On generating a flexible class of anisotropic spatial models using Gaussian predictive processes

Sujit Sahu

http://www.soton.ac.uk/~sks/
Co-author: Sabyasachi Mukhopadhyay

Rio, ISI 2015
On generating a flexible class of anisotropic spatial models using Gaussian predictive processes

Sujit Sahu

http://www.soton.ac.uk/~sks/
Co-author: Sabyasachi Mukhopadhyay

Rio, ISI 2015
Models for the covariance function

Types of anisotropy

How can predictive processes be used to generate anisotropic models?

Examples:
- Modelling scallop abundance data
- Modelling UK air pollution data for five years

Discussion
Models for the covariance function

Types of anisotropy

How can predictive processes be used to generate anisotropic models?

Examples:
- Modelling scallop abundance data
- Modelling UK air pollution data for five years

Discussion
Models for the covariance function

Types of anisotropy

How can predictive processes be used to generate anisotropic models?

Examples:
- Modelling scallop abundance data
- Modelling UK air pollution data for five years

Discussion
Models for the covariance function

Types of anisotropy

How can predictive processes be used to generate anisotropic models?

Examples:
- Modelling scallop abundance data
- Modelling UK air pollution data for five years

Discussion
Models for the covariance function

Types of anisotropy

How can predictive processes be used to generate anisotropic models?

Examples:
- Modelling scallop abundance data
- Modelling UK air pollution data for five years

Discussion
Models for the covariance function

Types of anisotropy

How can predictive processes be used to generate anisotropic models?

Examples:
- Modelling scallop abundance data
- Modelling UK air pollution data for five years

Discussion
Models for the covariance function

Types of anisotropy

How can predictive processes be used to generate anisotropic models?

Examples:
- Modelling scallop abundance data
- Modelling UK air pollution data for five years

Discussion
Example: UK air pollution data modelling

- Map of 323 local authorities in England for which we have health outcome data.
- Red dots define the corners of the 12 km square grid cells where we have AQUM output.
- Blue dots represent the 142 AURN air-quality monitoring sites.
Modelling setup: Suppose that we have random variables $Y(s_1), \ldots, Y(s_n)$ where each $s_i$ denotes a particular location.

In general, consider a real-valued spatial process $Y(s)$, where $s \in \mathbb{D}$ and $\mathbb{D}$ is the study region, usually a sub-space of $\mathbb{R}^2$, England in the above example!

There are 3 main concepts in spatial statistics (in the Matheron School):
1. Stationarity
2. Variogram
3. Isotropy

No formal model based inference for $Y(s)$ yet.
What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(s_1), \ldots, Y(s_n)$ where each $s_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(s)$, where $s \in \mathbb{D}$ and $\mathbb{D}$ is the study region, usually a sub-space of $\mathbb{R}^2$, England in the above example!

- There are 3 main concepts in spatial statistics (in the Matheron School):
  1. Stationarity
  2. Variogram
  3. Isotropy

- No formal model based inference for $Y(s)$ yet.
What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(s_1), \ldots, Y(s_n)$ where each $s_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(s)$, where $s \in D$ and $D$ is the study region, usually a sub-space of $\mathbb{R}^2$, England in the above example!

- There are 3 main concepts in spatial statistics (in the Matheron School):
  1. Stationarity
  2. Variogram
  3. Isotropy

- No formal model based inference for $Y(s)$ yet.
What is anisotropy?

Modelling setup: Suppose that we have random variables $Y(s_1), \ldots, Y(s_n)$ where each $s_i$ denotes a particular location.

In general, consider a real-valued spatial process $Y(s)$, where $s \in \mathbb{D}$ and $\mathbb{D}$ is the study region, usually a sub-space of $\mathbb{R}^2$, England in the above example!

There are 3 main concepts in spatial statistics (in the Matheron School):

1. Stationarity
2. Variogram
3. Isotropy

No formal model based inference for $Y(s)$ yet.
What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(s_1), \ldots, Y(s_n)$ where each $s_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(s)$, where $s \in \mathbb{D}$ and $\mathbb{D}$ is the study region, usually a sub-space of $\mathbb{R}^2$, England in the above example!

- There are 3 main concepts in spatial statistics (in the Matheron School):
  1. Stationarity
  2. Variogram
  3. Isotropy

- No formal model based inference for $Y(s)$ yet.
What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(s_1), \ldots, Y(s_n)$ where each $s_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(s)$, where $s \in \mathcal{D}$ and $\mathcal{D}$ is the study region, usually a sub-space of $\mathbb{R}^2$, England in the above example!

- There are 3 main concepts in spatial statistics (in the Matheron School):
  1. Stationarity
  2. Variogram
  3. Isotropy

- No formal model based inference for $Y(s)$ yet.
What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(s_1), \ldots, Y(s_n)$ where each $s_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(s)$, where $s \in \mathbb{D}$ and $\mathbb{D}$ is the study region, usually a sub-space of $\mathbb{R}^2$, England in the above example!

- There are 3 main concepts in spatial statistics (in the Matheron School):
  1. Stationarity
  2. Variogram
  3. Isotropy

- No formal model based inference for $Y(s)$ yet.
Suppose our spatial process has a mean, $\mu(s) = E(Y(s))$, and that the variance of $Y(s)$ exists for all $s$.

The process is said to be strictly stationary (also called strongly stationary) if, for any given $n \geq 1$, any set of $n$ sites $s_1, \ldots, s_n$ and any $h$ the distribution of $Y(s_1), \ldots, Y(s_n)$ is the same as that of $Y(s_1 + h), \ldots, Y(s_n + h)$.

A less restrictive condition is given by weak stationarity (also called second-order stationarity): A process is weakly stationary if $\mu(s) = \mu$ and $\text{Cov}(Y(s), Y(s + h)) = C(h)$ for all $h$ such that $s$ and $s + h$ both lie in $D$. 
• Suppose our spatial process has a mean, \( \mu(s) = E(Y(s)) \), and that the variance of \( Y(s) \) exists for all \( s \).

• The process is said to be strictly stationary (also called strongly stationary) if, for any given \( n \geq 1 \), any set of \( n \) sites \( s_1, \ldots, s_n \) and any \( h \) the distribution of \( Y(s_1), \ldots, Y(s_n) \) is the same as that of \( Y(s_1 + h), \ldots, Y(s_n + h) \).

• A less restrictive condition is given by weak stationarity (also called second-order stationarity): A process is weakly stationary if \( \mu(s) = \mu \) and \( \text{Cov}(Y(s), Y(s + h)) = C(h) \) for all \( h \) such that \( s \) and \( s + h \) both lie in \( D \).
Suppose our spatial process has a mean, $\mu(s) = E(Y(s))$, and that the variance of $Y(s)$ exists for all $s$.

The process is said to be strictly stationary (also called strongly stationary) if, for any given $n \geq 1$, any set of $n$ sites $s_1, \ldots, s_n$ and any $h$ the distribution of $Y(s_1), \ldots, Y(s_n)$ is the same as that of $Y(s_1 + h), \ldots, Y(s_n + h)$.

A less restrictive condition is given by weak stationarity (also called second-order stationarity): A process is weakly stationary if $\mu(s) = \mu$ and $\text{Cov}(Y(s), Y(s + h)) = C(h)$ for all $h$ such that $s$ and $s + h$ both lie in $D$. 
Weak stationarity says that the covariance between the values of the process at any two locations $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ can be summarized by a covariance function $C(h)$ (sometimes called a covariogram), and this function depends only on the separation vector $\mathbf{h}$.

