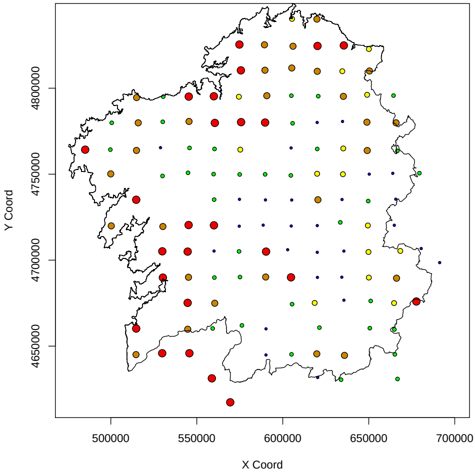
Linear geostatistical models

Dr Emanuele Giorgi
Lancaster University
e.giorgi@lancaster.ac.uk

Toowoomba 21-25 October 2019

Where we observe matters



Geostatistical lead pollution in Galicia

Geostatistical lead pollution in Galicia

- Y_i = lead concentration in a moss sample (ppm)

Geostatistical lead pollution in Galicia

- Y_i = lead concentration in a moss sample (ppm)

$$Y_i = \alpha + S(x_i) + Z_j$$

Geostatistical lead pollution in Galicia

- Y_i = lead concentration in a moss sample (ppm)

$$Y_i = \alpha + S(x_i) + Z_j$$

- $Z_i \sim N(0, \tau^2)$

Geostatistical lead pollution in Galicia

- Y_i = lead concentration in a moss sample (ppm)

$$Y_i = \alpha + S(x_i) + Z_j$$

- $Z_i \sim N(0, \tau^2)$
- We assume that $S(x)$ is a stationary and isotropic Gaussian process, i.e.

$$Cov(S(x), S(x')) = \sigma^2 \rho(u), \quad u = \|x - x'\|.$$

Geostatistical lead pollution in Galicia

- Y_i = lead concentration in a moss sample (ppm)

$$Y_i = \alpha + S(x_i) + Z_j$$

- $Z_i \sim N(0, \tau^2)$
- We assume that $S(x)$ is a stationary and isotropic Gaussian process, i.e.

$$Cov(S(x), S(x')) = \sigma^2 \rho(u), \quad u = \|x - x'\|.$$

- How do we choose $\rho(\cdot)$?

Geostatistical lead pollution in Galicia

- Y_i = lead concentration in a moss sample (ppm)

$$Y_i = \alpha + S(x_i) + Z_j$$

- $Z_i \sim N(0, \tau^2)$
- We assume that $S(x)$ is a stationary and isotropic Gaussian process, i.e.

$$Cov(S(x), S(x')) = \sigma^2 \rho(u), \quad u = \|x - x'\|.$$

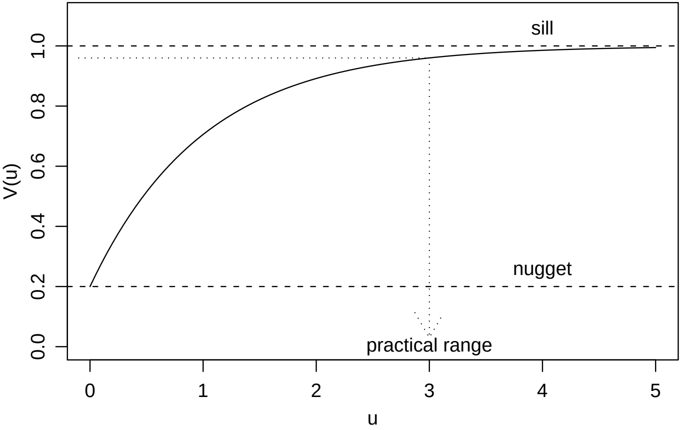
- How do we choose $\rho(\cdot)$?
- Example: $\rho(u) = \exp\{-u/\phi\}$

The theoretical variogram (continued)

