An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach
by Finn Lindgren, Havard Rue, & Johan Lindström, JRSSB, 73(4), 423–498 (2011)

Whitney Huang & Ji Hwan Oh
Department of Statistics
Purdue University

October 16, 2014
Outline

Motivation

Background

GP/GMRF connection: SPDE

Extensions: beyond classical Matérn models
Gaussian process (GP) geostatistics

Model:

\[ Y(s) = \mu(s) + \eta(s) + \epsilon(s), \quad s \in S \subset \mathbb{R}^d \]

where

\[ \mu(s) = X^T(s)\beta, \quad \{\eta(s)\}_{s \in S} \sim \text{GP} \left(0, C(\cdot, \cdot)\right) \]

\[ C(s, s') = \sigma^2 \rho_\theta \left(\|s - s'\|\right), \text{ and } \epsilon(s) \sim N(0, \tau^2) \quad \forall s \in S \]

Log-likelihood:

Given data \( Y = (Y(s_1), \cdots, Y(s_n))^T \)

\[ l_n(\beta, \theta, \sigma^2, \tau^2) \propto -\frac{1}{2} \log \left| C_{n,n}(\theta, \sigma^2) + \tau^2 I_n \right| \]

\[ -\frac{1}{2}(Y - X^T\beta)^T \left[ C_{n,n}(\theta, \sigma^2) + \tau^2 I_n \right]^{-1} (Y - X\beta) \]
“Big \( n \) Problem” in geostatistics

- Modern environmental instrument has produced a wealth of space–time data \( \Rightarrow n \) is big
- Evaluation of the likelihood function involves factorizing large covariance matrices that generally requires
  - \( \mathcal{O}(n^3) \) operations
  - \( \mathcal{O}(n^2) \) memory
- Modeling strategies are developed to deal with large spatial data set
  - covariance tapering (Furrer et al. 06)
  - low rank approximation (Cressie & Johannesson 08, Banerjee et al. 08)
  - likelihood approximation (Vecchia 88, Stein 04)
  - Gaussian Markov Random Fields (GMRF) approximation
Outline

Motivation

Background

GP/GMRF connection: SPDE

Extensions: beyond classical Matérn models
Matérn covariance functions

The Matérn covariance function family is the most commonly used covariance function in geostatistics.

Definition
The Matérn covariance function is isotropic and has the parametric form

\[ C(Y(s + h), Y(s)) = M(h|\nu, \alpha) = \frac{\sigma^2 (\alpha \|h\|)^\nu}{2^{\nu-1} \Gamma(\nu)} K_\nu (\alpha \|h\|) \]

where

- \( \|h\| \) denotes the euclidean distance
- \( \nu > 0 \) is the smoothness parameter
- \( \alpha > 0 \) is the scaling parameter (\( \frac{1}{\alpha} \) is the range parameter)
- \( K_\nu \) is the modified Bessel function of the second kind of order \( \nu \)
Markov Random Fields
Gaussian Markov Random Fields

Definition
Let the neighbors to a point \( i \) be the points \( \mathcal{N}_i \) that are “close” to \( i \). A Gaussian random field \( \mathbf{X} \sim \mathcal{N}(\mu, \Sigma = Q^{-1}) \) that satisfies

\[
p(X_i|X_j, j \neq i) = p(X_i|X_j : j \in \mathcal{N}_j)
\]

is a Gaussian Markov random field with \( Q_{ij} = 0 \) iff \( X_i \perp X_j|X_{\neg ij} \)

Example (AR(1) process)
\( X_t = \rho X_{t-1} + \varepsilon_t \) where \( \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \)
Remarks: GP vs. GMRF in geostatistical modeling

- **+**: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- **−**: Inference involves factorizing covariance matrices
- **+**: GMRF model is computationally efficient due to the sparse precision matrix
- **−**: Only for discretely indexed spatial data

Main point of this paper:

GP inference $\iff$ SPDE $\iff$ GMRF computation
Remarks: GP vs. GMRF in geostatistical modeling

- **+:** GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- **−:** Inference involves factorizing covariance matrices
- **+:** GMRF model is computationally efficient due to the sparse precision matrix
- **−:** Only for discretely indexed spatial data

Main point of this paper:

\[ \overbrace{\text{GP inference}}^{\text{SPDE}} \iff \overbrace{\text{GMRF computation}} \]
Remarks: GP vs. GMRF in geostatistical modeling

▶ +: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
▶ -: Inference involves factorizing covariance matrices
▶ +: GMRF model is computationally efficient due to the sparse precision matrix
▶ -: Only for discretely indexed spatial data

Main point of this paper:
Remarks: GP vs. GMRF in geostatistical modeling

- **+**: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- **−**: Inference involves factorizing covariance matrices
- **+**: GMRF model is computationally efficient due to the sparse precision matrix
- **−**: Only for discretely indexed spatial data

Main point of this paper:
Remarks: GP vs. GMRF in geostatistical modeling

- +: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- -: Inference involves factorizing covariance matrices
- +: GMRF model is computationally efficient due to the sparse precision matrix
- -: Only for discretely indexed spatial data

Main point of this paper:
Remarks: GP vs. GMRF in geostatistical modeling

- +: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- -: Inference involves factorizing covariance matrices
- +: GMRF model is computationally efficient due to the sparse precision matrix
- -: Only for discretely indexed spatial data

Main point of this paper:

\[
\begin{array}{c}
\text{GP} \\
\text{inference}
\end{array} \overset{\text{SPDE}}{\leftrightarrow} \begin{array}{c}
\text{GMRF} \\
\text{computation}
\end{array}
\]
Outline