Note that with all variances assumed to exist, strong stationarity implies weak stationarity.

The converse is not true in general, but it does hold for Gaussian processes.
Notes on Stationarity

Weak stationarity says that the covariance between the values of the process at any two locations \( s \) and \( s + h \) can be summarized by a covariance function \( C(h) \) (sometimes called a covariogram), and this function depends only on the separation vector \( h \).

Note that with all variances assumed to exist, strong stationarity implies weak stationarity.

The converse is not true in general, but it does hold for Gaussian processes.
Weak stationarity says that the covariance between the values of the process at any two locations $s$ and $s + h$ can be summarized by a covariance function $C(h)$ (sometimes called a covariogram), and this function depends only on the separation vector $h$.

Note that with all variances assumed to exist, strong stationarity implies weak stationarity.

The converse is not true in general, but it does hold for Gaussian processes.
Semi-Variogram is defined as:

\[
\gamma(h) = \frac{1}{2} \text{var}(Y(s + h) - Y(s))
\]

Simple calculation yields

\[
2\gamma(h) = 2 [C(0) - C(h)]
\]

So given the covariance function \(C(\cdot)\) we can determine the semivariogram.

But the converse is not true, we can add \(\pm a\) to \(C(\cdot)\) and obtain the same \(\gamma(\cdot)\).
Semi-Variogram is defined as:

\[ \gamma(h) = \frac{1}{2} \text{var}(Y(s + h) - Y(s)) \]

Simple calculation yields

\[ 2\gamma(h) = 2 [C(0) - C(h)] \]

So given the covariance function \( C(\cdot) \) we can determine the semivariogram.

But the converse is not true, we can add \( \pm a \) to \( C(\cdot) \) and obtain the same \( \gamma(\cdot) \).
Variogram

- Semi-Variogram is defined as:
  \[
  \gamma(h) = \frac{1}{2} \text{var}(Y(s + h) - Y(s))
  \]

- Simple calculation yields
  \[
  2\gamma(h) = 2 [C(0) - C(h)]
  \]

- So given the covariance function $C(\cdot)$ we can determine the semivariogram.

- But the converse is not true, we can add $\pm a$ to $C(\cdot)$ and obtain the same $\gamma(\cdot)$. 
Semi-Variogram is defined as:

$$\gamma(h) = \frac{1}{2} \text{var}(Y(s + h) - Y(s))$$

Simple calculation yields

$$2\gamma(h) = 2 [C(0) - C(h)]$$

So given the covariance function $C(\cdot)$ we can determine the semivariogram.

But the converse is not true, we can add $\pm a$ to $C(\cdot)$ and obtain the same $\gamma(\cdot)$. 
If the semivariogram $\gamma(h)$ depends upon the separation vector only through its length $||h||$ then we say that the process is isotropic.

For an isotropic process, $\gamma(h)$ is a real-valued function of a univariate argument, and can be written as $\gamma(||h||)$.

Isotropic processes are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for $\gamma(\cdot)$. 

Sujit Sahu
If the semivariogram $\gamma(h)$ depends upon the separation vector only through its length $||h||$ then we say that the process is isotropic.

For an isotropic process, $\gamma(h)$ is a real-valued function of a univariate argument, and can be written as $\gamma(||h||)$.

Isotropic processes are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for $\gamma(\cdot)$. 

Sujit Sahu
If the semivariogram $\gamma(h)$ depends upon the separation vector only through its length $||h||$ then we say that the process is isotropic.

For an isotropic process, $\gamma(h)$ is a real-valued function of a univariate argument, and can be written as $\gamma(||h||)$.

Isotropic processes are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for $\gamma(\cdot)$. 
If the semivariogram $\gamma(h)$ depends upon the separation vector only through its length $||h||$ then we say that the process is isotropic.

For an isotropic process, $\gamma(h)$ is a real-valued function of a univariate argument, and can be written as $\gamma(||h||)$.

Isotropic processes are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for $\gamma(\cdot)$. 
The Matérn correlation function is given by:

\[ C(t; \phi, \nu) = \frac{1}{2^{\nu-1} \Gamma(\nu)} (2 \sqrt{\nu} \phi t)^\nu K_\nu(2 \sqrt{\nu} \phi t), \quad \phi > 0, \nu > 0, \]

where \( \Gamma(\nu) \) is the standard gamma function, \( K_\nu \) is the modified Bessel function of second kind with order \( \nu \), and \( t = ||h|| \) is the distance between two sites.

- The parameter \( \phi \) controls the rate of decay of the correlation as the distance \( t \) increases.
- The parameter \( \nu \) controls smoothness of the random field \( Y(s) \).
  - \( \nu = 1/2 \implies C(t) = \sigma^2 \exp(-\phi t), \quad t > 0; \) Exponential Covariance Function
  - \( \nu = 3/2, \quad C(t) = \sigma^2 (1 + \phi t) \exp(-\phi t), \quad t > 0. \)
  - \( \nu \to \infty \implies C(t) = \sigma^2 \exp(-\phi^2 t^2), \quad t > 0; \) Gaussian
This is by far the most popular choice for modelers.

The correlation between two points distance $t$ apart is $\exp(-\phi t)$.

The *effective range*, $t_0$, as the distance at which this correlation becomes negligible, equal to 0.05.

Setting

$$\exp(-\phi t_0) = 0.05$$

$$\implies t_0 = -\log(0.05)/\phi$$

$$\implies t_0 \approx 3/\phi$$

since $\log(0.05) \approx -3$. 
Recall $\gamma(h) = \gamma(||h||) = C(0) - C(||h||)$.

So $\gamma(0) = 0$. But often there are micro-scale variation or measurement error even at very small distances.

To tackle that we define the nugget

$$\tau^2 \equiv \lim_{t \to 0^+} \gamma(t).$$

This introduces a discontinuity at 0 for the covariogram $\gamma(t)$. 
What happens to $\gamma(t)$ when $t \to \infty$?

This asymptotic value is called the sill.

In our notation sill is given by $\tau^2 + \sigma^2$.

The sill minus the nugget, $\sigma^2$, is called the partial sill.

The effective range is the smallest distance for which the semivariogram achieves the asymptotic sill.
Three closed form Matérn covariograms:

1. **Exponential**: \( \gamma(t) = \tau^2 + \sigma^2(1 - \exp(-\phi t)) \).
2. **Gaussian**: \( \gamma(t) = \tau^2 + \sigma^2(1 - \exp(-\phi^2 t^2)) \).
3. **Matérn with \( \nu = 1.5 \)**: \( \gamma(t) = \tau^2 + \sigma^2(1 - (1 + \phi t) \exp(-\phi t)) \).
Anisotropy is opposite of isotropy. For example,

- If the variogram depends on angle it is angular anisotropy.
- Similarly, sill and range anisotropy.
- Geometric anisotropy is obtained by stretching of an isotropic model: \( \gamma(h) = \gamma_0(\sqrt{h'Qh}) \) where \( \gamma_0(\cdot) \) is isotropic and \( Q \) is a positive definite matrix.
- Zonal anisotropy. Variogram only depends on some components of the vector \( h \). Also called stratified anisotropy.

See Chapter 2 of Chilès and Delfiner (2012).
How can we generate anisotropic processes?

- Answer depends on what type of anisotropy (e.g. geometric or zonal) we want.

- It is difficult to decide the type of anisotropy when all we have available is a realisation \( y(s_1), \ldots, y(s_n) \) along with the locations \( s_1, \ldots, s_n \).