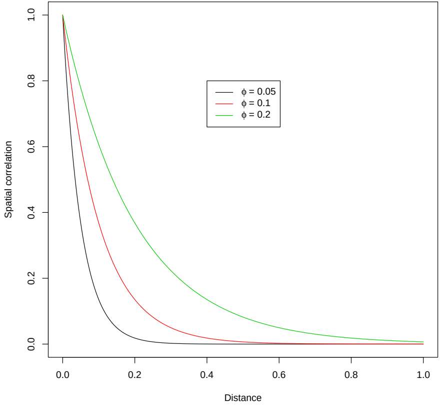
$$v(u) = \tau^2 + \sigma^2(1 - \rho(u))$$

The theoretical variogram (continued)

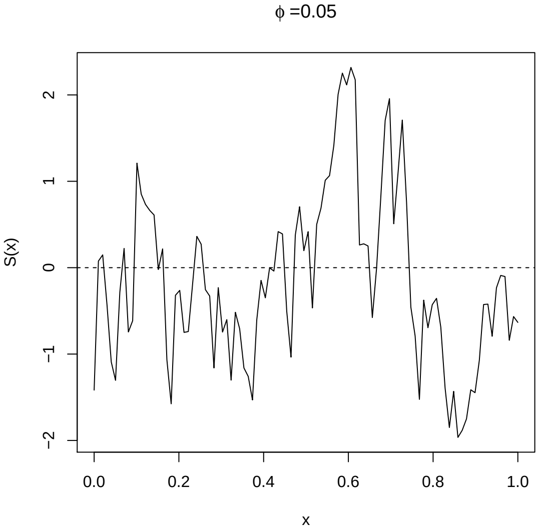
$$v(u) = \tau^2 + \sigma^2(1 - \rho(u))$$



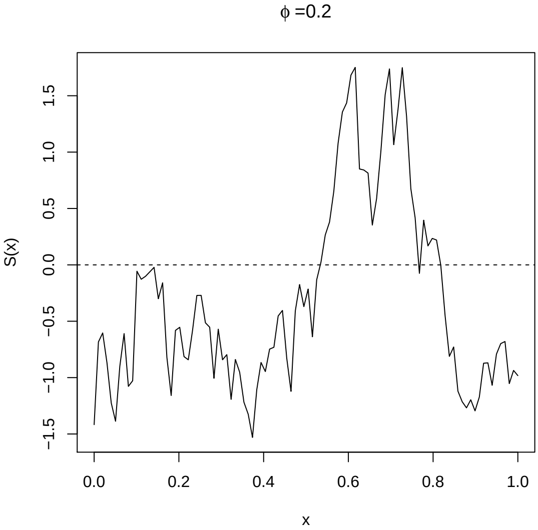
The scale of the spatial correlation



The scale of the spatial correlation



The scale of the spatial correlation



The Matérn correlation function

The Matérn correlation function

$$\rho(u) = 2^{\kappa-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi)$$

The Matérn correlation function

$$\rho(u) = 2^{\kappa-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi)$$

- $K_{\kappa}(\cdot)$: modified Bessel function of order κ

The Matérn correlation function

$$\rho(u) = 2^{\kappa-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi)$$

- $K_{\kappa}(\cdot)$: modified Bessel function of order κ
- Interpretation

The Matérn correlation function

$$\rho(u) = 2^{\kappa-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi)$$

- $K_{\kappa}(\cdot)$: modified Bessel function of order κ
- Interpretation
  - κ determines smoothness of underlying Gaussian process

$$\kappa > r \rightarrow S(x) \text{ is } r \text{ times differentiable}$$

The Matérn correlation function

$$\rho(u) = 2^{\kappa-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi)$$

- $K_{\kappa}(\cdot)$: modified Bessel function of order κ
- Interpretation
  - κ determines smoothness of underlying Gaussian process

    $$\kappa > r \rightarrow S(x) \text{ is } r \text{ times differentiable}$$

  - ϕ determines scale of the spatial correlation

The Matérn correlation function

$$\rho(u) = 2^{\kappa-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi)$$

- $K_{\kappa}(\cdot)$: modified Bessel function of order κ
- Interpretation
  - κ determines smoothness of underlying Gaussian process

