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

Sujit Sahu

University of Southampton

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

UNIVERSITY OF Southampton

http://www.soton.ac.uk/∼sks/

Co-author: Sabyasachi Mukhopadhyay

Rio, ISI 2015

Outline

- 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

Outline

- 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

Outline

- 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

Outline

- 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

Outline

- 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

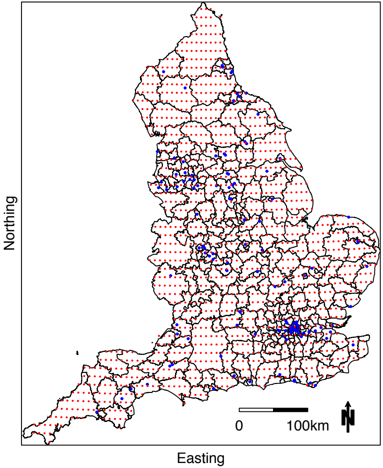
Outline

- 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

Outline

- 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.

What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ where each $\mathbf{s}_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(\mathbf{s})$, where $\mathbf{s} \in \mathbb{D}$ 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 $Y(\mathbf{s})$ yet.

What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ where each $\mathbf{s}_i$ denotes a particular location.
- In general, consider a real-valued spatial process $Y(\mathbf{s})$, where $\mathbf{s} \in \mathbb{D}$ 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 $Y(\mathbf{s})$ yet.

What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ where each $\mathbf{s}_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(\mathbf{s})$, where $\mathbf{s} \in \mathbb{D}$ 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 $Y(\mathbf{s})$ yet.

What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ where each $\mathbf{s}_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(\mathbf{s})$, where $\mathbf{s} \in \mathbb{D}$ 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 $Y(\mathbf{s})$ yet.

What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ where each $\mathbf{s}_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(\mathbf{s})$, where $\mathbf{s} \in \mathbb{D}$ 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 $Y(\mathbf{s})$ yet.

What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ where each $\mathbf{s}_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(\mathbf{s})$, where $\mathbf{s} \in \mathbb{D}$ 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 $Y(\mathbf{s})$ yet.

What is anisotropy?

- Modelling setup: Suppose that we have random variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ where each $\mathbf{s}_i$ denotes a particular location.

- In general, consider a real-valued spatial process $Y(\mathbf{s})$, where $\mathbf{s} \in \mathbb{D}$ 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 $Y(\mathbf{s})$ yet.

Stationarity

- Suppose our spatial process has a mean, $\mu(\mathbf{s}) = E(Y(\mathbf{s}))$, and that the variance of $Y(\mathbf{s})$ exists for all $\mathbf{s}$.

- The process is said to be strictly stationary (also called strongly stationary) if, for any given $n \geq 1$, any set of n sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and any $\mathbf{h}$ the distribution of $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ is the same as that of $Y(\mathbf{s}_1 + \mathbf{h}), \ldots, Y(\mathbf{s}_n + \mathbf{h})$.

- A less restrictive condition is given by weak stationarity (also called second-order stationarity): A process is weakly stationary if $\mu(\mathbf{s}) = \mu$ and $\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$ for all $\mathbf{h}$ such that $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ both lie in D.

Stationarity

- Suppose our spatial process has a mean, $\mu(\mathbf{s}) = E(Y(\mathbf{s}))$, and that the variance of $Y(\mathbf{s})$ exists for all $\mathbf{s}$.
- The process is said to be strictly stationary (also called strongly stationary) if, for any given $n \geq 1$, any set of n sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and any $\mathbf{h}$ the distribution of $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ is the same as that of $Y(\mathbf{s}_1 + \mathbf{h}), \ldots, Y(\mathbf{s}_n + \mathbf{h})$.
- A less restrictive condition is given by weak stationarity (also called second-order stationarity): A process is weakly stationary if $\mu(\mathbf{s}) = \mu$ and $\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$ for all $\mathbf{h}$ such that $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ both lie in D.

Stationarity

- Suppose our spatial process has a mean, $\mu(\mathbf{s}) = E(Y(\mathbf{s}))$, and that the variance of $Y(\mathbf{s})$ exists for all $\mathbf{s}$.

- The process is said to be strictly stationary (also called strongly stationary) if, for any given $n \geq 1$, any set of n sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and any $\mathbf{h}$ the distribution of $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ is the same as that of $Y(\mathbf{s}_1 + \mathbf{h}), \ldots, Y(\mathbf{s}_n + \mathbf{h})$.

- A less restrictive condition is given by weak stationarity (also called second-order stationarity): A process is weakly stationary if $\mu(\mathbf{s}) = \mu$ and $\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$ for all $\mathbf{h}$ such that $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ both lie in D.