- Hence it is difficult to specify a flexible covariance function \( C() \).

- Further problem arises due to the positive definiteness requirement of the implied covariance matrix of any \( n \) realisations \( Y(s) \).
Answer depends on what type of anisotropy (e.g. geometric or zonal) we want.

It is difficult to decide the type of anisotropy when all we have available is a realisation $y(s_1), \ldots, y(s_n)$ along with the locations $s_1, \ldots, s_n$.

Hence it is difficult to specify a flexible covariance function $C()$.

Further problem arises due to the positive definiteness requirement of the implied covariance matrix of any $n$ realisations $Y(s)$. 
Answer depends on what type of anisotropy (e.g. geometric or zonal) we want.

It is difficult to decide the type of anisotropy when all we have available is a realisation \( y(s_1), \ldots, y(s_n) \) along with the locations \( s_1, \ldots, s_n \).

Hence it is difficult to specify a flexible covariance function \( C() \).

Further problem arises due to the positive definiteness requirement of the implied covariance matrix of any \( n \) realisations \( Y(s) \).
Answer depends on what type of anisotropy (e.g. geometric or zonal) we want.

It is difficult to decide the type of anisotropy when all we have available is a realisation $y(s_1), \ldots, y(s_n)$ along with the locations $s_1, \ldots, s_n$.

Hence it is difficult to specify a flexible covariance function $C()$.

Further problem arises due to the positive definiteness requirement of the implied covariance matrix of any $n$ realisations $Y(s)$.
What’s available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O’hagan (2003).


- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).


What’s available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O’hagan (2003).


- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).


What’s available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O’hagan (2003).


- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).


What’s available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O’hagan (2003).


- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).


What’s available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O’hagan (2003).


- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).


Our main idea

- To use Gaussian predictive process to generate anisotropy.

- Suppose there are $m$ knot-locations $s^*_1, \ldots, s^*_m$. We shall choose these and $m$ later.

- Assume a latent Gaussian process $w(s)$ with realisations $w^* = (w(s^*_1), \ldots, w(s^*_m))$.

- At any other location $s$, define $\tilde{w}(s) = E[w(s)|w^*]$.

- This $\tilde{w}(s)$ defines a flexible anisotropic valid spatial process.
Our main idea

- To use Gaussian predictive process to generate anisotropy.

- Suppose there are \( m \) knot-locations \( s_1^*, \ldots, s_m^* \). We shall choose these and \( m \) later.

- Assume a latent Gaussian process \( w(s) \) with realisations \( w^* = (w(s_1^*), \ldots, w(s_m^*)) \).

- At any other location \( s \), define \( \tilde{w}(s) = E[w(s)|w^*] \).

- This \( \tilde{w}(s) \) defines a flexible anisotropic valid spatial process.
To use Gaussian predictive process to generate anisotropy.

Suppose there are $m$ knot-locations $s^*_1, \ldots, s^*_m$. We shall choose these and $m$ later.

Assume a latent Gaussian process $w(s)$ with realisations 
\[ w^* = (w(s^*_1), \ldots, w(s^*_m)). \]

At any other location $s$, define $\tilde{w}(s) = E[w(s)|w^*].$

This $\tilde{w}(s)$ defines a flexible anisotropic valid spatial process.
Our main idea

- To use Gaussian predictive process to generate anisotropy.

- Suppose there are $m$ knot-locations $s_1^*, \ldots, s_m^*$. We shall choose these and $m$ later.

- Assume a latent Gaussian process $w(s)$ with realisations $w^* = (w(s_1^*), \ldots, w(s_m^*))$.

- At any other location $s$, define $\tilde{w}(s) = E[w(s)|w^*]$.

- This $\tilde{w}(s)$ defines a flexible anisotropic valid spatial process.
Our main idea

- To use Gaussian predictive process to generate anisotropy.

- Suppose there are \( m \) knot-locations \( \mathbf{s}^*_1, \ldots, \mathbf{s}^*_m \). We shall choose these and \( m \) later.

- Assume a latent Gaussian process \( w(\mathbf{s}) \) with realisations \( w^* = (w(\mathbf{s}^*_1), \ldots, w(\mathbf{s}^*_m)) \).

- At any other location \( \mathbf{s} \), define \( w(\tilde{\mathbf{s}}) = E [w(\mathbf{s})|w^*] \).

- This \( w(\tilde{\mathbf{s}}) \) defines a flexible anisotropic valid spatial process.
An example

- Consider $\mathbb{D}$ to be $\mathbb{R}^1$, let $m = 1$ and $s_1^* = 0$, i.e. the single knot at the origin.

- Assume exponential covariance function with decay parameter $\phi > 0$ and variance 1.

- Then $\tilde{w}(s) = \exp(-\phi|s|) w^*(0)$ where $w^*(0) \sim N(0, 1)$.

- Now $\text{Cov}(\tilde{w}(s), \tilde{w}(s'))$ will depend not only on $|s - s'|$ but on both $s$ and $s'$.

- Further complexity is introduced by taking $m > 1$, and varying the positioning of the knots $s_1^*, \ldots, s_m^*$ at random or according to a specific clustering mechanism.
Consider $\mathbb{D}$ to be $\mathbb{R}^1$, let $m = 1$ and $s_1^* = 0$, i.e. the single knot at the origin.

Assume exponential covariance function with decay parameter $\phi > 0$ and variance 1.

Then $\tilde{w}(s) = \exp(-\phi|s|) w^*(0)$ where $w^*(0) \sim N(0, 1)$.

Now $\text{Cov}(\tilde{w}(s), \tilde{w}(s'))$ will depend not only on $|s - s'|$ but on both $s$ and $s'$.

Further complexity is introduced by taking $m > 1$, and varying the positioning of the knots $s_1^*, \ldots, s_m^*$ at random or according to a specific clustering mechanism.
Consider $\mathbb{D}$ to be $\mathbb{R}^1$, let $m = 1$ and $s^*_1 = 0$, i.e. the single knot at the origin.

Assume exponential covariance function with decay parameter $\phi > 0$ and variance 1.

Then $\tilde{w}(s) = \exp(-\phi|s|) w^*(0)$ where $w^*(0) \sim N(0, 1)$.

Now $\text{Cov}(\tilde{w}(s), \tilde{w}(s'))$ will depend not only on $|s - s'|$ but on both $s$ and $s'$.

Further complexity is introduced by taking $m > 1$, and varying the positioning of the knots $s^*_1, \ldots, s^*_m$ at random or according to a specific clustering mechanism.
Consider $D$ to be $\mathbb{R}^1$, let $m = 1$ and $s^*_1 = 0$, i.e. the single knot at the origin.

Assume exponential covariance function with decay parameter $\phi > 0$ and variance 1.

Then $\tilde{w}(s) = \exp(-\phi|s|) w^*(0)$ where $w^*(0) \sim N(0, 1)$.

Now $\text{Cov}(\tilde{w}(s), \tilde{w}(s'))$ will depend not only on $|s - s'|$ but on both $s$ and $s'$.

Further complexity is introduced by taking $m > 1$, and varying the positioning of the knots $s^*_1, \ldots, s^*_m$ at random or according to a specific clustering mechanism.
Consider $\mathbb{D}$ to be $\mathbb{R}^1$, let $m = 1$ and $s_1^* = 0$, i.e. the single knot at the origin.

Assume exponential covariance function with decay parameter $\phi > 0$ and variance 1.

Then $\tilde{w}(s) = \exp(-\phi|s|) w^*(0)$ where $w^*(0) \sim N(0, 1)$.