    $\kappa > r \rightarrow S(x)$ is r times differentiable

  - ϕ determines scale of the spatial correlation
- Special cases
  - $= 0.5$ gives $\rho(u) = \exp\{u/\phi\}$
  - $k \rightarrow \infty$ gives $\rho(u) = \exp\{(u/\phi)^2\}$

The Matérn correlation function

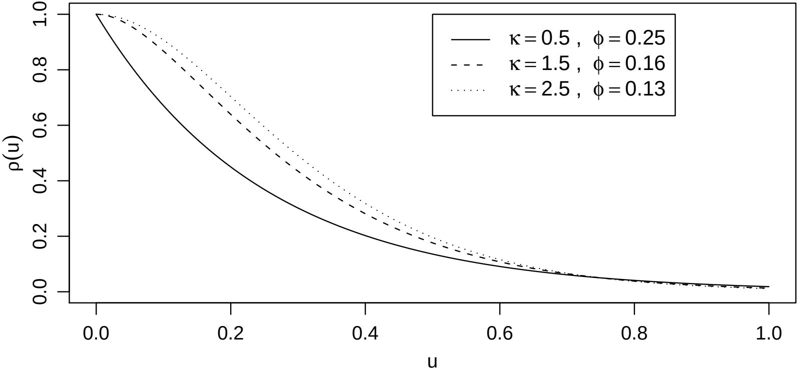
$$\rho(u) = 2^{\kappa-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi)$$

- $K_{\kappa}(\cdot)$: modified Bessel function of order κ
- Interpretation
  - κ determines smoothness of underlying Gaussian process

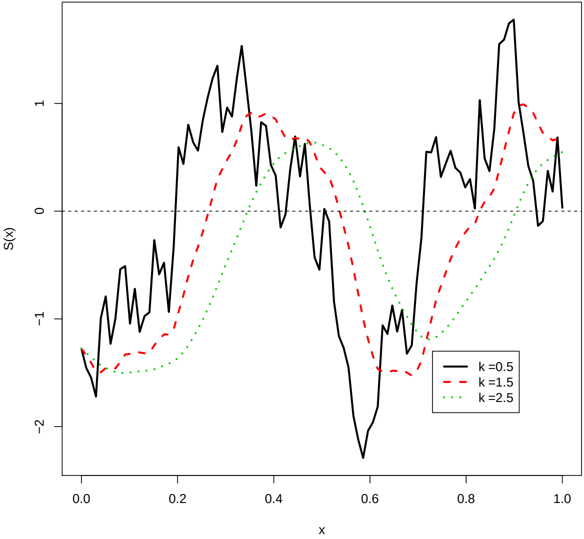
    $\kappa > r \rightarrow S(x)$ is r times differentiable

  - ϕ determines scale of the spatial correlation
- Special cases
  - $= 0.5$ gives $\rho(u) = \exp\{u/\phi\}$
  - $k \rightarrow \infty$ gives $\rho(u) = \exp\{(u/\phi)^2\}$
- Often sufficient to choose amongst $\kappa = 0.5, 1.5, 2.5$

The Matérn correlation function



The Matérn correlation function



Interpreting the nugget effect

Interpreting the nugget effect

1. Measurement error

Interpreting the nugget effect

1. Measurement error
2. Small scale spatial correlation

Interpreting the nugget effect

1. Measurement error
2. Small scale spatial correlation

$$
\begin{aligned}
Y_i &= \alpha + S(x_i) + Z(x_i) \\
\mathrm{cor}\{Z(x), Z(x')\} &= \begin{cases} \delta & \text{if } \|x - x'\| < u^* \\ 0 & \text{if } \|x - x'\| \geq u^* \end{cases}
\end{aligned}
$$