Notes on Stationarity

- 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 $C(\mathbf{h})$ (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 $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ can be summarized by a covariance function $C(\mathbf{h})$ (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 $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ can be summarized by a covariance function $C(\mathbf{h})$ (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

Variogram

- Semi-Variogram is defined as:

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

- Simple calculation yields

$$2\gamma(\mathbf{h}) = 2\left[C(\mathbf{0}) - C(\mathbf{h})\right]$$

- 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:

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

- Simple calculation yields

$$2\gamma(\mathbf{h}) = 2\left[C(\mathbf{0}) - C(\mathbf{h})\right]$$

- 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:

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

- Simple calculation yields

$$2\gamma(\mathbf{h}) = 2\left[C(\mathbf{0}) - C(\mathbf{h})\right]$$

- 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:

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

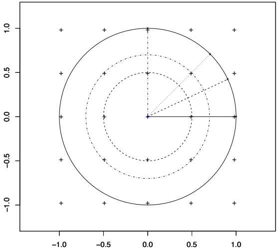
- Simple calculation yields

$$2\gamma(\mathbf{h}) = 2\left[C(\mathbf{0}) - C(\mathbf{h})\right]$$

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

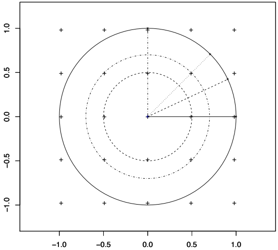
- But the converse is not true, we can add $\pm a$ to $C(\cdot)$ and obtain the same $\gamma(\cdot)$.

Isotropy



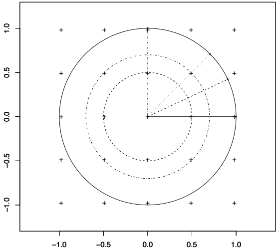
- If the semivariogram $\gamma(\mathbf{h})$ depends upon the separation vector only through its length $||\mathbf{h}||$ then we say that the process is isotropic.
- For an isotropic process, $\gamma(\mathbf{h})$ is a real-valued function of a univariate argument, and can be written as $\gamma(||\mathbf{h}||)$.
- Isotropic processes are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for $\gamma(\cdot)$.

Isotropy



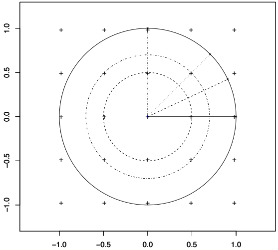
- If the semivariogram $\gamma(\mathbf{h})$ depends upon the separation vector only through its length $||\mathbf{h}||$ then we say that the process is isotropic.
- For an isotropic process, $\gamma(\mathbf{h})$ is a real-valued function of a univariate argument, and can be written as $\gamma(||\mathbf{h}||)$.
- Isotropic processes are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for $\gamma(\cdot)$.

Isotropy



- If the semivariogram $\gamma(\mathbf{h})$ depends upon the separation vector only through its length $||\mathbf{h}||$ then we say that the process is isotropic.
- For an isotropic process, $\gamma(\mathbf{h})$ is a real-valued function of a univariate argument, and can be written as $\gamma(||\mathbf{h}||)$.
- Isotropic processes are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for $\gamma(\cdot)$.

Isotropy



- If the semivariogram $\gamma(\mathbf{h})$ depends upon the separation vector only through its length $||\mathbf{h}||$ then we say that the process is isotropic.
- For an isotropic process, $\gamma(\mathbf{h})$ is a real-valued function of a univariate argument, and can be written as $\gamma(||\mathbf{h}||)$.
- 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 most common covariance function

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_ν is the modified Bessel function of second kind with order ν, and $t = ||\mathbf{h}||$ is the distance between two sites.

- The parameter ϕ controls the rate of decay of the correlation as the distance t increases
- The parameter ν controls smoothness of the random field $Y(\mathbf{s})$.
  - $\nu = 1/2 \implies C(t) = \sigma^2 \exp(-\phi t), t > 0$; Exponential Covariance Function
  - $\nu = 3/2$, $C(t) = \sigma^2(1 + \phi t)\exp(-\phi t), t > 0$.
  - $\nu \to \infty \implies C(t) = \sigma^2 \exp(-\phi^2 t^2), t > 0$; Gaussian

Exponential Covariance Function

- 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

$$\begin{aligned} \exp(-\phi t_0) &= 0.05 \\ \Longrightarrow\ t_0 &= -\log(0.05)/\phi \\ \Longrightarrow\ t_0 &\approx 3/\phi \end{aligned}$$

since $\log(0.05) \approx -3$.

Nugget

- Recall $\gamma(\mathbf{h}) = \gamma(||\mathbf{h}||) = C(\mathbf{0}) - C(||\mathbf{h}||)$.
- So $\gamma(0) = 0$. But often there are micro-scale variotion or measurement error even at very small distances.
- To tackle that we define the nugget

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

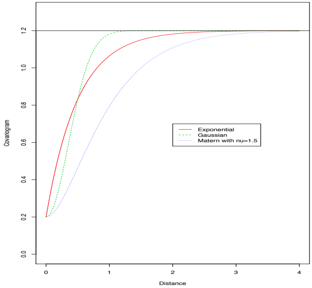
- This introduces a discontinuity at 0 for the covariogram $\gamma(t)$.

Sill

- 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, σ^2, is called the partial sill.
- The effective range is the smallest distance for which the semivariogram achieves the asymptotic sill.

