Gaussian Markov random fields: Efficient modelling of spatially dependent data

Johan Lindström\(^1\)  Finn Lindgren\(^2\)  Håvard Rue\(^2\)

\(^1\)Centre for Mathematical Sciences
Lund University

\(^2\)Department of Mathematical Sciences
Norwegian University of Science and Technology, Trondheim

Lund, April 28\(^{th}\), 2011
Spatial interpolation — Kriging

The standard problem

Given observations at some locations, \( Y(s_i), \ i = 1 \ldots n \), we want to make statements about the value at unobserved location(s), \( Y(s_0) \).

Data from Peter Jonsson (Engineering geology, Lund University).
Spatial interpolation — Kriging

The standard problem

Given observations at some locations, \( Y(s_i), i = 1 \ldots n \) we want to make statements about the value at unobserved location(s), \( Y(s_0) \).
Gaussian fields

A Gaussian model

We often assume that the observations come from a Gaussian field

$$\begin{bmatrix} Y_0 \\ Y \end{bmatrix} \in \mathbb{N}\left(\begin{bmatrix} \mu_0 \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma_{00} & \Sigma_{0n} \\ \Sigma_{0n}^T & \Sigma_{nn} \end{bmatrix}\right)$$

If the mean and covariance matrix are known the optimal predictor is the conditional expectation

$$E(Y_0|Y_1, \cdots, Y_n) = \mu_0 + \Sigma_{0n}\Sigma_{nn}^{-1}(Y - \mu).$$

However the mean and covariance matrix are typically not known.
We often assume that the observations come from a Gaussian field

\[
\begin{bmatrix} Y_0 \\ Y \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_0 \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma_{00} & \Sigma_{0n} \\ \Sigma_{0n}^T & \Sigma_{nn} \end{bmatrix} \right)
\]

If the mean and covariance matrix are known the optimal predictor is the conditional expectation

\[
E(Y_0|Y_1, \ldots, Y_n) = \mu_0 + \Sigma_{0n} \Sigma_{nn}^{-1} (Y - \mu).
\]

However the mean and covariance matrix are typically not known.
Parameter estimation

Assume a parametric form for the covariance and mean

$$\mathbf{Y} \sim \mathcal{N}(\mu(\theta), \Sigma(\theta)).$$

The log-likelihood becomes

$$l(\theta|\mathbf{Y}) = -\frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} \left( \mathbf{Y} - \mu(\theta) \right)^\top \Sigma(\theta)^{-1} \left( \mathbf{Y} - \mu(\theta) \right).$$

Problems:

1. The covariance matrix has $\mathcal{O}(n^2)$ unique elements.
2. Calculating $l(\theta|\mathbf{Y})$ takes $\mathcal{O}(n^3)$ time.

For the depth data:

- 11'705 observations
- Storing the covariance matrix $\sim$ 1 GB
- Evaluating $l(\theta|\mathbf{Y})$ once(!) 2.5 minutes.
Parameter estimation

Assume a parametric form for the covariance and mean

\[ Y \in \mathcal{N}(\mu(\theta), \Sigma(\theta)) \, . \]

The log-likelihood becomes

\[
l(\theta | Y) = -\frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} (Y - \mu(\theta))^\top \Sigma(\theta)^{-1} (Y - \mu(\theta))
\]

Problems:

1. The covariance matrix has \( \mathcal{O}(n^2) \) unique elements.
2. Calculating \( l(\theta | Y) \) takes \( \mathcal{O}(n^3) \) time.

For the depth data:

- 11'705 observations
- Storing the covariance matrix \( \sim 1 \text{ GB} \)
- Evaluating \( l(\theta | Y) \) once(!) 2.5 minutes.
Parameter estimation

Assume a parametric form for the covariance and mean

\[ Y \in \mathcal{N}(\mu(\theta), \Sigma(\theta)). \]

The log-likelihood becomes

\[
l(\theta|Y) = -\frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} (Y - \mu(\theta))^\top \Sigma(\theta)^{-1} (Y - \mu(\theta))\]

Problems:

1. The covariance matrix has \( \mathcal{O}(n^2) \) unique elements.
2. Calculating \( l(\theta|Y) \) takes \( \mathcal{O}(n^3) \) time.

For the depth data:

- 11’705 observations
- Storing the covariance matrix \( \sim 1 \) GB
- Evaluating \( l(\theta|Y) \) once(!) 2.5 minutes.
Gaussian Markov random fields

Let the neighbours $N_i$ to a point $s_i$ be the points $\{s_j, j \in N_i\}$ that are “close” to $s_i$.