Now $\text{Cov}(\tilde{w}(s), \tilde{w}(s'))$ will depend not only on $|s - s'|$ but on both $s$ and $s'$.

Further complexity is introduced by taking $m > 1$, and varying the positioning of the knots $s_1^*, \ldots, s_m^*$ at random or according to a specific clustering mechanism.
For any \( s \), \( \tilde{w}(s) = c^*(s)^T S_{w^*}^{-1} w^* \) where \( c^*(s) \) denotes the \( m \times 1 \) correlation vector between \( w(s) \) and \( w^* \), given by
\[
\left( C(|s - s^*_1|), \ldots, C(|s - s^*_m|) \right)^T
\]
and \( S_{w^*} \) is the correlation matrix of \( w^* \).

Consider two locations \( s \) and \( s + h \). Now:

\[
2\tilde{\gamma}(s, h) = \text{Var} [\tilde{w}(s) - \tilde{w}(s + h)]
\]
\[
= \mathbb{E} [\tilde{w}(s) - \tilde{w}(s + h)]^2
\]
\[
= (c^*(s) - c^*(s + h))^T S_{w^*}^{-1} (c^*(s) - c^*(s + h)).
\]

Depends on both \( s \) and \( h \).
For any \( s \), \( \tilde{w}(s) = c^*(s)^T S_{w^*}^{-1} w^* \) where \( c^*(s) \) denotes the \( m \times 1 \) correlation vector between \( w(s) \) and \( w^* \), given by

\[
\begin{pmatrix} C(|s - s_1^*|), \ldots, C(|s - s_m^*|) \end{pmatrix}^T
\]

and \( S_{w^*} \) is the correlation matrix of \( w^* \).

Consider two locations \( s \) and \( s + h \). Now:

\[
2\tilde{\gamma}(s, h) = \text{Var} [\tilde{w}(s) - \tilde{w}(s + h)]
\]
\[
= E [\tilde{w}(s) - \tilde{w}(s + h)]^2
\]
\[
= (c^*(s) - c^*(s + h))^T S_{w^*}^{-1} (c^*(s) - c^*(s + h)).
\]

Depends on both \( s \) and \( h \).
For any $s$, $\tilde{w}(s) = c^*(s)^T S_{w^*}^{-1} w^*$ where $c^*(s)$ denotes the $m \times 1$ correlation vector between $w(s)$ and $w^*$, given by

$\left( C(|s - s_1^*|), \ldots, C(|s - s_{ml}^*|) \right)^T$ and $S_{w^*}$ is the correlation matrix of $w^*$.

Consider two locations $s$ and $s + h$. Now:

$$2\tilde{\gamma}(s, h) = \text{Var} [\tilde{w}(s) - \tilde{w}(s + h)]$$
$$= E [\tilde{w}(s) - \tilde{w}(s + h)]^2$$
$$= (c^*(s) - c^*(s + h))^T S_{w^*}^{-1} (c^*(s) - c^*(s + h)).$$

Depends on both $s$ and $h$. 
Exploring correlation structure with $\tilde{\gamma}(s, h)$.

- **Is $\tilde{\gamma}(s, h)$ a legitimate semivariogram?**

  - No! It's not an even function of $h$, i.e. $\tilde{\gamma}(s, h) \neq \tilde{\gamma}(s, -h)$.

  - We still can treat this as a function of $|h|$ and study its properties for varying $s$ and $h$ and the knots.

  - We fix a central location $s^{**}$, assumed to be the centroid and then calculate distance between $s^{**}$ and $s^{**} + h$. 

Is $\tilde{\gamma}(s, h)$ a legitimate semivariogram?

No! It's not an even function of $h$, i.e. $\tilde{\gamma}(s, h) \neq \tilde{\gamma}(s, -h)$.

We still can treat this as a function of $|h|$ and study its properties for varying $s$ and $h$ and the knots.

We fix a central location $s^{**}$, assumed to be the centroid and then calculate distance between $s^{**}$ and $s^{**} + h$. 
Exploring correlation structure with $\tilde{\gamma}(s, h)$.

- Is $\tilde{\gamma}(s, h)$ a legitimate semivariogram?

- No! It's not an even function of $h$, i.e. $\tilde{\gamma}(s, h) \neq \tilde{\gamma}(s, -h)$.

- We still can treat this as a function of $|h|$ and study its properties for varying $s$ and $h$ and the knots.

- We fix a central location $s^{**}$, assumed to be the centroid and then calculate distance between $s^{**}$ and $s^{**} + h$. 

Sujit Sahu
Is $\tilde{\gamma}(s, h)$ a legitimate semivariogram?

No! It's not an even function of $h$, i.e. $\tilde{\gamma}(s, h) \neq \tilde{\gamma}(s, -h)$.

We still can treat this as a function of $|h|$ and study its properties for varying $s$ and $h$ and the knots.

We fix a central location $s^{**}$, assumed to be the centroid and then calculate distance between $s^{**}$ and $s^{**} + h$. 
How do we choose the knots?

- Space filling?
- Clustering?
- Random placement?

In general, assume a point process, i.e. a random distribution \( \pi(S^*_m) \) for \( s^*_1, \ldots, s^*_m \).

But how about \( m \)? We can assume that to be unknown as well.

Assume \( \pi(m) \) for \( m \) and think of \( \pi(S^*_m) \) conditional on \( m \).
How do we choose the knots?

- Space filling?

- Clustering?

- Random placement?

- In general, assume a point process, i.e. a random distribution $\pi(S^*_m)$ for $s^*_1, \ldots, s^*_m$.

- But how about $m$? We can assume that to be unknown as well.

- Assume $\pi(m)$ for $m$ and think of $\pi(S^*_m)$ conditional on $m$. 

Sujit Sahu
How do we choose the knots?

- Space filling?
- Clustering?
- Random placement?

In general, assume a point process, i.e. a random distribution $\pi(S^*_m)$ for $s^*_1, \ldots, s^*_m$.

But how about $m$? We can assume that to be unknown as well.

Assume $\pi(m)$ for $m$ and think of $\pi(S^*_m)$ conditional on $m$. 

Sujit Sahu
How do we choose the knots?

- Space filling?
- Clustering?
- Random placement?

- In general, assume a point process, i.e. a random distribution $\pi(S^*_m)$ for $s^*_1, \ldots, s^*_m$.

- But how about $m$? We can assume that to be unknown as well.

- Assume $\pi(m)$ for $m$ and think of $\pi(S^*_m)$ conditional on $m$. 

Sujit Sahu
How do we choose the knots?

- Space filling?
- Clustering?
- Random placement?

In general, assume a point process, i.e. a random distribution $\pi(S^*_{m})$ for $s^*_1, \ldots, s^*_m$.

- But how about $m$? We can assume that to be unknown as well.

- Assume $\pi(m)$ for $m$ and think of $\pi(S^*_{m})$ conditional on $m$. 
How do we choose the knots?

- Space filling?
- Clustering?
- Random placement?

In general, assume a point process, i.e. a random distribution $\pi(S_m^*)$ for $s^*_1, \ldots, s^*_m$.

But how about $m$? We can assume that to be unknown as well.

Assume $\pi(m)$ for $m$ and think of $\pi(S_m^*)$ conditional on $m$. 
Return to $\tilde{\gamma}(s, h)$

- It is random if $m$ and $S_m^*$ are.

- We can use the expected value. But that is not available in closed form.

- So, we use Monte Carlo to estimate.