Interpreting the nugget effect

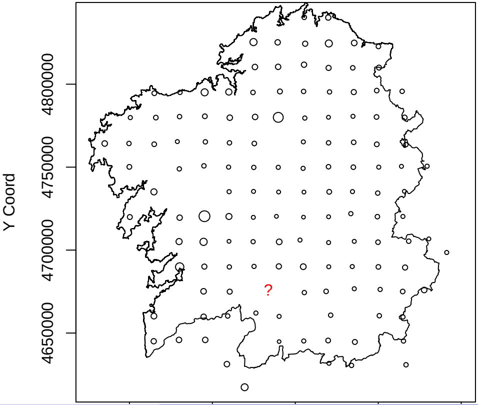
1. Measurement error
2. Small scale spatial correlation

$$
\begin{aligned}
Y_i &= \alpha + S(x_i) + Z(x_i) \\
\mathrm{cor}\{Z(x), Z(x')\} &= \begin{cases} \delta & \text{if } \|x - x'\| < u^* \\ 0 & \text{if } \|x - x'\| \geq u^* \end{cases}
\end{aligned}
$$

If $u^* < u_{min}$ then $Z(x_i)$ is pure noise.

The canonical geostatistical problem

Given a set of measurements $Y_i : i = 1, \ldots, n$ at locations x_i in a spatial region A, presumed to be (noisy) measurements of a spatially continuous phenomenon $S(x_i)$, what can we say about the realisation of $S(x)$ throughout A?



Getting initial parameter estimates

- Widely used, but not recommended except for initial analysis.
- $\theta = (\sigma^2, \phi, \tau^2)$
- Weighted least squares criterion:

$$W(\theta) = \sum_k n_k [\hat{v}(u_k) - v(u_k; \theta)]^2$$

Getting initial parameter estimates

- Widely used, but not recommended except for initial analysis.
- $\theta = (\sigma^2, \phi, \tau^2)$
- Weighted least squares criterion:

$$W(\theta) = \sum_k n_k [\hat{v}(u_k) - v(u_k; \theta)]^2$$

- Arbitrary binning and upper limit for u_k.

Getting initial parameter estimates

- Widely used, but not recommended except for initial analysis.
- $\theta = (\sigma^2, \phi, \tau^2)$
- Weighted least squares criterion:

$$W(\theta) = \sum_k n_k [\hat{v}(u_k) - v(u_k; \theta)]^2$$

- Arbitrary binning and upper limit for u_k.
- Standard errors not available.

Maximum likelihood estimation

Maximum likelihood estimation

- Multivariate Gaussian distribution: $Y \sim MVN(D\beta, \sigma^2 + \tau^2 I)$.
  - D matrix of covariates: $[D]_{ik} = d_k(x_i)$
  - R matrix of spatial correlation: $[R]_{ij} = \rho(u_{ij})$, with $u_{ij} = \|x_i - x_j\|$.

Maximum likelihood estimation

- Multivariate Gaussian distribution: $Y \sim MVN(D\beta, \sigma^2 + \tau^2 I)$.
  - D matrix of covariates: $[D]_{ik} = d_k(x_i)$
  - R matrix of spatial correlation: $[R]_{ij} = \rho(u_{ij})$, with $u_{ij} = \|x_i - x_j\|$.
- Fitting process
  1. Initialise β, e.g. using ordinary least squares.
  2. Initialise θ, e.g. using the empirical variogram
  3. Maximize

     $$l(\theta) = \log\{f(y; \beta, \theta)\}$$

     where $f(\cdot; \beta; \theta)$ denotes the density of the multivariate Gaussian distribution.

Validation of $\rho(\cdot)$

Validation of $\rho(\cdot)$

1. Obtain θ.

Validation of $\rho(\cdot)$

1. Obtain θ.
2. Simulate a large enough number of data-sets (e.g. 1,000) from the fitted model.

Validation of $\rho(\cdot)$

1. Obtain θ.
2. Simulate a large enough number of data-sets (e.g. 1,000) from the fitted model.
3. Fit a non-spatial mixed model (i.e. $S(x_i)$ is replaced by unstructured random effects U_i).

Validation of $\rho(\cdot)$

1. Obtain θ.
2. Simulate a large enough number of data-sets (e.g. 1,000) from the fitted model.
3. Fit a non-spatial mixed model (i.e. $S(x_i)$ is replaced by unstructured random effects U_i).
4. Compute the variogram using the estimated random effects associated with a location (i.e. $\hat{U}_i$).

