Spatial Matérn fields driven by non-Gaussian noise

David Bolin
Department of Mathematics and Mathematical Statistics
Umeå University
joint work with Jonas Wallin, Lund University

Montréal, August 4, 2013
Modeling spatial data

- Modeling spatial environmental data is a challenging problem:
  - Non-stationary covariance models are often needed.
  - Spatially irregular data on other domains than $\mathbb{R}^d$.
  - Large datasets.
  - Gaussianity can not always be assumed.

- A typical geostatistical model:
  \[ Y(s) = X(s) + \mathcal{E}(s) \]

  $X$ is Gaussian with some covariance function $r(s, t)$ and some mean, and $\mathcal{E}$ is Gaussian white noise.
The Matérn covariance function

- One of the most popular covariance functions for spatial data.
- Has three parameters, $\kappa$, $\nu$ and $\sigma$, and can be parametrized as

$$r(h) = \frac{2^{1-\nu}\sigma^2}{(4\pi)^{d/2}} \frac{(\kappa\|h\|)^\nu}{\Gamma(\nu + d/2)} K_{\nu}(\kappa\|h\|), \quad h \in \mathbb{R}^d,$$

where $K_{\nu}$ is a modified Bessel function of the second kind of order $\nu > 0$. 

David Bolin - david.bolin@math.umu.se
The SPDE approach

- Ordinary covariance-based representations are often too computationally expensive to use in spatial statistics.
- A Matérn field is a solution to the SPDE

\[(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X(s) = \sigma \mathcal{W}(s)\]

where \(\alpha = \nu + d/2\), \(\mathcal{W}(s)\) is Gaussian white noise and \(\Delta\) is the Laplacian (Lindgren et al 2011).

- Advantages with this representation:
  - Defines Matérn fields on general smooth manifolds.
  - Can be used to define non-stationary models by allowing the parameters in the SPDE to be spatially varying.
  - Facilitates computationally efficient approximations through Hilbert space approximations.

---

\[^1\text{An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion), Lindgren, Rue, and Lindström, JRSSB, 2011}\]
The SPDE approach

- Ordinary covariance-based representations are often too computationally expensive to use in spatial statistics.
- A Matérn field is a solution to the SPDE

\[(\kappa^2 - \Delta)^{\alpha/2} X(s) = \sigma \mathcal{W}(s)\]

where \(\alpha = \nu + d/2\), \(\mathcal{W}(s)\) is Gaussian white noise and \(\Delta\) is the Laplacian (Lindgren et al 2011).

- Advantages with this representation:
  - Defines Matérn fields on general smooth manifolds.
  - Can be used to define non-stationary models by allowing the parameters in the SPDE to be spatially varying.
  - Facilitates computationally efficient approximations through Hilbert space approximations.

---

1An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion), Lindgren, Rue, and Lindström, JRSSB, 2011
Hilbert space approximations

- We seek an approximate solution to the SPDE on the form

\[ \tilde{X}(s) = \sum_{i=1}^{n} w_i \varphi_i(s) \]

- The stochastic weights \( w \) are calculated by requiring the stochastic weak formulation of the SPDE to hold for only a specific set of test functions \( \{\psi_i, i = 1, \ldots, n\} \),

\[ \left\{ (\kappa^2 - \Delta)^{\frac{\alpha}{2}} X(\psi_i), i = 1, \ldots, n \right\} \overset{d}{=} \left\{ \phi W(\psi_i), i = 1, \ldots, n \right\}. \]

- For \( \alpha = 1, 2, \ldots \), \( w \) is mean-zero Gaussian with precision matrix \( Q_{\alpha} \). Let \( K = G + \kappa^2 C \), then

\[ Q_1 = K, \quad Q_2 = KC^{-1}K, \]

and for \( \alpha > 2 \):

\[ Q_{\alpha} = KC^{-1}Q_{\alpha-2}C^{-1}K. \]

Here \( C \) and \( G \) are matrices depending only on the basis.
non-Gaussian Matérn fields

Gaussian models are not appropriate in several applications

- Precipitation, vegetation, ocean waves, ozone ...