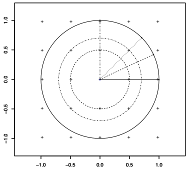
Three closed form Matérn covariograms:

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 $\nu = 1.5$. $\gamma(t) = \tau^2 + \sigma^2(1 - (1 + \phi t)\exp(-\phi t))$.



What is anisotropy?

- 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 by stretching of an isotropic model: $\gamma(h) = \gamma_0(\sqrt{\mathbf{h}' Q \mathbf{h}})$ where $\gamma_0(\cdot)$ is isotropic and Q is a positive definite matrix.
  - Zonal anisotropy. Variogram only depends on some components of the vector $\mathbf{h}$. 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 $y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$ along with the locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$).
- Hence it is difficult to specify a flexible covariance function $C(\dot{})$.
- Further problem arises due to the positive definiteness requirement of the implied covariance matrix of *any n* realisations $Y(\mathbf{s})$.

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 $y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$ along with the locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$).
- Hence it is difficult to specify a flexible covariance function $C(\dot{})$.
- Further problem arises due to the positive definiteness requirement of the implied covariance matrix of *any n* realisations $Y(\mathbf{s})$.

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 $y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$ along with the locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$).
- 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 $Y(\mathbf{s})$.

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 $y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$ along with the locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$).
- Hence it is difficult to specify a flexible covariance function $C(\dot{})$.
- Further problem arises due to the positive definiteness requirement of the implied covariance matrix of *any n* realisations $Y(\mathbf{s})$.

What's available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O'hagan (2003).
- Kernel mixing: Higdon (1998), Paciorek and Schervish (2006).
- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).
- Spatial basis functions and non-stationary Matérn covariance functions. Katzfuss (2013) and Konomi et al. (2014).
- More comprehensive literature citations in Section 3.2 of the Bayesian modelling book on spatial statistics: Banerjee, Carlin and Gelfand (2015).

What's available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O'hagan (2003).
- Kernel mixing: Higdon (1998), Paciorek and Schervish (2006).
- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).
- Spatial basis functions and non-stationary Matérn covariance functions. Katzfuss (2013) and Konomi et al. (2014).
- More comprehensive literature citations in Section 3.2 of the Bayesian modelling book on spatial statistics: Banerjee, Carlin and Gelfand (2015).

What's available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O'hagan (2003).
- Kernel mixing: Higdon (1998), Paciorek and Schervish (2006).
- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).
- Spatial basis functions and non-stationary Matérn covariance functions. Katzfuss (2013) and Konomi et al. (2014).
- More comprehensive literature citations in Section 3.2 of the Bayesian modelling book on spatial statistics: Banerjee, Carlin and Gelfand (2015).

What's available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O'hagan (2003).
- Kernel mixing: Higdon (1998), Paciorek and Schervish (2006).
- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).
- Spatial basis functions and non-stationary Matérn covariance functions. Katzfuss (2013) and Konomi et al. (2014).
- More comprehensive literature citations in Section 3.2 of the Bayesian modelling book on spatial statistics: Banerjee, Carlin and Gelfand (2015).

What's available in the literature?

- Large literature on constructing non-stationary models: Sampson and Guttorp (1992), Schmidt and O'hagan (2003).
- Kernel mixing: Higdon (1998), Paciorek and Schervish (2006).
- Spatially varying cross-covariance models. Guhaniyogi et al. (2013).
- Spatial basis functions and non-stationary Matérn covariance functions. Katzfuss (2013) and Konomi et al. (2014).
- More comprehensive literature citations in Section 3.2 of the Bayesian modelling book on spatial statistics: Banerjee, Carlin and Gelfand (2015).

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 $\mathbf{w}^* = (w(\mathbf{s}_1^*), \ldots, w(\mathbf{s}_m^*))$.
- At any other location $\mathbf{s}$, define $\tilde{w}(\mathbf{s}) = E\left[w(\mathbf{s})|\mathbf{w}^*\right]$.
- This $\tilde{w}(\mathbf{s})$ 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 $\mathbf{w}^* = (w(\mathbf{s}_1^*), \ldots, w(\mathbf{s}_m^*))$.
- At any other location $\mathbf{s}$, define $\tilde{w}(\mathbf{s}) = E\left[w(\mathbf{s})|\mathbf{w}^*\right]$.
- This $\tilde{w}(\mathbf{s})$ 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 $\mathbf{w}^* = (w(\mathbf{s}_1^*), \ldots, w(\mathbf{s}_m^*))$.
- At any other location $\mathbf{s}$, define $\tilde{w}(\mathbf{s}) = E\left[w(\mathbf{s})|\mathbf{w}^*\right]$.
- This $\tilde{w}(\mathbf{s})$ 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 $\mathbf{w}^* = (w(\mathbf{s}_1^*), \ldots, w(\mathbf{s}_m^*))$.
- At any other location $\mathbf{s}$, define $\tilde{w(\mathbf{s})} = E\left[w(\mathbf{s})|\mathbf{w}^*\right]$.
- This $\tilde{w(\mathbf{s})}$ 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 $\mathbf{w}^* = (w(\mathbf{s}_1^*), \ldots, w(\mathbf{s}_m^*))$.
- At any other location $\mathbf{s}$, define $\tilde{w}(\mathbf{s}) = E\left[w(\mathbf{s})|\mathbf{w}^*\right]$.
- This $\tilde{w}(\mathbf{s})$ defines a flexible anisotropic valid spatial process.