- We generate an $m_\ell$ from $\pi(m)$ and and a set of $m_\ell$ random knots $S_{m_l}^*$ from $\pi(S_{m_\ell}^*)$.

- Conditional on these values, evaluate the inner expectation $E \left[ (\tilde{w}(s) - \tilde{w}(s + h))^2 | m_l, S_{m_l}^* \right]$.

- Finally, we approximate $\tilde{\gamma}(s, h)$ by

$$\frac{1}{2L} \sum_{\ell=1}^{L} E \left[ (\tilde{w}(s) - \tilde{w}(s + h))^2 | m_l, S_{m_l}^* \right].$$
It is random if \( m \) and \( S_m^* \) are.

We can use the expected value. But that is not available in closed form.

So, we use Monte Carlo to estimate.

We generate an \( m_\ell \) from \( \pi(m) \) and a set of \( m_\ell \) random knots \( S_{m_\ell}^* \) from \( \pi(S_{m_\ell}^*) \).

Conditional on these values, evaluate the inner expectation
\[
E \left[ \{\tilde{w}(s) - \tilde{w}(s + h)\}^2 \mid m_\ell, S_{m_\ell}^* \right].
\]

Finally, we approximate \( \tilde{\gamma}(s, h) \) by
\[
\frac{1}{2L} \sum_{\ell=1}^{L} E \left[ \{\tilde{w}(s) - \tilde{w}(s + h)\}^2 \mid m_\ell, S_{m_\ell}^* \right].
\]
It is random if \( m \) and \( S^*_m \) are.

We can use the expected value. But that is not available in closed form.

So, we use Monte Carlo to estimate.

We generate an \( m_\ell \) from \( \pi(m) \) and and a set of \( m_\ell \) random knots \( S^*_{m_\ell} \) from \( \pi(S^*_{m_\ell}) \).

Conditional on these values, evaluate the inner expectation
\[
E \left[ \{ \tilde{w}(s) - \tilde{w}(s + h) \}^2 \mid m_\ell, S^*_{m_\ell} \right].
\]

Finally, we approximate \( \tilde{\gamma}(s, h) \) by
\[
\frac{1}{2L} \sum_{\ell=1}^{L} E \left[ \{ \tilde{w}(s) - \tilde{w}(s + h) \}^2 \mid m_\ell, S^*_{m_\ell} \right].
\]
It is random if $m$ and $S_m^*$ are.

We can use the expected value. But that is not available in closed form.

So, we use Monte Carlo to estimate.

We generate an $m_\ell$ from $\pi(m)$ and and a set of $m_\ell$ random knots $S_{m_\ell}^*$ from $\pi(S_{m_\ell}^*)$.

Conditional on these values, evaluate the inner expectation
\[
E \left[ \{\tilde{\nu}(s) - \tilde{\nu}(s + h)\}^2 | m_\ell, S_{m_\ell}^* \right].
\]

Finally, we approximate $\tilde{\gamma}(s, h)$ by
\[
\frac{1}{2L} \sum_{\ell=1}^{L} E \left[ \{\tilde{\nu}(s) - \tilde{\nu}(s + h)\}^2 | m_\ell, S_{m_\ell}^* \right].
\]
It is random if \( m \) and \( S_m^* \) are.

We can use the expected value. But that is not available in closed form.

So, we use Monte Carlo to estimate.

We generate an \( m_\ell \) from \( \pi(m) \) and and a set of \( m_\ell \) random knots \( S_{m_\ell}^* \) from \( \pi(S_{m_\ell}^*) \).

Conditional on these values, evaluate the inner expectation

\[
\mathbb{E} \left[ \{ \tilde{\nu}(s) - \tilde{\nu}(s + h) \}^2 \mid m_\ell, S_{m_\ell}^* \right].
\]

Finally, we approximate \( \tilde{\gamma}(s, h) \) by

\[
\frac{1}{2L} \sum_{\ell=1}^{L} \mathbb{E} \left[ \{ \tilde{\nu}(s) - \tilde{\nu}(s + h) \}^2 \mid m_\ell, S_{m_\ell}^* \right].
\]
It is random if $m$ and $S_m^*$ are.

We can use the expected value. But that is not available in closed form.

So, we use Monte Carlo to estimate.

We generate an $m_\ell$ from $\pi(m)$ and and a set of $m_\ell$ random knots $S_{m_\ell}^*$ from $\pi(S_{m_\ell}^*)$.

Conditional on these values, evaluate the inner expectation $E \left[ \{\tilde{w}(s) - \tilde{w}(s + h)\}^2 | m_\ell, S_{m_\ell}^* \right]$.

Finally, we approximate $\tilde{\gamma}(s, h)$ by

$$\frac{1}{2L} \sum_{\ell=1}^L E \left[ \{\tilde{w}(s) - \tilde{w}(s + h)\}^2 | m_\ell, S_{m_\ell}^* \right].$$
Consider $\mathbb{D} = [-1, 1]$ in one and $\mathbb{D} = [-1, 1] \times [-1, 1]$ in two dimensions.

1. **Space filling with $m = 25$**
2. Complete Spatial Randomness (CSR) with $m = 25$.
3. All the knots clustered within the central quarter: $[-0.25, 0.25]$ with $m = 25$.
4. CSR but with $m$ following uniform between 1 to 25.
Consider $\mathbb{D} = [-1, 1]$ in one and $\mathbb{D} = [-1, 1] \times [-1, 1]$ in two dimensions.

1. Space filling with $m = 25$
2. Complete Spatial Randomness (CSR) with $m = 25$.
3. All the knots clustered within the central quarter: $[-0.25, 0.25]$ with $m = 25$.
4. CSR but with $m$ following uniform between 1 to 25.
Consider $\mathbb{D} = [-1, 1]$ in one and $\mathbb{D} = [-1, 1] \times [-1, 1]$ in two dimensions.

1. Space filling with $m = 25$

2. Complete Spatial Randomness (CSR) with $m = 25$.

3. All the knots clustered within the central quarter: $[-0.25, 0.25]$ with $m = 25$.

4. CSR but with $m$ following uniform between 1 to 25.
Consider $D = [-1, 1]$ in one and $D = [-1, 1] \times [-1, 1]$ in two dimensions.

1. Space filling with $m = 25$
2. Complete Spatial Randomness (CSR) with $m = 25$.
3. All the knots clustered within the central quarter: $[-0.25, 0.25]$ with $m = 25$.
4. CSR but with $m$ following uniform between 1 to 25.
Solid line: Semi variogram in positive and dotted line is in the negative direction.

Compare with the figure for isotropic correlation structure shown before.

Effect of the space filling knots are seen in the top left.

Knots clustered in a smaller sub-region is seen in the bottom left panel.

The correlation curves become ‘more’ smooth when knots are placed at random.
Semivariogram plots against radial distance.

The shape of the variogram depends on where the knots are placed.

Shows angular anisotropy as well.
Semivariogram plots against angle.
There may not be any sill.
Hence, the GPP can generate very flexible anisotropic processes.
Consider the scallop data set from Ecker and Gelfand.
Top left: Theoretical contours for an isotropic model.
Top right: Empirical Semivariogram Contour (ESC) plot of the observed data.
Bottom left: Theoretical SC plot for a fixed space filling knot design with 100 knots.
Bottom right: TSC plot for a random design.
Hierarchical modeling

- **Basic Model:**

\[ Y(s) = x^T(s)\beta + \tilde{w}(s) + \epsilon(s) \]

- The residual is partitioned into two pieces: one spatial, \( \tilde{w}(s) \), and one non-spatial, \( \epsilon(s) \).