Goal: Formulate a model with Matérn covariances and non-Gaussian marginal distributions.

Idea: Replace the Gaussian noise with a non-Gaussian process $\dot{M}$:

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

- Existence of the solution can be shown for a general class of $L_2$-valued random measures $M$.
- Efficient estimation and simulation methods can be derived if $M$ is a type G Lévy process.
non-Gaussian Matérn fields

Gaussian models are not appropriate in several applications

- Precipitation, vegetation, ocean waves, ozone ...

Goal: Formulate a model with Matérn covariances and non-Gaussian marginal distributions.

Idea: Replace the Gaussian noise with a non-Gaussian process $\dot{M}$:

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

- Existence of the solution can be shown for a general class of $L_2$-valued random measures $M$.
- Efficient estimation and simulation methods can be derived if $M$ is a type G Lévy process.
The most well-known subclass of the type-G processes are the generalized hyperbolic processes. We need the distribution to be closed under convolution. For likelihood-based estimation and prediction and for interpretability during change of support. Only two special cases have this property: Generalized asymmetric Laplace (GAL) fields and Normal inverse Gaussian (NIG) fields. The Hilbert space approximation for these models is

\[ Kw \sim N(\gamma \tau a + \mu G, \text{diag}(G)) \]

where \( G_i \sim \Gamma(a_i \tau, 1) \) for GAL and \( G_i \sim IG(\nu^2 a_i, 2) \) for NIG.
Generalized hyperbolic fields

- The most well known subclass of the type-G processes are the generalized hyperbolic processes.
- We need the distribution to be closed under convolution.
  - for likelihood-based estimation and prediction
  - for interpretability during change of support
- Only two special cases have this property
  - Generalized asymmetric Laplace (GAL) fields
  - Normal inverse Gaussian (NIG) fields
- The Hilbert space approximation for these models is

\[
Kw \sim N(\gamma \tau a + \mu G, \text{diag}(G))
\]

where \( G_i \sim \Gamma(a_i \tau, 1) \) for GAL and \( G_i \sim IG(\nu^2 a_i, 2) \) for NIG.
A simulated example
Latent non-Gaussian models

- When using these models in practice, we want to allow for measurement noise and covariates for the mean.
- This is incorporated in a hierarchical model for the data \( \mathbf{y} \)

\[
\mathbf{y} = \mathbf{B}_x \mathbf{\beta}_x + \mathbf{A} \mathbf{w} + \mathbf{\varepsilon},
\]

\[
\mathbf{w} = \mathbf{K}_\alpha^{-1} \mathbf{C}^{-1} \left( \gamma \mathbf{t} \mathbf{a} + \mu \mathbf{G} + \sigma \sqrt{\mathbf{G}} \mathbf{Z} \right),
\]

\[
\mathbf{G} \sim \pi(\mathbf{G}).
\]

- Here \( \mathbf{B}_x \) contains the covariates, \( \mathbf{A} \) is the observation matrix, and \( \mathbf{\varepsilon} \) is the measurement noise.
- To recover the latent field \( X(s) \) at the measurement locations, one calculates

\[
X = \mathbf{B}_x \mathbf{\beta}_x + \mathbf{A} \mathbf{w}
\]
Parameter estimation

- We need to estimate the following model parameters: $\beta_x, \kappa, \sigma_\epsilon, \gamma, \mu, \sigma$ and $\tau$ (GAL) or $\nu$ (NIG).

- Using the SPDE formulation, one can do the parameter estimation in a likelihood framework using a Monte Carlo Expectation Maximization (MCEM) algorithm.

- The EM algorithm contains an E step where one calculates the function

$$Q \left( \Theta, \Theta^{(p)} \right) = E_{\Theta^{(p)}} \left[ \log \pi(Y, w, G|\Theta)|Y, \Theta^{(p-1)} \right],$$

where $\Theta^{(p)}$ denotes the current parameter estimate.