- 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 $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$ at random or according to a specific clustering mechanism.

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 $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$ at random or according to a specific clustering mechanism.

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 $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$ at random or according to a specific clustering mechanism.

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 $\mathrm{Cov}(\tilde{w}(s), \tilde{w}(s'))$ 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 $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$ at random or according to a specific clustering mechanism.

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 $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$ at random or according to a specific clustering mechanism.

Some details

- For any $\mathbf{s}$, $\tilde{w}(\mathbf{s}) = \mathbf{c}^*(\mathbf{s})^T S_{w^*}^{-1} \mathbf{w}^*$ where $\mathbf{c}^*(\mathbf{s})$ denotes the $m \times 1$ correlation vector between $w(\mathbf{s})$ and $\mathbf{w}^*$, given by $\left(C(|\mathbf{s} - \mathbf{s}_1^*|), \ldots, C(|\mathbf{s} - \mathbf{s}_m^*|)\right)^T$ and S_{w^*} is the correlation matrix of $\mathbf{w}^*$.
- Consider two locations $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$. Now:

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

- Depends on both $\mathbf{s}$ and $\mathbf{h}$.

Some details

- For any $\mathbf{s}$, $\tilde{w}(\mathbf{s}) = \mathbf{c}^*(\mathbf{s})^T S_{w^*}^{-1} \mathbf{w}^*$ where $\mathbf{c}^*(\mathbf{s})$ denotes the $m \times 1$ correlation vector between $w(\mathbf{s})$ and $\mathbf{w}^*$, given by $\left(C(|\mathbf{s} - \mathbf{s}_1^*|), \ldots, C(|\mathbf{s} - \mathbf{s}_m^*|)\right)^T$ and S_{w^*} is the correlation matrix of $\mathbf{w}^*$.
- Consider two locations $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$. Now:

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

- Depends on both $\mathbf{s}$ and $\mathbf{h}$.

Some details

- For any $\mathbf{s}$, $\tilde{w}(\mathbf{s}) = \mathbf{c}^*(\mathbf{s})^T S_{w^*}^{-1} \mathbf{w}^*$ where $\mathbf{c}^*(\mathbf{s})$ denotes the $m \times 1$ correlation vector between $w(\mathbf{s})$ and $\mathbf{w}^*$, given by $\left(C(|\mathbf{s} - \mathbf{s}_1^*|), \ldots, C(|\mathbf{s} - \mathbf{s}_m^*|)\right)^T$ and S_{w^*} is the correlation matrix of $\mathbf{w}^*$.
- Consider two locations $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$. Now:

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

- Depends on both $\mathbf{s}$ and $\mathbf{h}$.

Exploring correlation structure with $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$.

- Is $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ a legitimate semivariogram?
- No! Its not an even function of $\mathbf{h}$, i.e. $\tilde{\gamma}(\mathbf{s}, \mathbf{h}) \neq \tilde{\gamma}(\mathbf{s}, -\mathbf{h})$.
- We still can treat this as a function of $|\mathbf{h}|$ and study its properties for varying $\mathbf{s}$ and $\mathbf{h}$ and the knots.
- We fix a central location $\mathbf{s}^{**}$, assumed to be the centroid and then calculate distance between $\mathbf{s}^{**}$ and $\mathbf{s}^{**} + \mathbf{h}$.

Exploring correlation structure with $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$.

- Is $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ a legitimate semivariogram?
- No! Its not an even function of $\mathbf{h}$, i.e. $\tilde{\gamma}(\mathbf{s}, \mathbf{h}) \neq \tilde{\gamma}(\mathbf{s}, -\mathbf{h})$.
- We still can treat this as a function of $|\mathbf{h}|$ and study its properties for varying $\mathbf{s}$ and $\mathbf{h}$ and the knots.
- We fix a central location $\mathbf{s}^{**}$, assumed to be the centroid and then calculate distance between $\mathbf{s}^{**}$ and $\mathbf{s}^{**} + \mathbf{h}$.

Exploring correlation structure with $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$.

- Is $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ a legitimate semivariogram?
- No! Its not an even function of **h**, i.e. $\tilde{\gamma}(\mathbf{s}, \mathbf{h}) \neq \tilde{\gamma}(\mathbf{s}, -\mathbf{h})$.
- We still can treat this as a function of $|\mathbf{h}|$ and study its properties for varying **s** and **h** and the knots.
- We fix a central location $\mathbf{s}^{**}$, assumed to be the centroid and then calculate distance between $\mathbf{s}^{**}$ and $\mathbf{s}^{**} + \mathbf{h}$.

Exploring correlation structure with $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$.