**Gaussian Markov random field (GMRF)**

Gaussian random field $x \sim N(\mu, \Sigma)$ that satisfies

$$p(x_i|\{x_j : j \neq i\}) = p(x_i|\{x_j : j \in N_i\})$$

is a Gaussian Markov random field.

Using the precision matrix, $Q = \Sigma^{-1}$, the conditional expectation becomes

$$E(x_i|\{x_j : j \neq i\}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j \neq i} Q_{ij}(x_j - \mu_j)$$

A GMRF has a sparse precision matrix $j \in N_i \iff Q_{i,j} \neq 0$. 

---

Johan Lindström - johanl@maths.lth.se
Gaussian Markov random fields

Let the neighbours $\mathcal{N}_i$ to a point $s_i$ be the points $\{s_j, j \in \mathcal{N}_i\}$ that are “close” to $s_i$.

Gaussian Markov random field (GMRF)

Gaussian random field $x \sim N(\mu, \Sigma)$ that satisfies

$$p(x_i|\{x_j : j \neq i\}) = p(x_i|\{x_j : j \in \mathcal{N}_i\})$$

is a Gaussian Markov random field.

Using the precision matrix, $Q = \Sigma^{-1}$, the conditional expectation becomes

$$E(x_i|\{x_j : j \neq i\}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j \neq i} Q_{ij}(x_j - \mu_j)$$

A GMRF has a sparse precision matrix $j \in \mathcal{N}_i \iff Q_{i,j} \neq 0$. 

Johan Lindström - johanl@maths.lth.se
Gaussian Markov random fields

Let the **neighbours** $\mathcal{N}_i$ to a point $s_i$ be the points $\{s_j, j \in \mathcal{N}_i\}$ that are “close” to $s_i$.

**Gaussian Markov random field (GMRF)**

Gaussian random field $x \sim N(\mu, \Sigma)$ that satisfies

$$p(x_i|\{x_j : j \neq i\}) = p(x_i|\{x_j : j \in \mathcal{N}_i\})$$

is a Gaussian Markov random field.

Using the precision matrix, $Q = \Sigma^{-1}$, the conditional expectation becomes

$$E(x_i|\{x_j : j \neq i\}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j \neq i} Q_{ij} (x_j - \mu_j)$$

A GMRF has a **sparse** precision matrix $j \in \mathcal{N}_i \iff Q_{i,j} \neq 0$. 
The Markov property

The Markov property **does not** imply that events far apart are independent. Only that if we know what happens close by we can ignore things further away.

**Example**

We want to predict the temperature in Lund

1. If we know the temperature in Copenhagen, the temperature in London contributes very little information.

2. However, if we don’t know the temperature in Copenhagen, the temperature in London would help.

The sparse precision matrix makes GMRF computationally effective but it is hard to construct reasonable precision matrices.
The Markov property

The Markov property **does not** imply that events far apart are independent. Only that if we know what happens close by we can ignore things further away.

**Example**

We want to predict the temperature in Lund

1. If we know the temperature in Copenhagen, the temperature in London contributes very little information.
2. However, if we don’t know the temperature in Copenhagen, the temperature in London would help.

The sparse precision matrix makes GMRF computationally effective but it is hard to construct reasonable precision matrices.
The Markov property

The Markov property **does not** imply that events far apart are independent. Only that if we know what happens close by we can ignore things further away.

**Example**

We want to predict the temperature in Lund

1. If we know the temperature in Copenhagen, the temperature in London contributes very little information.

2. However, if we don’t know the temperature in Copenhagen, the temperature in London would help.

The sparse precision matrix makes GMRF computationally effective but it is hard to construct reasonable precision matrices.
The Markov property

The Markov property does not imply that events far apart are independent. Only that if we know what happens close by we can ignore things further away.

