Non-stationary nested SPDE models applied to global ozone data

David Bolin, Finn Lindgren

Centre for Mathematical Sciences
Lund University

Smögen, August 16, 2010
Modeling spatial data

- Modeling spatial environmental data is a challenging problem:
  - Non-stationary covariance models are often needed.
  - Computational efficiency is an increasingly important property.
- One such data set is the Total Ozone Mapping Spectrometer (TOMS) atmospheric ozone data.
- Collected by a TOMS instrument onboard the near-polar, Sun-synchronous orbiting satellite Nimbus-7.
During the sunlit portions of the satellite’s orbit, the instrument collected data in scans perpendicular to the orbital plane.

The data is spatially and temporally irregular, requires non-stationary covariance structures, and efficient computational techniques due to the massive amounts of data.
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

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

where $K_\nu$ is a modified Bessel function of the second kind of order $\nu > 0$. 
The Matérn covariance family

There are no oscillating covariance functions in the Matérn family.

Computational cost for spatial prediction of one location given $m$ observations is $\mathcal{O}(m^3)$. 

David Bolin - bolin@maths.lth.se
Nested SPDE models

A Matérn field is the solution to a stochastic partial differential equation (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 Laplace operator.

An easy generalization of the generating SPDE is

\[(\kappa^2 - \Delta)^{\alpha/2} X(s) = (b + B^\top \nabla) \mathcal{W}(s)\] (1)

where \(\nabla\) is the gradient and \(B\) is a vector.

Since the two operators are commutative, (1) is equivalent to the following system of nested SPDEs

\[(\kappa^2 - \Delta)^{\alpha/2} X_0(s) = \mathcal{W}(s),\]

\[X(s) = (b + B^\top \nabla) X_0(s).\]
Nested SPDE models

- To get a larger class of random fields, the orders of the operators can be increased.
- The class of nested SPDE models is defined as solutions to

\[
\left( \prod_{i=1}^{n_1} (\kappa_i^2 - \Delta)^{\frac{\alpha_i}{2}} \right) X(s) = \left( \prod_{i=1}^{n_2} (b_i + B_i^T \nabla) \right) W(s)
\]

for \( \kappa_i > 0, b_i \in \mathbb{R}, \alpha_i \in \mathbb{N}, \) and \( B_i \in \mathbb{R}^d \).

- Differentiations in several directions \( B_i \).
- 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 $\mathbb{R}^d$

- **Spectral density**

**Proposition 2.1** The spectral density for $X$ is given by

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

- **Sample path regularity**

**Proposition 2.2** $X$ has almost surely continuous sample functions if $2 \sum_{i=1}^{n_1} \alpha_i - 2n_2 > d$.

- **Sample path differentiability**

**Corollary 2.3** The $m$:th order directional derivative of $X$ has almost surely continuous sample functions given that $2 \sum_{i=1}^{n_1} \alpha_i - 2n_2 - d > m$. 

David Bolin - bolin@maths.lth.se

Nested SPDE models applied to global ozone data
Covariance functions

- Wide class of explicit covariance functions.
Realizations

- Corresponding realizations.
Hilbert space approximations

- For large data sets, a computationally efficient model representation is needed.
- For Matérn fields, Lindgren, Rue and Lindström (2010) derived a Hilbert space approximation technique for obtaining high-rank basis expansions $X(s) \approx \sum_{i=1}^{n} \varphi_i(s)v_i$.
- The method uses the stochastic weak formulation of the SPDE
  $$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X(s) = \sigma \mathcal{W}(s),$$
  with respect to some test functions to calculate the weights $v$.
- For certain bases, the weights form a Gaussian Markov Random Field:
  $$v \sim N(0, Q^{-1})$$
- The precision matrix, $Q$, is sparse.

David Bolin - bolin@maths.lth.se
Nested SPDE models applied to global ozone data
Hilbert space approximations

\[
\left( \prod_{i=1}^{n_1} (\kappa_i^2 - \Delta)^{\alpha_i / 2} \right) X(s) = \left( \prod_{i=1}^{n_2} (b_i + B_i^\top \nabla) \right) \mathcal{W}(s)
\]

- The method can be extended to give a similar basis expansion as for Matérn fields: \( X(s) \approx \sum_{i=1}^{n} \varphi_i(s) w_i \).
- \( w \) will not be a GMRF, but we have

\[
w \sim N(0, HQ^{-1}H^\top)
\]

where \( H \) and \( Q \) are sparse matrices.