- In the M-step, $Q \left( \Theta, \Theta^{(p)} \right)$ is maximized with respect to $\Theta$. 

David Bolin - david.bolin@math.umu.se

Non-Gaussian spatial Matérn fields
We need to estimate the following model parameters: $\beta_x, \kappa, \sigma_\varepsilon, \gamma, \mu, \sigma$ and $\tau$ (GAL) or $\nu$ (NIG).

Using the SPDE formulation, one can do the parameter estimation in a likelihood framework using an Monte Carlo Expectation Maximization (MCEM) algorithm.

The EM algorithm contains an E step where one calculates the function

$$Q \left( \Theta, \Theta^{(p)} \right) = E_{\Theta^{(p)}} \left[ \log \pi (Y, w, G|\Theta)|Y, \Theta^{(p-1)} \right],$$

where $\Theta^{(p)}$ denotes the current parameter estimate.

In the M-step, $Q \left( \Theta, \Theta^{(p)} \right)$ is maximized with respect to $\Theta$. 

David Bolin - david.bolin@math.umu.se
Non-Gaussian spatial Matérn fields
Parameter estimation

- We cannot calculate $Q(\Theta, \Theta^{(p)})$ analytically, so we replace it with an MC estimate:

$$Q^{MC}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \log \pi(Y, G^{(i)}, w^{(i)} | \Theta)$$

- We want to minimize the number of MC samples in the E-step since each sample requires a Cholesky factorization of $\hat{Q}$.
- This can be achieved using Rao-Blackwellization by replacing $Q^{MC}$ with

$$Q^{RB}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{E} \left[ \log \pi(Y, G, w^{(i)} | \Theta) | \star \right]$$

- We do not need to save all MC samples, only a few sufficient statistics are needed.
- Spatial predictions are also obtained using Gibbs sampling.
Parameter estimation

- We cannot calculate $Q(\Theta, \Theta^{(p)})$ analytically, so we replace it with an MC estimate:

$$Q^{MC}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \log \pi(Y, G^{(i)}, w^{(i)}|\Theta)$$

- We want to minimize the number of MC samples in the E-step since each sample requires a Cholesky factorization of $\hat{Q}$.
- This can be achieved using Rao-Blackwellization by replacing $Q^{MC}$ with

$$Q^{RB}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{E} \left[ \log \pi(Y, G, w^{(i)}|\Theta)|\star \right],$$

- We do not need to save all MC samples, only a few sufficient statistics are needed.
- Spatial predictions are also obtained using Gibbs sampling.
Parameter estimation

- We cannot calculate $Q(\Theta, \Theta^{(p)})$ analytically, so we replace it with an MC estimate:

$$Q^{MC}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \log \pi(Y, G^{(i)}, w^{(i)} | \Theta)$$

- We want to minimize the number of MC samples in the E-step since each sample requires a Cholesky factorization of $\hat{Q}$.
- This can be achieved using Rao-Blackwellization by replacing $Q^{MC}$ with

$$Q^{RB}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{E} \left[ \log \pi(Y, G, w^{(i)} | \Theta) | \star \right],$$

- We do not need to save all MC samples, only a few sufficient statistics are needed.
- Spatial predictions are also obtained using Gibbs sampling.
Data from the National Climatic Data Center.

11,918 unique sites throughout the United States with a temporal coverage for the period 1895-1997.

We want to do spatial infilling to produce high-resolution maps.

Because of the skewness in the distribution for precipitation data, Gaussian models cannot be used directly.