- \( \tilde{w}(s) \) is a non-stationary and anisotropic Gaussian process depending on the parameters \( \sigma^2_w \), decay parameter \( \phi \), and smoothness \( \nu \) and the number and positioning of the knot locations.

- \( \epsilon(s) \) adds the nugget \( (\tau^2) \) effect.

- \( \tilde{w}(s) \) reduces dimension if \( n > m \). Otherwise, it may increase it to achieve flexibility.
Hierarchical modeling

**Basic Model:**

\[ Y(s) = x^T(s)\beta + \tilde{w}(s) + \epsilon(s) \]

The residual is partitioned into two pieces: one spatial, \( \tilde{w}(s) \), and one non-spatial, \( \epsilon(s) \).

\( \tilde{w}(s) \) is a non-stationary and anisotropic Gaussian process depending on the parameters \( \sigma_w^2 \), decay parameter \( \phi \), and smoothness \( \nu \) and the number and positioning of the knot locations.

\( \epsilon(s) \) adds the nugget \( (\tau^2) \) effect.

\( \tilde{w}(s) \) reduces dimension if \( n > m \). Otherwise, it may increase it to achieve flexibility.
Hierarchical modeling

- **Basic Model:**

\[
Y(s) = x^T(s)\beta + \tilde{w}(s) + \epsilon(s)
\]

- The residual is partitioned into two pieces: one spatial, \(\tilde{w}(s)\), and one non-spatial, \(\epsilon(s)\).

- \(\tilde{w}(s)\) is a non-stationary and anisotropic Gaussian process depending on the parameters \(\sigma^2_w\), decay parameter \(\phi\), and smoothness \(\nu\) and the number and positioning of the knot locations.

- \(\epsilon(s)\) adds the nugget \((\tau^2)\) effect.

- \(\tilde{w}(s)\) reduces dimension if \(n > m\). Otherwise, it may increase it to achieve flexibility.
Hierarchical modeling

- Basic Model:

\[ Y(s) = x^T(s)\beta + \tilde{w}(s) + \epsilon(s) \]

- The residual is partitioned into two pieces: one spatial, \( \tilde{w}(s) \), and one non-spatial, \( \epsilon(s) \).

- \( \tilde{w}(s) \) is a non-stationary and anisotropic Gaussian process depending on the parameters \( \sigma^2_w \), decay parameter \( \phi \), and smoothness \( \nu \) and the number and positioning of the knot locations.

- \( \epsilon(s) \) adds the nugget \( (\tau^2) \) effect.

- \( \tilde{w}(s) \) reduces dimension if \( n > m \). Otherwise, it may increase it to achieve flexibility.
Hierarchical modeling

- Basic Model:

\[ Y(s) = x^T(s)\beta + \tilde{w}(s) + \epsilon(s) \]

- The residual is partitioned into two pieces: one spatial, \( \tilde{w}(s) \), and one non-spatial, \( \epsilon(s) \).

- \( \tilde{w}(s) \) is a non-stationary and anisotropic Gaussian process depending on the parameters \( \sigma^2_w \), decay parameter \( \phi \), and smoothness \( \nu \) and the number and positioning of the knot locations.

- \( \epsilon(s) \) adds the nugget \( \tau^2 \) effect.

- \( \tilde{w}(s) \) reduces dimension if \( n > m \). Otherwise, it may increase it to achieve flexibility.
pure error term; model is not perfectly spatial;

\( \tau^2 \) and \( \sigma_w^2 \) are known as variance components.

measurement error or replication variability causing discontinuity in spatial surface \( Y(s) \);

microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.
pure error term; model is not perfectly spatial;

\( \tau^2 \) and \( \sigma_w^2 \) are known as variance components.

measurement error or replication variability causing discontinuity in spatial surface \( Y(s) \);

microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.
Interpretations attached to $\epsilon(s)$

- pure error term; model is not perfectly spatial;

- $\tau^2$ and $\sigma^2_w$ are known as variance components.

- measurement error or replication variability causing discontinuity in spatial surface $Y(s)$;

- microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.
Interpretations attached to $\epsilon(s)$

- pure error term; model is not perfectly spatial;

- $\tau^2$ and $\sigma^2_w$ are known as variance components.

- measurement error or replication variability causing discontinuity in spatial surface $Y(s)$;

- microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.
Conditional on $m$ assume a non-homogenous Poisson Process model for the knots $S_m^*$. 

$$
\pi(S_m^*) = (\lambda(D))^{-m} \prod_{j=1}^{m} \lambda(s_j),
$$

where $\lambda(D) = \int_D \lambda(s)ds$ and $\lambda(s)$ is a given intensity function which is constant for CSR.

The logarithm of the full posterior distribution, $
\log(\pi(m, S_m^*, w(S_m^*), \theta|z)),$ is given by:

$$
\propto -\frac{n}{2} \log(\tau^2) \\
- \frac{1}{2\tau^2} \sum_{i=1}^{n} \left( z(s_i) - x(s_i)^T \beta - \tilde{w}(s_i) \right)^2 \\
- m \log(\lambda(D)) + \sum_{j=1}^{m} \log(\lambda(s_j)) \\
- \frac{m}{2} \log(\sigma_w^2) - \frac{1}{2} \log|S_w| - \frac{1}{2\sigma_w^2} (w^*)^T S_w^{-1} w \\
+ \log(\pi(\theta))
$$

where $\theta = (\beta, \tau^2, \sigma_w^2, \nu, \phi)^T$ and $\pi(\theta)$ denotes the prior.
Conditional on $m$ assume a non-homogenous Poisson Process model for the knots $S_m^*$.

$$\pi(S_m^*) = (\lambda(D))^{-m} \prod_{j=1}^{m} \lambda(s_j),$$

where $\lambda(D) = \int_D \lambda(s) ds$ and $\lambda(s)$ is a given intensity function which is constant for CSR.

The logarithm of the full posterior distribution, $\log(\pi(m, S_m^*, w(S_m^*), \theta|z))$, is given by:

$$\propto -\frac{n}{2} \log(\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^{n} \left( z(s_i) - x(s_i)^T \beta - \tilde{w}(s_i) \right)^2 - m \log(\lambda(D)) + \sum_{j=1}^{m} \log(\lambda(s_j)) - \frac{m}{2} \log(\sigma_w^2) - \frac{1}{2} \log|S_w| - \frac{1}{2\sigma_w^2} (w^*)^T S_w^{-1} w + \log(\pi(\theta))$$

where $\theta = (\beta, \tau^2, \sigma_w^2, \nu, \phi)^T$ and $\pi(\theta)$ denotes the prior.
Informativeness: \( \pi(\beta) \) can be a flat (improper)

Without nugget, \( \tau^2 \), can’t identify both \( \sigma^2_w \) and \( \phi \) (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.

With \( \tau^2 \), then \( \phi \) and at least one of \( \sigma^2_w \) and \( \tau^2 \) require informative priors.

Assume a Matérn covariance function with known \( \nu \). If the prior on \( \beta, \sigma^2_w, \phi \) is of the form \( \frac{\pi(\phi)}{(\sigma^2_w)^{a+1}} \) with \( \pi(\cdot) \) uniform, then we get improper posterior if \( a < \frac{1}{2} \).

Shows the problem with using IG(\( \epsilon, \epsilon \)) priors for \( \sigma^2_w \) – nearly improper. Safer is IG(\( a, b \)) with \( a \geq 1 \).
Informativeness: $\pi(\beta)$ can be a flat (improper)

Without nugget, $\tau^2$, can’t identify both $\sigma^2_w$ and $\phi$ (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.