- Is $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ a legitimate semivariogram?
- No! Its not an even function of **h**, i.e. $\tilde{\gamma}(\mathbf{s}, \mathbf{h}) \neq \tilde{\gamma}(\mathbf{s}, -\mathbf{h})$.
- We still can treat this as a function of $|\mathbf{h}|$ and study its properties for varying **s** and **h** and the knots.
- We fix a central location $\mathbf{s}^{**}$, assumed to be the centroid and then calculate distance between $\mathbf{s}^{**}$ and $\mathbf{s}^{**} + \mathbf{h}$.

How do we choose the knots?

- Space filling?
- Clustering?
- Random placement?
- In general, assume a point process, i.e. a random distribution $\pi(\mathbf{S}_m^*)$ for $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$.
- But how about m? We can assume that to be unknown as well.
- Assume $\pi(m)$ for m and think of $\pi(\mathbf{S}_m^*)$ 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 $\pi(\mathbf{S}_m^*)$ for $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$.
- But how about m? We can assume that to be unknown as well.
- Assume $\pi(m)$ for m and think of $\pi(\mathbf{S}_m^*)$ 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 $\pi(\mathbf{S}_m^*)$ for $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$.
- But how about m? We can assume that to be unknown as well.
- Assume $\pi(m)$ for m and think of $\pi(\mathbf{S}_m^*)$ 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 $\pi(\mathbf{S}_m^*)$ for $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$.
- But how about m? We can assume that to be unknown as well.
- Assume $\pi(m)$ for m and think of $\pi(\mathbf{S}_m^*)$ 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 $\pi(\mathbf{S}_m^*)$ for $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$.
- But how about m? We can assume that to be unknown as well.
- Assume $\pi(m)$ for m and think of $\pi(\mathbf{S}_m^*)$ 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 $\pi(\mathbf{S}_m^*)$ for $\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*$.
- But how about m? We can assume that to be unknown as well.
- Assume $\pi(m)$ for m and think of $\pi(\mathbf{S}_m^*)$ conditional on m.

Return to $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$

- It is random if m and $\mathbf{S}^*_m$ 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_ℓ from $\pi(m)$ and and a set of m_ℓ random knots $\mathbf{S}^*_{m_l}$ from $\pi(\mathbf{S}^*_{m_\ell})$.
- Conditional on these values, evaluate the inner expectation $E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right]$.
- Finally, we approximate $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ by

$$\frac{1}{2L}\sum_{\ell=1}^{L} E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right].$$

Return to $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$

- It is random if m and $\mathbf{S}^*_m$ 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_ℓ from $\pi(m)$ and and a set of m_ℓ random knots $\mathbf{S}^*_{m_l}$ from $\pi(\mathbf{S}^*_{m_\ell})$.
- Conditional on these values, evaluate the inner expectation $E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right]$.
- Finally, we approximate $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ by

$$\frac{1}{2L}\sum_{\ell=1}^{L} E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right].$$

Return to $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$

- It is random if m and $\mathbf{S}^*_m$ 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_ℓ from $\pi(m)$ and and a set of m_ℓ random knots $\mathbf{S}^*_{m_l}$ from $\pi(\mathbf{S}^*_{m_\ell})$.
- Conditional on these values, evaluate the inner expectation $E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right]$.
- Finally, we approximate $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ by

$$\frac{1}{2L} \sum_{\ell=1}^{L} E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right].$$

Return to $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$

- It is random if m and $\mathbf{S}^*_m$ 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_ℓ from $\pi(m)$ and and a set of m_ℓ random knots $\mathbf{S}^*_{m_l}$ from $\pi(\mathbf{S}^*_{m_\ell})$.
- Conditional on these values, evaluate the inner expectation $E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right]$.
- Finally, we approximate $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ by

$$\frac{1}{2L}\sum_{\ell=1}^{L} E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right].$$

Return to $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$

- It is random if m and $\mathbf{S}^*_m$ 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_ℓ from $\pi(m)$ and and a set of m_ℓ random knots $\mathbf{S}^*_{m_l}$ from $\pi(\mathbf{S}^*_{m_\ell})$.
- Conditional on these values, evaluate the inner expectation $E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right]$.
- Finally, we approximate $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ by

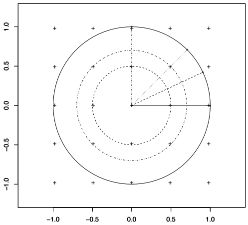
$$\frac{1}{2L}\sum_{\ell=1}^{L} E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right].$$

Return to $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$

- It is random if m and $\mathbf{S}^*_m$ 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_ℓ from $\pi(m)$ and and a set of m_ℓ random knots $\mathbf{S}^*_{m_l}$ from $\pi(\mathbf{S}^*_{m_\ell})$.
- Conditional on these values, evaluate the inner expectation $E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right]$.
- Finally, we approximate $\tilde{\gamma}(\mathbf{s}, \mathbf{h})$ by

$$\frac{1}{2L}\sum_{\ell=1}^{L} E\left[\{\tilde{w}(\mathbf{s}) - \tilde{w}(\mathbf{s}+\mathbf{h})\}^2 | m_l, \mathbf{S}^*_{m_l}\right].$$

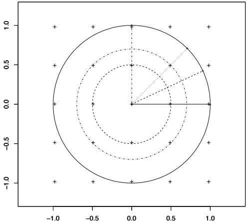
Four designs for knot-locations



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.