We compare the non-Gaussian models to a Gaussian model using a square-root transformation of the data.
Data from the National Climatic Data Center.
- 11918 unique sites throughout the United States with a temporal coverage for the period 1895-1997.
- We want to do spatial infilling to produce high-resolution maps.
- Because of the skewness in the distribution for precipitation data, Gaussian models cannot be used directly.
- We compare the non-Gaussian models to a Gaussian model using a square-root transformation of the data.
Models and model selection

- Transformed Gaussian model:
  \[ \sqrt{Y_i} = \beta + X(s_i) + \varepsilon_i \]

- \( X \) is a mean-zero Gaussian Matérn field
- \( \varepsilon \) is Gaussian measurement noise.

- Models for data without transformation:
  \[ Y_i = \beta + X(s_i) + \varepsilon_i \]

- \( X \) is a Matérn field driven by Gaussian noise, NIG noise, or GAL noise.
- We compare the four models using cross-validation.
- The data is divided into 10 groups and the data from each group is predicted using the rest of the data.
- The residuals \( Y_k - E(Y_k|Y_{-k}) \) are calculated for each group.
- We also use the models to predict the variance of the residuals.
Models and model selection

- Transformed Gaussian model:
  \[ \sqrt{Y_i} = \beta + X(s_i) + \varepsilon_i \]

- \( X \) is a mean-zero Gaussian Matérn field
- \( \varepsilon \) is Gaussian measurement noise.

- Models for data without transformation:
  \[ Y_i = \beta + X(s_i) + \varepsilon_i \]

- \( X \) is a Matérn field driven by Gaussian noise, NIG noise, or GAL noise.

- We compare the four models using cross-validation.
- The data is divided into 10 groups and the data from each group is predicted using the rest of the data.
- The residuals \( Y_k - E(Y_k|Y_{-k}) \) are calculated for each group.
- We also use the models to predict the variance of the residuals.
Models and model selection

- Transformed Gaussian model:
  \[
  \sqrt{Y_i} = \beta + X(s_i) + \varepsilon_i
  \]

- \(X\) is a mean-zero Gaussian Matérn field
- \(\varepsilon\) is Gaussian measurement noise.
- Models for data without transformation:
  \[
  Y_i = \beta + X(s_i) + \varepsilon_i
  \]

- \(X\) is a Matérn field driven by Gaussian noise, NIG noise, or GAL noise.
- We compare the four models using cross-validation.
- The data is divided into 10 groups and the data from each group is predicted using the rest of the data.
- The residuals \(Y_k - E(Y_k|Y_{-k})\) are calculated for each group.
- We also use the models to predict the variance of the residuals.
### Cross validation results

<table>
<thead>
<tr>
<th></th>
<th>January</th>
<th></th>
<th>June</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tGauss</td>
<td>Gauss</td>
<td>NIG</td>
<td>GAL</td>
<td>tGauss</td>
<td>Gauss</td>
</tr>
<tr>
<td>V(r_s)</td>
<td>1.42</td>
<td>1.46</td>
<td>1.7</td>
<td><strong>1.38</strong></td>
<td>0.78</td>
<td>3.19</td>
</tr>
<tr>
<td>E(r)</td>
<td>3.6</td>
<td><strong>0.13</strong></td>
<td>-1.00</td>
<td>-0.50</td>
<td>2.12</td>
<td><strong>-0.05</strong></td>
</tr>
<tr>
<td>V(r)</td>
<td>1423</td>
<td>1415</td>
<td>2494</td>
<td>1436</td>
<td>1048</td>
<td>1046</td>
</tr>
<tr>
<td>E(</td>
<td>r</td>
<td>)</td>
<td>21.5</td>
<td>20.5</td>
<td>22.4</td>
<td><strong>20.3</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>es</td>
<td></td>
<td></td>
<td>2221</td>
<td>2190</td>
</tr>
<tr>
<td>CRPS</td>
<td><strong>15.6</strong></td>
<td>17.8</td>
<td>18.0</td>
<td>16.8</td>
<td><strong>16.30</strong></td>
<td>17.48</td>
</tr>
</tbody>
</table>