Example

We want to predict the temperature in Lund

1. If we know the temperature in Copenhagen, the temperature in London contributes very little information.

2. However, if we don’t know the temperature in Copenhagen, the temperature in London would help.

The sparse precision matrix makes GMRF computationally effective but it is hard to construct reasonable precision matrices.
A common covariance function

The Matérn covariance family

The covariance between two points at distance $\|u\|$ is

$$C(\|u\|) = \sigma^2 \frac{2^{1-\nu}}{I(\nu)} (\chi \|u\|)^\nu K_\nu(\chi \|u\|)$$

The Matérn family includes:

- Exponential, if $\nu = 1/2$.
- Gaussian, when $\nu \to \infty$. 
A common covariance function

The Matern covariance family

The covariance between two points at distance $\|u\|$ is

$$C(\|u\|) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\nu \|u\|)^\nu K_\nu(\nu \|u\|)$$

Fields with Matérn covariances are solutions to an SPDE (Whittle, 1954, 1963),

$$(\nu^2 - \Delta)^{\alpha/2} x(s) = \mathcal{W}(s).$$

Here $\mathcal{W}(s)$ is white noise, $\Delta = \sum_i \frac{\partial^2}{\partial s_i^2}$, and $\alpha = \nu + d/2$. 

Johan Lindström - johanl@maths.lth.se
Constructing a GMRF from the SPDE

We want to construct a (GMRF) solution to the SPDE.

The basic idea

Construct the solution as a finite basis expansion

\[ x(s) = \sum_k \psi_k(s) x_k, \]

with a suitable distribution for the weights \( \{ x_k \} \).

A stochastic weak solution is given by weights, \( \{ x_i \} \), such that the joint distribution fulfills

\[ \sum_i \langle \psi_j, (\nabla^2 - \Delta)^{\alpha/2} \psi_i x_i \rangle^D \equiv \langle \psi_j, \mathcal{W} \rangle \quad \forall j \]
Constructing a GMRF from the SPDE

We want to construct a (GMRF) solution to the SPDE.

The basic idea

Construct the solution as a finite basis expansion

\[ x(s) = \sum_k \psi_k(s) x_k, \]

with a suitable distribution for the weights \( \{x_k\} \).

A **stochastic weak solution** is given by weights, \( \{x_i\} \), such that the joint distribution fulfills

\[ \sum_i \left\langle \psi_j, (x^2 - \Delta)^{\alpha/2} \psi_i x_i \right\rangle \overset{D}{=} \left\langle \psi_j, \mathcal{W} \right\rangle \quad \forall j \]
Possible basis functions

Several possible basis functions exist:

- Harmonic functions give a spectral representation.
- Eigenfunctions to the covariance matrix give Karhunen-Loéve.
- A piecewise linear basis gives (almost) a GMRF.
Possible basis functions

Several possible basis functions exist:

- Harmonic functions give a spectral representation.
- Eigenfunctions to the covariance matrix give Karhunen-Loéve.
- A piecewise linear basis gives (almost) a GMRF.
Construction of the precision matrix

We have the SPDE

$$
\sum_i \left\langle \psi_j, (\kappa^2 - \Delta)^{\alpha/2} \psi_i x_i \right\rangle \overset{D}{=} \left\langle \psi_j, \mathcal{W} \right\rangle
$$

SPDE solution

The distribution of the weights is $x \in \mathcal{N} \left( 0, Q^{-1} \right)$

$$
Q_{1,\kappa} = K \quad Q_{2,\kappa} = KC^{-1}K \\
Q_{\alpha,\kappa} = KC^{-1}Q_{\alpha-2,\kappa}C^{-1}K, \quad \alpha = 3, 4, \ldots
$$

where

$$
C_{i,j} = \left\langle \psi_i, \psi_j \right\rangle, \\
K_{i,j} = \left\langle \psi_i, (\kappa^2 - \Delta) \psi_j \right\rangle = \kappa^2 \left\langle \psi_i, \psi_j \right\rangle - \left\langle \psi_i, \Delta \psi_j \right\rangle = \kappa^2 C_{i,j} + G_{i,j},
$$
Markov approximation

Since our basis functions have compact support the $C$ and $K$ matrices are sparse. However, $C^{-1}$ is dense making $Q_{2,x} = KC^{-1}K$ dense.

GMRF approximation

To obtain sparse precision matrices we replace the $C$-matrix with a diagonal matrix $\tilde{C}$ with elements

$$\tilde{C}_{i,i} = \langle \psi_i, 1 \rangle$$

The resulting approximation error is small (Bolin & Lindgren, 2009, and Simpson, Lindgren & Rue, 2010, tech.reps.)
Markov approximation

Since our basis functions have compact support the $C$ and $K$ matrices are sparse. However, $C^{-1}$ is dense making $Q_{2,x} = KC^{-1}K$ dense.