With $\tau^2$, then $\phi$ and at least one of $\sigma^2_w$ and $\tau^2$ require informative priors.

Assume a Matérn covariance function with known $\nu$. If the prior on $\beta, \sigma^2_w, \phi$ is of the form $\frac{\pi(\phi)}{(\sigma^2_w)^{a+1}}$ with $\pi(\cdot)$ uniform, then we get improper posterior if $a < \frac{1}{2}$.

Shows the problem with using IG($\epsilon, \epsilon$) priors for $\sigma^2_w$ – nearly improper. Safer is IG($a, b$) with $a \geq 1$. 
Informativeness: $\pi(\beta)$ can be a flat (improper)

Without nugget, $\tau^2$, can’t identify both $\sigma^2_w$ and $\phi$ (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.

With $\tau^2$, then $\phi$ and at least one of $\sigma^2_w$ and $\tau^2$ require informative priors.

Assume a Matérn covariance function with known $\nu$. If the prior on $\beta, \sigma^2_w, \phi$ is of the form $\frac{\pi(\phi)}{(\sigma^2_w)^{a+1}}$ with $\pi(\cdot)$ uniform, then we get improper posterior if $a < \frac{1}{2}$.

Shows the problem with using IG($\varepsilon, \varepsilon$) priors for $\sigma^2_w$ – nearly improper. Safer is IG($a, b$) with $a \geq 1$. 

Sujit Sahu
Informativeness: $\pi(\beta)$ can be a flat (improper)

Without nugget, $\tau^2$, can’t identify both $\sigma_w^2$ and $\phi$ (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.

With $\tau^2$, then $\phi$ and at least one of $\sigma_w^2$ and $\tau^2$ require informative priors.

Assume a Matérn covariance function with known $\nu$. If the prior on $\beta, \sigma_w^2, \phi$ is of the form $\frac{\pi(\phi)}{(\sigma_w^2)^{a+1}}$ with $\pi(\cdot)$ uniform, then we get improper posterior if $a < \frac{1}{2}$.

Shows the problem with using IG$(\epsilon, \epsilon)$ priors for $\sigma_w^2$ – nearly improper. Safer is IG$(a, b)$ with $a \geq 1$. 

Sujit Sahu
Prediction of $Y(s_0)$ at a new site $s_0$ with associated covariates $x_0 = x(s_0)$.

Predictive distribution $\pi(y(s_0)|y) = \int \pi(y(s_0)|m, S_m^*, w^*, \theta, y)\pi(m, S_m^*, w^*, \theta|y)dm dS_m^* dw^* d\theta$

$\Longrightarrow$ easy Monte Carlo estimate using composition with Gibbs draws $\theta^{(1)}, \ldots, \theta^{(G)}$:

For each $\theta^{(g)}$ drawn from $\pi(\theta|y, X)$ draw $Y(s_0)^{(g)}$ from $f(y(s_0|y, \theta^{(g)}, X, x_0))$. 

Sujit Sahu
Results for NO$_2$ modelling and validation.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSPE</th>
<th>MAPE</th>
<th>Bias</th>
<th>RBias</th>
<th>N Cov(%)</th>
<th>G</th>
<th>P</th>
<th>G+P</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQUM</td>
<td>26.96</td>
<td>19.45</td>
<td>16.93</td>
<td>0.34</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Kriging</td>
<td>20.12</td>
<td>15.26</td>
<td>3.48</td>
<td>0.07</td>
<td>96.13</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Linear</td>
<td>13.66</td>
<td>10.45</td>
<td>–1.35</td>
<td>–0.03</td>
<td>99.83</td>
<td>105733</td>
<td>8002</td>
<td>113735</td>
</tr>
<tr>
<td>GP</td>
<td>15.14</td>
<td>12.39</td>
<td>2.48</td>
<td>0.05</td>
<td>98.32</td>
<td>2918</td>
<td>18594</td>
<td>21511</td>
</tr>
<tr>
<td>$M_1$</td>
<td>13.54</td>
<td>10.23</td>
<td>2.84</td>
<td>0.06</td>
<td>98.33</td>
<td>3828</td>
<td>51684</td>
<td>55512</td>
</tr>
<tr>
<td>$M_2$</td>
<td>10.78</td>
<td>8.17</td>
<td>1.12</td>
<td>0.02</td>
<td>99.16</td>
<td>4897</td>
<td>62710</td>
<td>67607</td>
</tr>
<tr>
<td>$M_3$</td>
<td>13.29</td>
<td>10.10</td>
<td>2.32</td>
<td>0.05</td>
<td>98.83</td>
<td>4765</td>
<td>61756</td>
<td>66521</td>
</tr>
<tr>
<td>$M_4$</td>
<td>14.72</td>
<td>10.93</td>
<td>4.51</td>
<td>0.09</td>
<td>94.34</td>
<td>5000</td>
<td>62603</td>
<td>67603</td>
</tr>
</tbody>
</table>

Table: Model choice measures for NO$_2$. Fitted $n = 4822$, validation $n = 601 \approx 12.4\%$. $M_1, \ldots, M_4$ are models with fixed range parameters at 3000, 600, 300 and 100 kilometres respectively. G and P are goodness-of-fit and Penalty according to the predictive model choice criteria (Gelfand and Ghosh, 1998).
Results for $O_3$ modelling and validation.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSPE</th>
<th>MAPE</th>
<th>Bias</th>
<th>RBias</th>
<th>NCoV(%)</th>
<th>G</th>
<th>P</th>
<th>G+P</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQUM</td>
<td>16.06</td>
<td>13.28</td>
<td>-8.79</td>
<td>-0.14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kriging</td>
<td>8.95</td>
<td>7.08</td>
<td>-3.44</td>
<td>-0.06</td>
<td>93.31</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Linear</td>
<td>9.01</td>
<td>7.29</td>
<td>-0.60</td>
<td>-0.01</td>
<td>99.45</td>
<td>76384</td>
<td>2010</td>
<td>78394</td>
</tr>
<tr>
<td>GP</td>
<td>9.60</td>
<td>7.90</td>
<td>2.36</td>
<td>0.03</td>
<td>100.0</td>
<td>1149</td>
<td>5992</td>
<td>7141</td>
</tr>
<tr>
<td>$M_1$</td>
<td>6.77</td>
<td>5.25</td>
<td>0.72</td>
<td>0.01</td>
<td>94.50</td>
<td>1387</td>
<td>18107</td>
<td>19494</td>
</tr>
<tr>
<td>$M_2$</td>
<td>6.53</td>
<td>5.12</td>
<td>0.72</td>
<td>0.01</td>
<td>96.70</td>
<td>1371</td>
<td>18716</td>
<td>20087</td>
</tr>
<tr>
<td>$M_3$</td>
<td>6.68</td>
<td>5.17</td>
<td>0.56</td>
<td>0.009</td>
<td>95.33</td>
<td>1366</td>
<td>18870</td>
<td>20236</td>
</tr>
<tr>
<td>$M_4$</td>
<td>8.09</td>
<td>5.98</td>
<td>0.41</td>
<td>0.006</td>
<td>96.42</td>
<td>1285</td>
<td>19139</td>
<td>20424</td>
</tr>
<tr>
<td>$M_5$</td>
<td>6.53</td>
<td>5.12</td>
<td>0.72</td>
<td>0.01</td>
<td>96.70</td>
<td>1370</td>
<td>18706</td>
<td>20076</td>
</tr>
<tr>
<td>$M_6$</td>
<td>6.82</td>
<td>5.27</td>
<td>0.74</td>
<td>0.01</td>
<td>93.95</td>
<td>1388</td>
<td>17941</td>
<td>19329</td>
</tr>
</tbody>
</table>

Table: Model choice measures for $O_3$. Fitting $n = 3269$, validation $n = 364$. $M_1, ..., M_4$ are models with fixed range parameters as before and $M_5$ and $M_6$ are models with uniform and gamma prior distributions for the decay parameter $\phi$. 