Validation of $\rho(\cdot)$

1. Obtain θ.
2. Simulate a large enough number of data-sets (e.g. 1,000) from the fitted model.
3. Fit a non-spatial mixed model (i.e. $S(x_i)$ is replaced by unstructured random effects U_i).
4. Compute the variogram using the estimated random effects associated with a location (i.e. $\hat{U}_i$).
5. Repeat 3 to 4, for all the simulated data-sets.

Validation of $\rho(\cdot)$

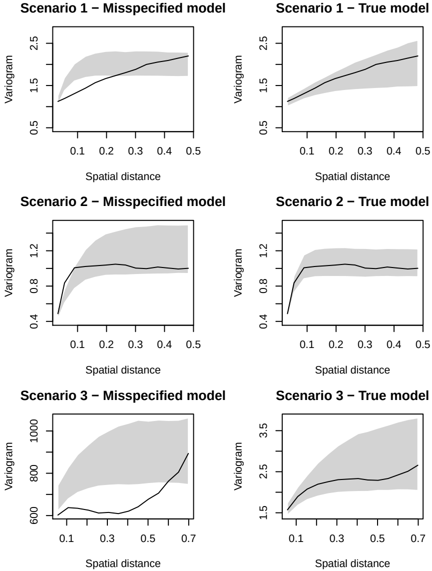
1. Obtain θ.
2. Simulate a large enough number of data-sets (e.g. 1,000) from the fitted model.
3. Fit a non-spatial mixed model (i.e. $S(x_i)$ is replaced by unstructured random effects U_i).
4. Compute the variogram using the estimated random effects associated with a location (i.e. $\hat{U}_i$).
5. Repeat 3 to 4, for all the simulated data-sets.
6. Summarize the results by computing the 95% confidence intervals for each distance u.

Validation of $\rho(\cdot)$

1. Obtain θ.
2. Simulate a large enough number of data-sets (e.g. 1,000) from the fitted model.
3. Fit a non-spatial mixed model (i.e. $S(x_i)$ is replaced by unstructured random effects U_i).
4. Compute the variogram using the estimated random effects associated with a location (i.e. $\hat{U}_i$).
5. Repeat 3 to 4, for all the simulated data-sets.
6. Summarize the results by computing the 95% confidence intervals for each distance u.

The generated 95% bandwidth from the last step indicates the band of variation for the variogram under the assumed model.

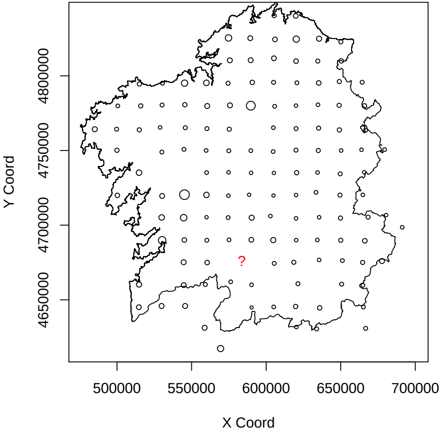
Examples



Scenario 1 − Misspecified model
Scenario 1 − True model
Scenario 2 − Misspecified model
Scenario 2 − True model
Scenario 3 − Misspecified model
Scenario 3 − True model
Variogram
Spatial distance
0.5
1.5
2.5
0.1 0.2 0.3 0.4 0.5
0.4
0.8
1.2
600
800
1000
1.5
2.5
3.5
0.1 0.3 0.5 0.7

The canonical geostatistical problem

Given a set of measurements $Y_i : i = 1, \ldots, n$ at locations x_i in a spatial region A, presumed to be (noisy) measurements of a spatially continuous phenomenon $S(x_i)$, what can we say about the realisation of $S(x)$ throughout A?



Prediction

The answer to any prediction problem is a probability distribution

Peter McCullagh, FRS

- $T =$ any quantity of scientific interest
- $Y =$ data that can tell us something about T.