- r are the cross validation residuals.
- r_s are the residuals standardized by the predicted variance.
- ||es|| denotes the energy norm of r and CRPS denotes the continuous ranked probability score of r.
- Best results are marked with bold script.
- GAL and tGauss performs the best over all.
Kriging estimates $E(X|Y, \theta)$

January 1997

GAL

June 1997

transformed Gaussian

difference

David Bolin - david.bolin@math.umu.se  Non-Gaussian spatial Matérn fields
There is a large difference in the predicted standard deviation of the kriging errors. The reason is that the actual values of the measurements enter the variance of the transformed model.
Conclusions

- Non-Gaussian SPDE models are an interesting alternative to using transformed Gaussian models for non-Gaussian data.
- We can handle latent non-Gaussian models with measurement noise and partial observations.
- For precipitation data non-stationary models would improve the results.
- An R package is in development.

References:
- Bolin, Spatial Matérn fields driven by non-Gaussian noise, SJS 2013 (accepted with minor revision)
- Wallin and Bolin, Non-Gaussian Matérn fields with an application to precipitation modeling, arxiv preprint 2013.

Thanks for listening!
Conclusions

- Non-Gaussian SPDE models are an interesting alternative to using transformed Gaussian models for non-Gaussian data.
- We can handle latent non-Gaussian models with measurement noise and partial observations.
- For precipitation data non-stationary models would improve the results.
- An R package is in development.
- References:
  - Bolin, Spatial Matérn fields driven by non-Gaussian noise, SJS 2013 (accepted with minor revision)
  - Wallin and Bolin, Non-Gaussian Matérn fields with an application to precipitation modeling, arxiv preprint 2013.

Thanks for listening!
E-step

- We cannot calculate $Q(\Theta, \Theta^{(p)})$ analytically, so we replace it with an MC estimate:

$$Q^{MC}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \log \pi(Y, G^{(i)}, w^{(i)} | \Theta)$$

- $G^{(i)}, w^{(i)}$ are sampled using Gibbs sampling.
- $w^{(i)}|\{G^{(i-1)}, Y, \Theta^{(p-1)}\} \sim N(\hat{m}, \hat{Q}^{-1})$ where $\hat{m}$ and $\hat{Q}^{-1}$ depends on $\Theta, G^{(i)},$ and $Y$ ($\hat{Q}$ is sparse).
- $G^{(i)}|\{w^{i}, Y, \Theta^{(p-1)}\} \sim GIG(p, a, b)$ where

<table>
<thead>
<tr>
<th>GAL</th>
<th>NIG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>$h\tau - 1/2$</td>
</tr>
<tr>
<td>$a$</td>
<td>$(B_{\mu\mu})^2/\sigma^2 + 2$</td>
</tr>
<tr>
<td>$b$</td>
<td>$(K_\alpha w^{(i)} - B_{\gamma\gamma})^2/\sigma^2$</td>
</tr>
</tbody>
</table>
The log-likelihood can be written as

\[
\log \pi(Y, G^{(i)}, w^{(i)}|\Theta) = \log \pi(Y^{(i)}|w^{(i)}, \beta_x, \sigma_e) \\
+ \log \pi(w^{(i)}|G^{(i)}, \gamma, \mu, \sigma, \kappa) \\
+ \log \pi(G^{(i)}|\tau, \nu).
\]

the maximization over \(\Theta\) can be split into three independent steps maximizing \(\{\beta_x, \sigma_e\}\), \(\{\gamma, \mu, \sigma, \kappa\}\), and \(\{\tau, \nu\}\).

Closed form solutions can be found for \(\{\beta_x, \sigma_e\}\) and \(\{\tau, \nu\}\).

\(\kappa\) is found using numerical optimization and closed form solutions for \(\{\gamma, \mu, \sigma\}\) exists given \(\kappa\).
We want to minimize the number of MC samples in the E-step since each sample requires a Cholesky factorization of $\hat{Q}$.