GMRF approximation

To obtain sparse precision matrices we replace the $C$-matrix with a diagonal matrix $\tilde{C}$ with elements

$$\tilde{C}_{i,i} = \langle \psi_i, 1 \rangle$$

The resulting approximation error is small (Bolin & Lindgren, 2009, and Simpson, Lindgren & Rue, 2010, tech.reps.)
Lattice on $\mathbb{R}^2$, step size $h$, regular triangulation

- **Order $\alpha = 1$:**
  \[
  \chi^2 h^2 \begin{bmatrix} 1 \\ \end{bmatrix} + \begin{bmatrix} -1 & -1 \\ 4 & -1 \\ \end{bmatrix} \approx -\Delta
  \]

- **Order $\alpha = 2$:**
  \[
  \chi^4 h^2 \begin{bmatrix} 1 \\ \end{bmatrix} + 2 \chi^2 \begin{bmatrix} -1 & -1 \\ 4 & -1 \\ \end{bmatrix} + \frac{1}{h^2} \begin{bmatrix} 1 & 2 & -8 & 2 \\ -8 & 20 & -8 & 1 \\ 2 & -8 & 2 & 1 \\ \end{bmatrix} \approx \Delta^2
  \]

- **Not restricted to regular lattices; easy on triangulations**
Example of neighbours for $\alpha = 2$
Observations

Point observations

\[ y(u) = x(u) + \varepsilon = \sum_{k} \psi_k(u)x_k + \varepsilon \]

Area observations

\[ y(M) = \int_{M} x(u) \, du + \varepsilon = \sum_{k} x_k \int_{M} \psi_k(u) \, du + \varepsilon \]
Observations

Point observations

\[ y(u) = x(u) + \varepsilon = \sum_k \psi_k(u)x_k + \varepsilon \]

Area observations

\[ y(M) = \int_M x(u) \, du + \varepsilon = \sum_k x_k \int_M \psi_k(u) \, du + \varepsilon \]
Conditional expectation

Introducing a matrix $A$ with rows

$$A_i. = \begin{bmatrix} \psi_1(u_i) & \cdots & \psi_N(u_i) \end{bmatrix} \quad \text{or} \quad A_i. = \begin{bmatrix} \int_{M_i} \psi_1(u) \, du & \cdots \end{bmatrix}$$

we can write the observation equation on matrix form as

$$y = Ax + \varepsilon \quad x \in N\left(\mu, Q^{-1}\right) \quad \varepsilon \in N\left(0, Q_\varepsilon^{-1}\right)$$

Kriging with GMRF

$$E(x|y) = \mu + Q_{x|y}^{-1} A^\top Q_\varepsilon (y - A\mu)$$

$$V(x|y) = Q_{x|y}^{-1} = \left(Q + A^\top Q_\varepsilon A\right)^{-1}$$
Conditional expectation

Introducing a matrix $A$ with rows

$$A_i. = \begin{bmatrix} \psi_1(u_i) & \cdots & \psi_N(u_i) \end{bmatrix} \quad \text{or} \quad A_i. = \begin{bmatrix} \int_{M_i} \psi_1(u) \, du & \cdots \end{bmatrix}$$

we can write the observation equation on matrix form as

$$y = Ax + \epsilon \quad x \in \mathcal{N} \left( \mu, Q^{-1} \right) \quad \epsilon \in \mathcal{N} \left( 0, Q_\epsilon^{-1} \right)$$

Kriging with GMRF

$$E(x|y) = \mu + Q_{x|y}^{-1} A^\top Q_\epsilon (y - A \mu)$$

$$V(x|y) = Q_{x|y}^{-1} = \left( Q + A^\top Q_\epsilon A \right)^{-1}$$
Example of an oscillating field

\[(\chi^2 \exp(i\pi\theta) - \Delta) (x_1(u) + ix_2(u)) = \mathcal{W}_1(u) + i\mathcal{W}_2(u)\]
Example of a non-stationary field

$$(\chi^2(u) - \Delta)(\tau(u)x(u)) = W(u)$$
Features and extensions, past, current and future work

- Just as easy for general manifolds, e.g. the globe, as for $\mathbb{R}^d$.
- Oscillating fields
- Non-stationary versions, such as

$$
(\chi^2(u) - \Delta)^{\alpha/2} (\tau(u)x(u)) = \mathcal{W}(u)
$$

$$
(\chi^2(u) + \nabla \cdot m(u) - \nabla \cdot M(u)\nabla)^{\alpha/2} x(u) = \tau(u)\mathcal{W}(u)
$$

(Includes the Sampson & Guttorp deformation method)
- Non-stationarity that depends on covariates.
- Multivariate versions
- Spatio-temporal versions
- Works well with INLA (current & next generation) but is also generally applicable