Motivation

Background

GP/GMRF connection: SPDE

Extensions: beyond classical Matérn models

Gaussian process $Y(s)$ with Matern covariance function is a stationary solution to the linear fractional stochastic partial differential equation:

$$(\alpha^2 - \Delta)^{\frac{\kappa}{2}} Y(s) = \mathcal{W}(s), \quad \kappa = \nu + \frac{d}{2}, \nu > 0$$

where

- $\mathcal{W}(s)$ is a spatial Gaussian white noise
- $\Delta = \sum_i \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator
- $d$ is the dimension of the spatial domain
The spectral density of Matérn covariance:

\[ f(w) \propto \frac{1}{(\alpha + \|w\|^2)^\kappa} \]

Spectral density for Markov Fields

A stationary fields is a Markov if and only if the spectral density is a reciprocal of a polynomial (Rozanov, 1977)

For the SPDE this implies \( \kappa \in \mathbb{Z} \) (or \( \nu = \kappa - 1 \in \mathbb{Z} \) for \( \mathbb{R}^2 \))
Basic idea

Construct a discrete approximation of the continuous spatial stochastic process using basis function \( \{ \psi_k \} \), and weights \( \{ w_k \} \),

\[
Y(s) = \sum_k \psi_k(s) w_k
\]

Find the distribution of \( w_k \) by solving

\[
(\alpha^2 - \Delta)^{\frac{\kappa}{2}} Y(s) = \mathcal{W}(s)
\]
Solving the SPDE

A stochastic weak solution to the SPDE is given by

$$\left[ \langle \varphi_k, (\alpha^2 - \Delta)^{\frac{\kappa}{2}} Y \rangle \right]_k \overset{L}{=} \left[ \langle \varphi_k, W \rangle \right]_k$$

for each set of test functions $\varphi_k$

Replacing $Y$ with $\sum_k \psi_k w_k$ gives

$$\left[ \langle \varphi_i, (\alpha^2 - \Delta)^{\frac{\kappa}{2}} \psi_j \rangle \right]_{i,j} w \overset{L}{=} \left[ \langle \varphi_k, W \rangle \right]_k$$

For the case of $\kappa = 2$ and $\varphi_i = \psi_i$ (Galerkin)

$$\left( \alpha^2 \underbrace{[\langle \psi_i, \psi_j \rangle]}_{C} + \underbrace{[\langle \psi_i, -\Delta \psi_j \rangle]}_{G} \right) w \overset{L}{=} \left[ \langle \varphi_k, W \rangle \right]_k$$

$$\overset{N(0,C)}{\text{Galerkin}}$$
Solution to SPDE

A weak solution to the SPDE ($\kappa = 2$)

$$ (\alpha^2 - \Delta) \ Y(s) = \mathcal{W}(s) $$

is given by $Y(s) = \sum_k \psi_k(s)w_k$ where

$$ (\alpha^2C + G)w \sim N(0, C) $$

Actually, there exists a recursive formula of $Q$, the precision matrix, for general $\kappa \in \mathbb{Z}$

- $\kappa = 1$: $Q_1 = \alpha^2C + G$
- $\kappa = 2$: $Q_2 = (\alpha^2C + G)^T C^{-1} (\alpha^2C + G)$
- $\kappa > 2 \in \mathbb{Z}$: $Q_\kappa = (\alpha^2C + G)^T C^{-1} Q_{\kappa-2} (\alpha^2C + G)$
Choice of the basis functions

- **Harmonic functions**: gives a spectral representation
- **Eigenfunction**: leads to the Karhunen-Loéve expansion
- **Piece-wise linear basis**: gives a (almost) GMRF

Remark: using piece-wise linear basis only neighboring basis overlap $\Rightarrow C_{ij}, G_{ij}$ are sparse and to approximate $C^{-1}$ so that $\tilde{C}^{-1}$ is sparse too.
Outline

Motivation

Background

GP/GMRF connection: SPDE

Extensions: beyond classical Matérn models
Extension: model on manifold

\[(\alpha^2 - \Delta)^{\frac{\kappa}{2}} Y(s) = \mathcal{W}(s), \quad s \in S^2\]
Further Extensions

- non-stationary model on a sphere

\[(\alpha^2(s) + \Delta)^{\frac{\kappa}{2}} \tau(s) Y(s) = \mathcal{W}(s), \quad s \in S^2\]

- non-separable anisotropic space-time model

\[\left(\frac{\partial}{\partial t} + (\alpha^2 + \mathbf{m} \cdot \nabla - \nabla \cdot \mathbf{H} \nabla)^{\frac{\kappa}{2}}\right) Y(s, t) = \mathcal{W}(s, t)\]

where \((s, t) \in S^2 \times \mathbb{R}\)
Inference: INLA for Approximate Bayesian inference

We can formulate the GP geostatistical model

\[ Y(s) = \mu(s) + \eta(s) + \epsilon(s), \quad s \in S \subset \mathbb{R}^d \]

in a hierarchical fashion and to be a Bayesian

- Data: \( p(y|x, \tau^2) \)
- Latent process: \( p(x|\beta, \theta, \sigma) \)
- Prior: \( p(\tau), p(\sigma), p(\beta), p(\theta) \)

Inference target:

- Posterior for the parameters: \( p(\sigma, \tau, \beta, \theta|y) \)
- Posterior for the latent field: \( p(x|y) \)
- Integrated nested Laplace approximation (INLA) approach for doing Bayesian inference (Rue et al., 2009)