Rao-Blackwellization is based on the fact that for any function $h$ and any two random variables $X$ and $Y$, one has that $\text{Var}[\mathbb{E}[h(X)|Y]] \leq \text{Var}[h(X)]$.

To apply Rao-Blackwellization to $Q^{MC}$, we note that

$$Q\left(\Theta, \Theta^{(p)}\right) = \mathbb{E}\left[ \mathbb{E}\left[ \log \pi(Y, G, w|\Theta) | \star \right] | Y, \Theta^{(p)} \right],$$

where $\star$ denotes $\{Y, w, \Theta^{(p)}\}$, the inner expectation is taken over $G$, and the outer expectation is taken over $w$. 
The Rao-Blackwellization is

\[
Q^{RB}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{E} \left[ \log \pi(Y, G, w^{(i)}|\Theta) | \star \right],
\]

\[
\mathbb{E} \left[ \log \pi(Y, G, w|\Theta) | \star \right]
\]

is

\[
- \frac{1}{2\sigma^2} \left( (K_\alpha w - B\beta)^T \mathbb{E}[I_{G-1} | \star] (K_\alpha w - B\beta) \\
+ \mu^T D^T \mathbb{E}[I_G | \star] D\mu \right) - \mathbb{E} \left[ \log \pi(G|\tau, \nu^2) | \star \right],
\]

If \( X \sim GIG(p, a, b) \), we have

\[
\mathbb{E}[X^\lambda] = \frac{K_{p+\lambda} \left( \sqrt{ab} \right)}{K_p \left( \sqrt{ab} \right)} \left( \frac{b}{a} \right)^{\lambda/2}, \quad \lambda \in \mathbb{R}
\]

\[
\mathbb{E}[\log(X)] = \log(\sqrt{a/b}) + \frac{\partial \log K_p}{\partial p} \left( \sqrt{ab} \right),
\]
Rao-Blackwellization

The Rao-Blackwellization is

\[ Q^{RB}(\Theta, \Theta^{(p)}) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{E} \left[ \log \pi(Y, G, w^{(i)} | \Theta) \right] , \]

\[ \mathbb{E}[\log \pi(Y, G, w | \Theta)] \]

is

\[ -\frac{1}{2\sigma^2} ((K\alpha w - B\beta)^T \mathbb{E}[I_G^{-1} | \star](K\alpha w - B\beta) + \mu^T D^T \mathbb{E}[I_G | \star] D\mu) - \mathbb{E}[\log \pi(G | \tau, \nu^2)] , \]

If \( X \sim GIG(p, a, b) \), we have

\[ \mathbb{E}[X^\lambda] = \frac{b/a^{\lambda/2} K_{p+\lambda} \left( \sqrt{ab} \right)}{K_p \left( \sqrt{ab} \right)} , \quad \lambda \in \mathbb{R} \]

\[ \mathbb{E}[\log(X)] = \log(\sqrt{a/b}) + \frac{\partial \log K_p}{\partial p} \left( \sqrt{ab} \right) , \]
Spatial prediction

- Assume that we want to do spatial prediction to a set of locations $s_1, \ldots, s_m$, and denote $X_i = X(s_i)$.
- Let $A_p$ have elements $A_{p,ij} = \varphi_i(s_j)$.
- The distribution for $A_p w$ is not known, so we estimate its mean and variance using MC simulation:

  $$
  E[X_i|Y, \Theta] \approx \frac{1}{k} \sum_{i=1}^{k} A_{p,i} w^{(i)},
  $$
  
  $$
  V[X_i|Y, \Theta] \approx \frac{1}{k-1} \sum_{i=1}^{k} (A_{p,i} w^{(i)} - E[A_{p,i} w|Y])^2,
  $$

  where $w^{(i)}$ are generated using the Gibbs-sampler and $A_{p,i}$ denotes the $i$:th row in $A_p$.
- Rao-Blackwellization is used to reduce the MC variance.
The SPDE representation is used for $X$.

- Piecewise linear basis functions induced by a triangulation of the domain.