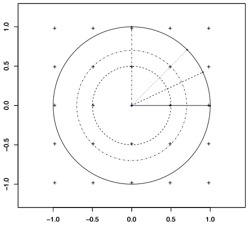
Four designs for knot-locations



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.

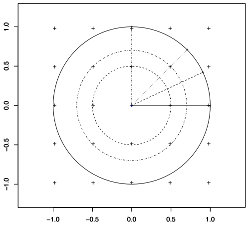
Four designs for knot-locations



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.

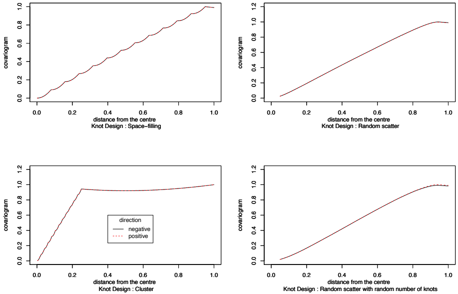
Four designs for knot-locations



Consider $\mathbb{D} = [-1, 1]$ in one and $\mathbb{D} = [-1, 1] \times [-1, 1]$ in two dimensions.

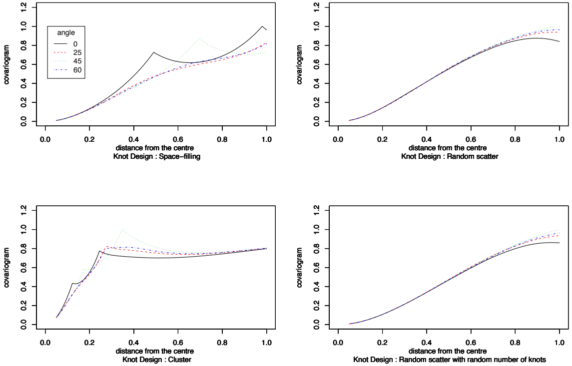
1. Space filling with $m = 25$
2. Complete Spatial Randomness (CSR) with $m = 25$.
3. All the knots clustered within the central quarter: $[-0.25, 0.25]$ with $m = 25$.
4. CSR but with m following uniform between 1 to 25.

One dimensional example



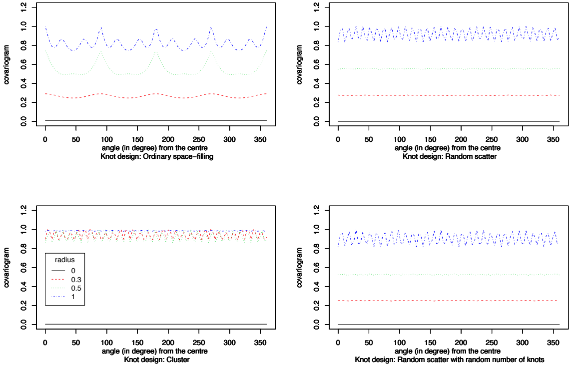
- Solid line: Semi variogram in positive and dotted line is in the negative direction.
- Compare with the figure for isotropic correlation structure shown before.
- Effect of the space filling knots are seen in the top left.
- Knots clustered in a smaller sub-region is seen in the bottom left panel.
- The correlation curves become 'more' smooth when knots are placed at random.

Two dimensional example



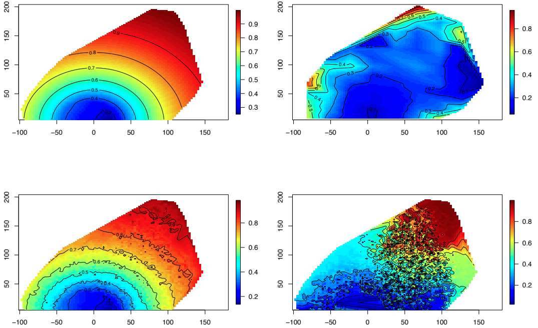
- Semivariogram plots against radial distance.
- The shape of the variogram depends on where the knots are placed.
- Shows angular anisotropy as well.

Two dimensional example ...



- Semivariogram plots against angle.
- There may not be any sill.
- Hence, the GPP can generate very flexible anisotropic processes.

Generating zonal anisotropy



- 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:

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

- The residual is partitioned into two pieces: one spatial, $\tilde{w}(\mathbf{s})$, and one non-spatial, $\epsilon(\mathbf{s})$.
- $\tilde{w}(\mathbf{s})$ is a non-stationary and anisotropic Gaussian process depending on the parameters σ_w^2, decay parameter ϕ, and smoothness ν and the number and positioning of the knot locations.
- $\epsilon(\mathbf{s})$ adds the nugget (τ^2) effect.
- $\tilde{w}(\mathbf{s})$ reduces dimension if $n > m$. Otherwise, it may increase it to achieve flexibility.

Hierarchical modeling

- Basic Model:

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

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

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

- $\epsilon(\mathbf{s})$ adds the nugget (τ^2) effect.