The predictive distribution of T is the conditional probability distribution of T given Y

Geostatistical prediction

Let $S^* = \{S(x_1^*), ..., S(x_M^*)\}$ for any set of locations $\{x_1^*, ..., x_M^*\}$

- $Y \sim$ multivariate Normal
- for the Gaussian linear model $S^*|Y \sim$ multivariate Normal
- hence simulate samples of S^* conditional on Y
- corresponding $T^* = \mathcal{T}(S^*)$ are samples from predictive distribution of T

Minimum mean square error prediction

Model

- $[S^*]$ = probability distribution of underlying spatial process
- $[Y|S^*]$ = probability distribution of data conditional on underlying spatial process
- Bayes' theorem then gives us the predictive distribution $[S^*|Y]$

Mean square error

- $\hat{T} = t(Y)$ is a point predictor
- $\text{MSE}(\hat{T}) = \mathrm{E}[(\hat{T} - T)^2]$ is the mean square error

Theorem

1. $MSE(\hat{T})$ takes its minimum value when $\hat{T} = \mathrm{E}(T|Y)$.
2. $\mathrm{Var}(T|Y)$ estimates the achieved mean square error

Simple and ordinary kriging

$$Y \sim \mathrm{MVN}(\mu 1, \sigma^2 V)$$

$$V = R + (\tau^2/\sigma^2) \qquad R_{ij} = \rho(\|x_i - x_j\|)$$

Target for prediction is $T = S(x)$

Write $r = (r_1, ..., r_n)$ where

$$r_i = \rho(\|x - x_i\|)$$

Standard results on multivariate Normal then give $[T|Y]$ as multivariate Gaussian with mean and variance

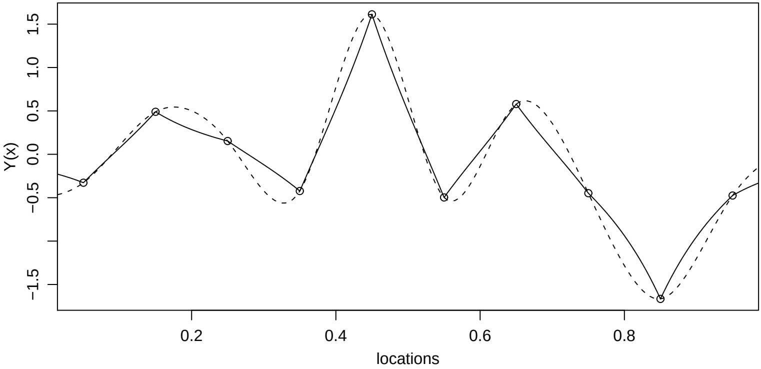
$$\hat{T} = \mu + r' V^{-1}(Y - \mu\mathbf{1})$$

$$\mathrm{Var}(T|Y) = \sigma^2(1 - r' V^{-1} r)$$

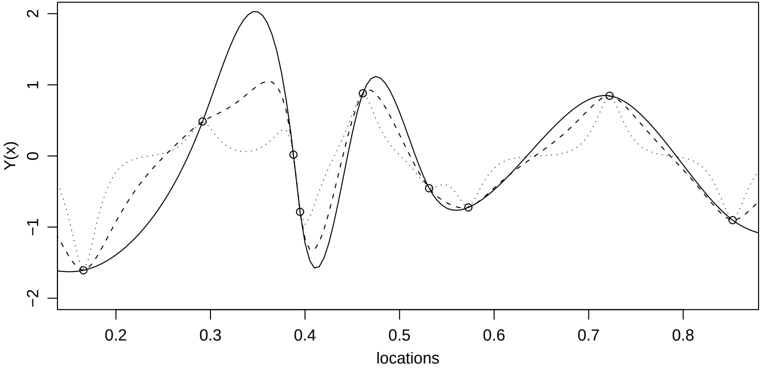
Simple kriging: $\hat{\mu} = \bar{Y}$ Ordinary kriging: $\hat{\mu} = (1' V^{-1} 1)^{-1} 1' V^{-1} Y$