Sujit Sahu
Results for PM$_{10}$ modelling and validation.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSPE</th>
<th>MAPE</th>
<th>Bias</th>
<th>RBias</th>
<th>NCov(%)</th>
<th>G</th>
<th>P</th>
<th>G+P</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQUM</td>
<td>11.85</td>
<td>10.32</td>
<td>10.27</td>
<td>0.51</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Kriging</td>
<td>3.82</td>
<td>2.99</td>
<td>0.09</td>
<td>0.005</td>
<td>88.60</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Linear</td>
<td>5.65</td>
<td>4.69</td>
<td>0.32</td>
<td>0.02</td>
<td>89.23</td>
<td>91873</td>
<td>91973</td>
<td>183846</td>
</tr>
<tr>
<td>GP</td>
<td>5.71</td>
<td>4.72</td>
<td>1.10</td>
<td>0.05</td>
<td>85.34</td>
<td>721</td>
<td>3928</td>
<td>4649</td>
</tr>
<tr>
<td>$M_1$</td>
<td>3.29</td>
<td>2.55</td>
<td>–0.03</td>
<td>–0.002</td>
<td>89.70</td>
<td>595</td>
<td>7617</td>
<td>8212</td>
</tr>
<tr>
<td>$M_2$</td>
<td>3.45</td>
<td>2.65</td>
<td>–0.14</td>
<td>–0.007</td>
<td>89.03</td>
<td>585</td>
<td>8023</td>
<td>8608</td>
</tr>
<tr>
<td>$M_3$</td>
<td>3.56</td>
<td>2.72</td>
<td>–0.24</td>
<td>–0.01</td>
<td>89.70</td>
<td>554</td>
<td>7755</td>
<td>8309</td>
</tr>
<tr>
<td>$M_4$</td>
<td>4.81</td>
<td>3.39</td>
<td>–0.13</td>
<td>–0.007</td>
<td>91.36</td>
<td>539</td>
<td>8331</td>
<td>8870</td>
</tr>
<tr>
<td>$M_5$</td>
<td>3.46</td>
<td>2.67</td>
<td>–0.20</td>
<td>–0.01</td>
<td>91.02</td>
<td>574</td>
<td>7779</td>
<td>8353</td>
</tr>
<tr>
<td>$M_6$</td>
<td>3.28</td>
<td>2.55</td>
<td>–0.04</td>
<td>–0.002</td>
<td>89.70</td>
<td>593</td>
<td>7614</td>
<td>8207</td>
</tr>
</tbody>
</table>

**Table:** Model choice measures for PM$_{10}$. Fitting $n = 2463$, validation $n = 301$. $M_1$, ..., $M_4$ are models with fixed range parameters as before and $M_5$ and $M_6$ are models with uniform and gamma prior distributions for the decay parameter $\phi$. 

Sujit Sahu
Results for PM$_{2.5}$ modelling and validation.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSPE</th>
<th>MAPE</th>
<th>Bias</th>
<th>RBias</th>
<th>NCov(%)</th>
<th>G</th>
<th>P</th>
<th>G+P</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQUM</td>
<td>7.26</td>
<td>5.41</td>
<td>4.77</td>
<td>0.34</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Kriging</td>
<td>2.81</td>
<td>1.92</td>
<td>–0.76</td>
<td>–0.05</td>
<td>82.53</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Linear</td>
<td>5.17</td>
<td>4.24</td>
<td>–0.43</td>
<td>–0.03</td>
<td>81.45</td>
<td>46590</td>
<td>46679</td>
<td>93268</td>
</tr>
<tr>
<td>GP</td>
<td>5.18</td>
<td>4.35</td>
<td>1.51</td>
<td>0.11</td>
<td>81.45</td>
<td>595</td>
<td>5466</td>
<td>6061</td>
</tr>
<tr>
<td>$M_1$</td>
<td>2.72</td>
<td>1.93</td>
<td>–0.52</td>
<td>–0.04</td>
<td>83.11</td>
<td>330</td>
<td>2765</td>
<td>3095</td>
</tr>
<tr>
<td>$M_2$</td>
<td>2.81</td>
<td>1.98</td>
<td>–0.62</td>
<td>–0.04</td>
<td>82.68</td>
<td>318</td>
<td>2819</td>
<td>3137</td>
</tr>
<tr>
<td>$M_3$</td>
<td>2.91</td>
<td>2.05</td>
<td>–0.56</td>
<td>–0.04</td>
<td>82.25</td>
<td>304</td>
<td>2883</td>
<td>3186</td>
</tr>
<tr>
<td>$M_4$</td>
<td>4.50</td>
<td>3.01</td>
<td>–0.57</td>
<td>–0.04</td>
<td>84.84</td>
<td>289</td>
<td>3126</td>
<td>3415</td>
</tr>
<tr>
<td>$M_5$</td>
<td>2.82</td>
<td>1.98</td>
<td>–0.62</td>
<td>–0.04</td>
<td>83.11</td>
<td>318</td>
<td>2821</td>
<td>3139</td>
</tr>
<tr>
<td>$M_6$</td>
<td><strong>2.70</strong></td>
<td>1.92</td>
<td>–0.77</td>
<td>–0.05</td>
<td>83.54</td>
<td>314</td>
<td>2651</td>
<td><strong>2966</strong></td>
</tr>
</tbody>
</table>

Table: Model choice measures for PM$_{2.5}$. Fitting $n = 1820$, validation $n = 231$. $M_1, ..., M_4$ are models with fixed range parameters as before and $M_5$ and $M_6$ are models with uniform and gamma prior distributions for the decay parameter $\phi$. 
Figure: Annual map of ozone levels in 2011
Conclusions

1. We have proposed flexible anisotropic models for spatial and spatio-temporal data.
2. We can generate all sorts of anisotropy: sill, nugget and zonal anisotropy.
3. Spatio-temporal models perform better out of sample predictions as we have illustrated with air pollution data.
4. A separate talk/paper discusses air pollution modelling and links pollution to health outcome data.
1. We have proposed flexible anisotropic models for spatial and spatio-temporal data.

2. We can generate all sorts of anisotropy: sill, nugget and zonal anisotropy.

3. Spatio-temporal models perform better out of sample predictions as we have illustrated with air pollution data.

4. A separate talk/paper discusses air pollution modelling and links pollution to health outcome data.
We have proposed flexible anisotropic models for spatial and spatio-temporal data.

We can generate all sorts of anisotropy: sill, nugget and zonal anisotropy.

Spatio-temporal models perform better out of sample predictions as we have illustrated with air pollution data.

A separate talk/paper discusses air pollution modelling and links pollution to health outcome data.
1. We have proposed flexible anisotropic models for spatial and spatio-temporal data.

2. We can generate all sorts of anisotropy: sill, nugget and zonal anisotropy.

3. Spatio-temporal models perform better out of sample predictions as we have illustrated with air pollution data.

4. A separate talk/paper discusses air pollution modelling and links pollution to health outcome data.