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

Hierarchical modeling

- Basic Model:

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

- The residual is partitioned into two pieces: one spatial, $\tilde{w}(\mathbf{s})$, and one non-spatial, $\epsilon(\mathbf{s})$.
- $\tilde{w}(\mathbf{s})$ is a non-stationary and anisotropic Gaussian process depending on the parameters σ_w^2, decay parameter ϕ, and smoothness ν and the number and positioning of the knot locations.
- $\epsilon(\mathbf{s})$ adds the nugget (τ^2) effect.
- $\tilde{w}(\mathbf{s})$ reduces dimension if $n > m$. Otherwise, it may increase it to achieve flexibility.

Hierarchical modeling

- Basic Model:

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

- The residual is partitioned into two pieces: one spatial, $\tilde{w}(\mathbf{s})$, and one non-spatial, $\epsilon(\mathbf{s})$.
- $\tilde{w}(\mathbf{s})$ is a non-stationary and anisotropic Gaussian process depending on the parameters σ_w^2, decay parameter ϕ, and smoothness ν and the number and positioning of the knot locations.
- $\epsilon(\mathbf{s})$ adds the nugget (τ^2) effect.
- $\tilde{w}(\mathbf{s})$ reduces dimension if $n > m$. Otherwise, it may increase it to achieve flexibility.

Hierarchical modeling

- Basic Model:

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

- The residual is partitioned into two pieces: one spatial, $\tilde{w}(\mathbf{s})$, and one non-spatial, $\epsilon(\mathbf{s})$.
- $\tilde{w}(\mathbf{s})$ is a non-stationary and anisotropic Gaussian process depending on the parameters σ_w^2, decay parameter ϕ, and smoothness ν and the number and positioning of the knot locations.
- $\epsilon(\mathbf{s})$ adds the nugget (τ^2) effect.
- $\tilde{w}(\mathbf{s})$ reduces dimension if $n > m$. Otherwise, it may increase it to achieve flexibility.

Interpretations attached to $\epsilon(\mathbf{s})$

- pure error term; model is not perfectly spatial;
- τ^2 and σ_w^2 are known as variance components.
- measurement error or replication variability causing discontinuity in spatial surface $Y(\mathbf{s})$;
- microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.

Interpretations attached to $\epsilon(\mathbf{s})$

- pure error term; model is not perfectly spatial;
- τ^2 and σ_w^2 are known as variance components.
- measurement error or replication variability causing discontinuity in spatial surface $Y(\mathbf{s})$;
- microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.

Interpretations attached to $\epsilon(\mathbf{s})$

- pure error term; model is not perfectly spatial;
- τ^2 and σ_w^2 are known as variance components.
- measurement error or replication variability causing discontinuity in spatial surface $Y(\mathbf{s})$;
- microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.

Interpretations attached to $\epsilon(\mathbf{s})$

- pure error term; model is not perfectly spatial;
- τ^2 and σ_w^2 are known as variance components.
- measurement error or replication variability causing discontinuity in spatial surface $Y(\mathbf{s})$;
- microscale uncertainty; distances smaller than the smallest inter-location distance, independence assumed.

Likelihood and priors

- Conditional on m assume a non-homogenous Poisson Process model for the knots $\mathbf{S}_m^*$.

$$\pi\left(\mathbf{S}_m^*\right) = (\lambda(D))^{-m} \prod_{j=1}^{m} \lambda(\mathbf{s}_j),$$

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

- The logarithm of the full posterior distribution, $\log\left(\pi\left(m, \mathbf{S}_m^*, \mathbf{w}\left(\mathbf{S}_m^*\right), \theta|\mathbf{z}\right)\right)$, is given by:

$$\begin{aligned} \propto \quad & -\frac{n}{2}\log(\tau^2) \\ & -\frac{1}{2\tau^2}\sum_{i=1}^{n}\left(z(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^T\beta - \tilde{w}(\mathbf{s}_i)\right)^2 \\ & -m\log(\lambda(D)) + \sum_{j=1}^{m}\log(\lambda(\mathbf{s}_j)) \\ & -\frac{m}{2}\log(\sigma_w^2) - \frac{1}{2}\log|S_w| - \frac{1}{2\sigma_w^2}(\mathbf{w}^*)^T S_w^{-1}\mathbf{w} \\ & +\log(\pi(\theta)) \end{aligned}$$

where $\theta = (\beta, \tau^2, \sigma_w^2, \nu, \phi)^T$ and $\pi(\theta)$ denotes the prior.

Likelihood and priors

- Conditional on m assume a non-homogenous Poisson Process model for the knots $\mathbf{S}^*_m$.

$$\pi\left(\mathbf{S}^*_m\right) = (\lambda(D))^{-m} \prod_{j=1}^{m} \lambda(\mathbf{s}_j),$$

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

- The logarithm of the full posterior distribution, $\log\left(\pi\left(m, \mathbf{S}^*_m, \mathbf{w}\left(\mathbf{S}^*_m\right), \boldsymbol{\theta}|\mathbf{z}\right)\right)$, is given by:

$$\begin{aligned} \propto\quad & -\frac{n}{2}\log(\tau^2) \\ & -\frac{1}{2\tau^2}\sum_{i=1}^{n}\left(z(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^T\boldsymbol{\beta} - \tilde{w}(\mathbf{s}_i)\right)^2 \\ & -m\log(\lambda(D)) + \sum_{j=1}^{m}\log(\lambda(\mathbf{s}_j)) \\ & -\frac{m}{2}\log(\sigma^2_w) - \frac{1}{2}\log|S_w| - \frac{1}{2\sigma^2_w}(\mathbf{w}^*)^T S_w^{-1}\mathbf{w} \\ & +\log(\pi(\boldsymbol{\theta})) \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau^2, \sigma^2_w, \nu, \phi)^T$ and $\pi(\boldsymbol{\theta})$ denotes the prior.

