s}_1, \ldots, \mathbf{s}_N$ 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

$$\mathbf{y} \sim \mathsf{N}(\mathbf{A}\mathbf{w}, \sigma^2 \mathbf{I})$$
$$\mathbf{w} \sim \mathsf{N}(\mathbf{0}, \mathbf{Q}^{-1})$$

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

The RW2 model for irregular locations

Before we continue with the general construction, it may be instructive to consider the RW2 model.

Recall the representation of the model as the solution to a stochastic differential equation

$$\Delta x(t) = W(t)$$

We see that this is the special case $\alpha = 2$, $\kappa = 0$, and $d = 1$ in the general construction.

The idea behind the construction of the precision matrix for irregular locations was to use piecewise linear basis functions in the FEM construction.

We continue next time!