Simple kriging: three examples

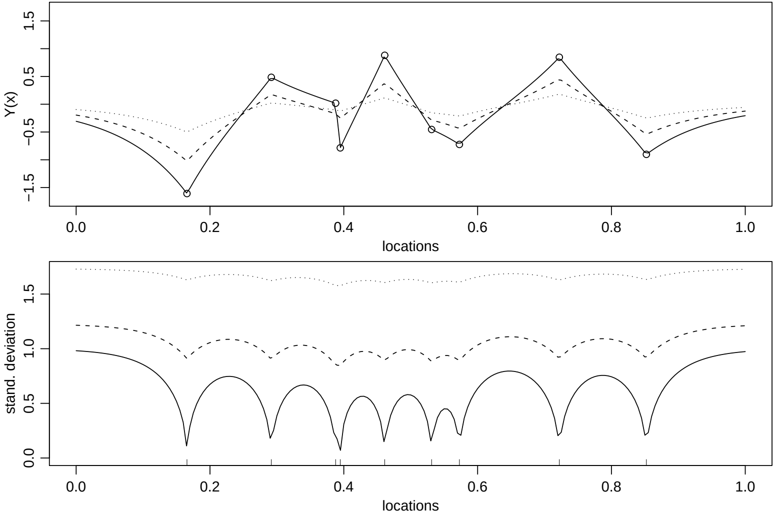
1. Varying κ (smoothness of $S(x)$)



2. Varying ϕ (range of spatial correlation



3. Varying τ^2/σ^2 (noise-to-signal ratio)



Trans-Gaussian models

- assume Gaussian model holds after point-wise transformation
- Box-Cox family is widely used

$$Y_i^* = h_\lambda(Y_i) = \begin{cases} (Y_i^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log(Y_i) & \text{if } \lambda = 0 \end{cases}$$

Example: log-Gaussian kriging

- $T(x) = \exp\{S(x)\} \quad \hat{T}(x) = \exp\{\hat{S}(x) + v(x)/2\}$
- $S_1, ..., S_m$ are a sample from $[S|Y]$
- $T_i = \exp(S_i) \Rightarrow T_1, ..., T_m$ are a sample from $[T|Y]$

Reminder: Predicting non-linear functionals

- minimum mean square error prediction is not invariant under non-linear transformation
- the complete answer to a prediction problem is the predictive distribution, $[T|Y]$
- Recommended strategy:
  - draw repeated samples from $[S^*|Y]$
  - calculate required summaries

Bayesian inference

Model specification

$$[Y, \theta] = [\theta][Y|\theta]$$

- $[Y|\theta]$ probability distribution of Y given parameter value θ
- $[\theta]$ prior probability distribution for θ
(before you collect any data)

Parameter estimation

- Bayes' Theorem gives posterior distribution for θ
(adding information from data)

$$[\theta|Y] = [Y|\theta][\theta]/[Y]$$

where $[Y] = \int [Y|\theta][\theta] d\theta$

Bayesian inference for geostatistical models

Model specification

$$[Y, S, \theta] = [\theta][S|\theta][Y|S, \theta]$$

- $[S]$ is an unobserved spatial stochastic process, representng the spatial phenomenon of scientifc interest

Parameter estimation

- integration gives likelihood function

$$[Y, \theta] = \int [Y, S, \theta] dS = [\theta][Y|\theta]$$

- as before, Bayes' Theorem gives posterior distribution

$$[\theta|Y] = [Y|\theta][\theta]/[Y]$$

where $[Y] = \int [Y|\theta][\theta] d\theta$

Bayesian inference for geostatistical models (2)

Prediction

S denotes the spatial process of interest at data-locations

S^* denotes the same process at data and prediction locations

- expand model specification to

$$[Y, S^*, \theta] = [\theta][S|\theta][Y|S, \theta][S^*|S, \theta]$$

- plug-in predictive distribution is

$$[S^*|Y, \hat{\theta}]$$

- Bayesian predictive distribution is

$$[S^*|Y] = \int [S^*|Y, \theta][\theta|Y]d\theta$$

- for any target $T = t(S^*)$, required predictive distribution $[T|Y]$ follows by direct calculation

Notes

- likelihood function is central to both classical and Bayesian inference
- Bayesian prediction is a weighted average of plug-in predictions, with different plug-in values of θ weighted according to their conditional probabilities given the observed data.
- Bayesian prediction is usually more conservative than plug-in prediction