Priors...

- Informativeness: $\pi(\beta)$ can be a flat (improper)
- Without nugget, τ^2, can't identify both σ_w^2 and ϕ (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.
- With τ^2, then ϕ and at least one of σ_w^2 and τ^2 require informative priors.
- Assume a Matérn covariance function with known ν. If the prior on β, σ_w^2, ϕ 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(ϵ, ϵ) priors for σ_w^2 – nearly improper. Safer is IG(a, b) with $a \geq 1$.

Priors...

- Informativeness: $\pi(\beta)$ can be a flat (improper)
- Without nugget, τ^2, can't identify both σ_w^2 and ϕ (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.
- With τ^2, then ϕ and at least one of σ_w^2 and τ^2 require informative priors.
- Assume a Matérn covariance function with known ν. If the prior on β, σ_w^2, ϕ 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(ϵ, ϵ) priors for σ_w^2 – nearly improper. Safer is IG(a, b) with $a \geq 1$.

Priors...

- Informativeness: $\pi(\boldsymbol{\beta})$ can be a flat (improper)
- Without nugget, τ^2, can't identify both σ_w^2 and ϕ (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.
- With τ^2, then ϕ and at least one of σ_w^2 and τ^2 require informative priors.
- Assume a Matérn covariance function with known ν. If the prior on $\boldsymbol{\beta}, \sigma_w^2, \phi$ 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(ϵ, ϵ) priors for σ_w^2 – nearly improper. Safer is IG(a, b) with $a \geq 1$.

Priors...

- Informativeness: $\pi(\boldsymbol{\beta})$ can be a flat (improper)

- Without nugget, τ^2, can't identify both σ_w^2 and ϕ (Zhang, 2004). With Matérn, can identify the product. So an informative prior on at least one of these parameters.

- With τ^2, then ϕ and at least one of σ_w^2 and τ^2 require informative priors.

- Assume a Matérn covariance function with known ν. If the prior on $\boldsymbol{\beta}, \sigma_w^2, \phi$ 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(ϵ, ϵ) priors for σ_w^2 – nearly improper. Safer is IG(a, b) with $a \geq 1$.

Spatial prediction (Bayesian kriging)

- Prediction of $Y(\mathbf{s}_0)$ at a new site $\mathbf{s}_0$ with associated covariates $\mathbf{x}_0 \equiv \mathbf{x}(\mathbf{s}_0)$.
- Predictive distribution $\pi(y(\mathbf{s}_0)|\mathbf{y}) =$

$$\int \pi(y(\mathbf{s}_0)|m, \mathbf{S}_m^*, \mathbf{w}^*, \boldsymbol{\theta}, \mathbf{y})\pi(m, \mathbf{S}_m^*, \mathbf{w}^*, \boldsymbol{\theta}|\mathbf{y})dmd\mathbf{S}_m^* d\mathbf{w}^* d\boldsymbol{\theta}$$

- $\Longrightarrow$ easy Monte Carlo estimate using composition with Gibbs draws $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(G)}$:
- For each $\boldsymbol{\theta}^{(g)}$ drawn from $\pi(\boldsymbol{\theta}|\mathbf{y}, X)$ draw $Y(\mathbf{s}_0)^{(g)}$ from $f(y(\mathbf{s}_0|\mathbf{y}, \boldsymbol{\theta}^{(g)}, X, \mathbf{x}_0)$.

Results for NO_2 modelling and validation.

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

Table: Model choice measures for NO_2. Fitted $n = 4822$, validation $n = 601 \approx 12.4\%$. $M_1, ..., M_4$ 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.

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

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 ϕ.

Results for PM_{10} modelling and validation.

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

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 ϕ.

Results for $PM_{2.5}$ modelling and validation.

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

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 ϕ.

Example of a local authority aggregated map

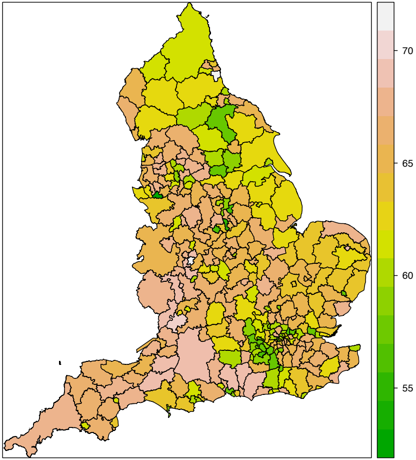


Figure: Annual map of ozone levels in 2011

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.

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.

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.

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.