- We have lost the Markov structure, but the computational benefits can be preserved using the fact that \( w = Hv \), where \( v \) is a GMRF.

- The computational cost for Kriging in \( \mathbb{R}^2 \) is \( \mathcal{O}(n^{3/2}) \) instead of \( \mathcal{O}(m^3) \) for covariance based models.
A non-stationary nested SPDE model

- We analyze the data from October 1, 1988.
- This data set contains approximately 180,000 measurements.
- Assume that the data are measurements of the latent ozone field $X(s)$ under Gaussian measurement noise.

$$Y(s) = X(s) + \epsilon(s), \quad \epsilon(s) \sim N(0, \sigma^2)$$

- We let $X(s)$ have some mean value $\mu(s)$, and, inspired by Jun and Stein (2008) assume that $Z(s) = X(s) - \mu(s)$ follows the simplest nested SPDE model:¹

$$(\kappa^2(s) - \Delta)Z_0(s) = W(s)$$

$$Z(s) = (b(s) + B(s)^\top \nabla)Z_0(s)$$

---

The Hilbert space approximation technique is extended and used for the class of non-stationary nested SPDE models.

A basis of 9002 piecewise linear functions induced by a triangulation of the earth is used.
Parameter estimation

We assume a constant but unknown mean $\mu(s) = \mu$ and use the following regression models for the varying parameters.

\[
b(s) = \exp \left( \sum_{k,m} b_{k,m} Y_{k,m}(s) \right), \quad B(s) = \sum_k b_k F_k(s)
\]

\[
\kappa^2(s) = \exp \left( \sum_{k,m} \kappa_{k,m} Y_{k,m}(s) \right),
\]

where $Y_{k,m}$ is the spherical harmonic of order $k$ and mode $m$ and \{\(F_k\)\} is a basis for vector fields on the sphere.

The regression parameters are estimated by numerical optimization of the marginal posterior $\pi(\kappa, \phi, b, \sigma^2 | Y)$. 

David Bolin - bolin@maths.lth.se

Nested SPDE models applied to global ozone data
The number of basis functions are chosen using AIC and BIC.

Best model has constant range parameter $\kappa(s)$ and spatially varying vector field $B(s)$ and variance scaling $b(s)$. 

$b(s)$

$|B(s)|$
Kriging estimate

David Bolin - bolin@maths.lth.se

Nested SPDE models applied to global ozone data
Standard errors

Nested SPDE models applied to global ozone data
Kriging residuals

David Bolin - bolin@maths.lth.se

Nested SPDE models applied to global ozone data
Conclusions

- The nested SPDE models is a wide en flexible model class that
  - easily can be made non-stationary.
  - is applicable to data on general smooth manifolds.
  - is computationally efficient.
- Good model fit for the ozone data
  - Non-stationary covariance structures necessary.
  - Likelihood estimation is not computationally feasible using covariance-based modeling.
  - The temporal aspect of the data should be modeled.
- Further work:
  - Spatio-temporal extensions.
  - Non-Gaussian versions.

Thanks for listening!
Bolin, D. and Lindgren, F. Spatial models generated by nested stochastic partial differential equations
*To appear in Annals of Applied Statistics*


Models for ozone data

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa^2(s)$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>$b(s)$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$B(s)$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>8</td>
<td>11</td>
<td>14</td>
<td>18</td>
<td>26</td>
<td>32</td>
<td>34</td>
<td>47</td>
<td>50</td>
<td>62</td>
<td>75</td>
<td>98</td>
</tr>
</tbody>
</table>

**Table:** Maximal orders of the spherical harmonics used in the bases for the different parameters and total number of covariance parameters in the different models for $X(s)$. The actual number of basis functions for $\kappa^2(s)$ and $b(s)$ are given by $(ord + 1)^2$, and for $B(s)$, the actual number is $2(ord + 1)^2 - 2$, where $ord$ is the maximal order indicated in the table.
AIC and BIC

Varying $\kappa^2$ and $b$

Varying $B$ and $b$

Varying $\kappa^2$, $B$, and $b$

All models

David Bolin - bolin@maths.lth.se

Nested SPDE models applied to global ozone data