

Lecture 15

Interpolation of Spatial Data II

DSA 8020 Statistical Methods II

April 25-29, 2022

Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data

Whitney Huang
Clemson University

Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data

1 Review: Spatial Interpolation

2 Parameter estimation

3 A Case Study of Paraná State Precipitation Data

If

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

Then

$$[\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2] \sim N(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2})$$

where

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \Sigma_{12} \Sigma_{22}^{-1} (\mathbf{y}_2 - \boldsymbol{\mu}_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

If $\{Y(\mathbf{s})\}_{\mathbf{s} \in \mathcal{S}}$ follows a GP, then

$$\begin{pmatrix} Y_0 \\ \mathbf{Y} \end{pmatrix} \sim N \left(\begin{pmatrix} m_0 \\ \mathbf{m} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & k^T \\ k & \Sigma \end{pmatrix} \right)$$

We have

$$[Y_0 | \mathbf{Y} = \mathbf{y}] \sim N(m_{Y_0 | \mathbf{Y} = \mathbf{y}}, \sigma_{Y_0 | \mathbf{Y} = \mathbf{y}}^2)$$

where

$$\begin{aligned} m_{Y_0 | \mathbf{Y} = \mathbf{y}} &= m_0 + k^T \Sigma^{-1} (\mathbf{y} - \mathbf{m}) \\ \sigma_{Y_0 | \mathbf{Y} = \mathbf{y}}^2 &= \sigma_0^2 - k^T \Sigma^{-1} k \end{aligned}$$

Next, we are going to revisit our toy examples

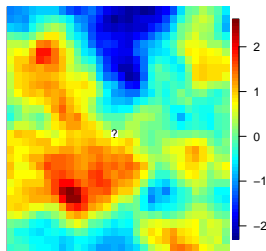
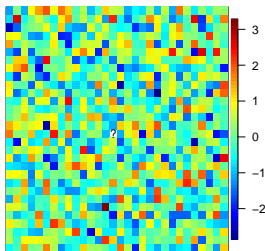
Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data

Toy Examples Revisited

For simplicity, we assume $m(s) = 0$ for $s \in \mathcal{S}$, the spatial covariance only depends on distance



$$m_{Y_0|Y=y} = 0 + k^T \Sigma^{-1} (y - 0), \quad \sigma_{Y_0|Y=y}^2 = \sigma_0^2 - k^T \Sigma^{-1} k$$

Spatial uncorrelated field:

- $m_{Y_0|Y} = 0$
- $\sigma_{Y_0|Y=y}^2 = \sigma_0^2$

Spatial correlated field:

- $m_{Y_0|Y} = k^T \Sigma^{-1} y$
- $\sigma_{Y_0|Y=y}^2 = \sigma_0^2 - k^T \Sigma^{-1} k$

Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data

Interpolating Multiple Points in Space

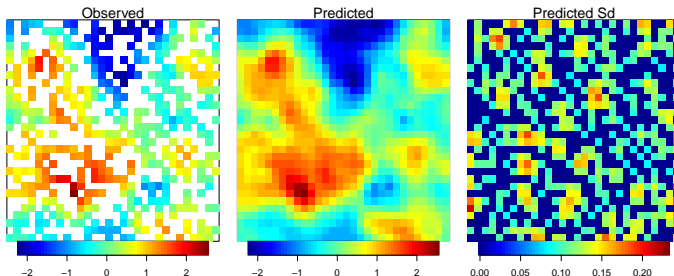
In practice, we would like to predict the values at many locations. The Gaussian conditional distribution formula can still be used:

$$[Y_0|Y = y] \sim N(m_{Y_0|Y=y}, \Sigma_{Y_0|Y=y})$$

where

$$m_{Y_0|Y=y} = m_0 + k^T \Sigma^{-1} (y - m)$$

$$\Sigma_{Y_0|Y=y} = \Sigma_0 - k^T \Sigma^{-1} k$$



If $\{Y(s)\}_{s \in \mathcal{S}}$ follows a GP, then

$$\begin{pmatrix} Y_0 \\ \mathbf{Y} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{m}_0 \\ \mathbf{m} \end{pmatrix}, \begin{pmatrix} \Sigma_0 & \mathbf{k}^T \\ \mathbf{k} & \Sigma \end{pmatrix} \right)$$

We have

$$[Y_0 | \mathbf{Y} = \mathbf{y}] \sim N(\mathbf{m}_{Y_0 | \mathbf{Y} = \mathbf{y}}, \Sigma_{Y_0 | \mathbf{Y} = \mathbf{y}})$$

where

$$\mathbf{m}_{Y_0 | \mathbf{Y} = \mathbf{y}} = \mathbf{m}_0 + \mathbf{k}^T \Sigma^{-1} (\mathbf{y} - \mathbf{m})$$

$$\Sigma_{Y_0 | \mathbf{Y} = \mathbf{y}} = \Sigma_0 - \mathbf{k}^T \Sigma^{-1} \mathbf{k}$$

Question: what if we don't know $m(s; \beta), c(h; \theta)$?

\Rightarrow We need to estimate the mean and covariance from the data \mathbf{y} .

Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data

Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data

1 Review: Spatial Interpolation

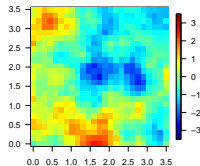
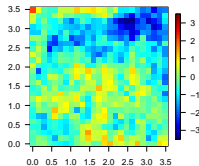
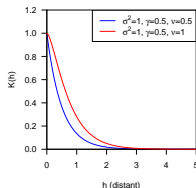
2 Parameter estimation

3 A Case Study of Paraná State Precipitation Data

Recap: Gaussian Process

We assume that the observed data $\{y(s_i)\}_{i=1}^n$ is one partial realization of a (continuously indexed) spatial stochastic process $\{Y(s)\}_{s \in \mathcal{S}}$.

- Gaussian Processes $GP(m(\cdot), K(\cdot, \cdot))$ are widely used in modeling spatial stochastic processes
- Spatial statisticians often focus on the covariance function.
e.g. $K(h) = \sigma^2 \frac{(\sqrt{2\nu}h/\gamma)^\nu \mathcal{K}_\nu(\sqrt{2\nu}h/\gamma)}{\Gamma(\nu)2^{\nu-1}}$



Under the stationary and isotropic assumptions

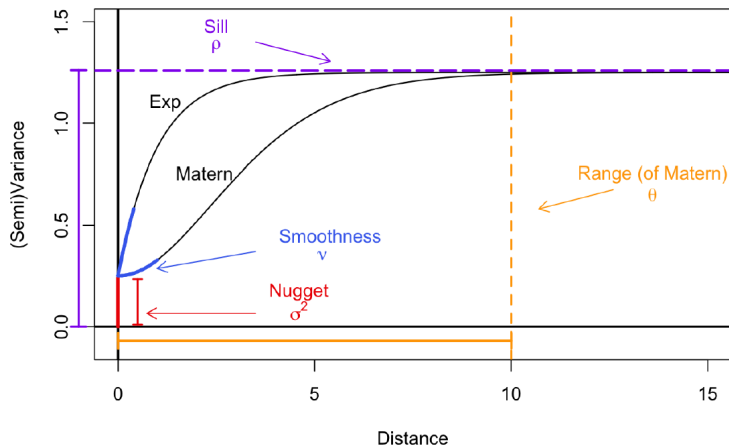
Variogram:

$$\begin{aligned}2\gamma(\mathbf{s}_i, \mathbf{s}_j) &= \text{Var}(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)) \\&= \text{E} \left\{ ((Y(\mathbf{s}_i) - \mu(\mathbf{s}_i)) - (Y(\mathbf{s}_j) - \mu(\mathbf{s}_j)))^2 \right\} \\&= \text{E} \left\{ (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2 \right\} \\&= 2\gamma(\|\mathbf{s}_i - \mathbf{s}_j\|) = 2\gamma(h)\end{aligned}$$

Semivariogram and covariance function:

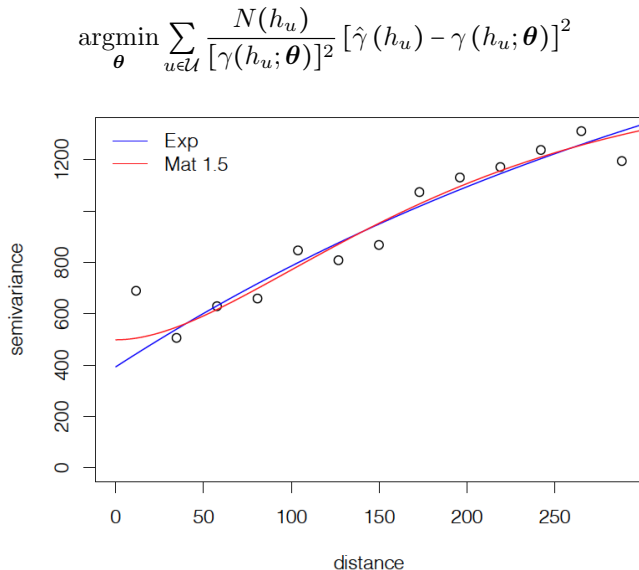
$$\gamma(h) = C(0) - C(h)$$

Semivariogram $\{\frac{1}{2}\text{Var}(\varepsilon(s_i) - \varepsilon(s_j))\}_{i,j}$



Source: `fields` vignette by Wiens and Krock, 2019

Estimation: Weighted Least Squares Method



Log-likelihood:

Given data $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))^T$

$$\ell_n(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) \propto -\frac{1}{2} \log |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T [\boldsymbol{\Sigma}_{\boldsymbol{\theta}}]_{n \times n}^{-1} (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(i, j) = \sigma^2 \rho_{\rho, \nu}(\|\mathbf{s}_i - \mathbf{s}_j\|) + \tau^2 1_{\{\mathbf{s}_i = \mathbf{s}_j\}}, i, j = 1, \dots, n$

Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data

Maximum Likelihood Estimation (MLE)

Log-likelihood:

Given data $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))^T$

$$\ell_n(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) \propto -\frac{1}{2} \log |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T [\boldsymbol{\Sigma}_{\boldsymbol{\theta}}]_{n \times n}^{-1} (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(i, j) = \sigma^2 \rho_{\rho, \nu}(\|\mathbf{s}_i - \mathbf{s}_j\|) + \tau^2 1_{\{\mathbf{s}_i = \mathbf{s}_j\}}, i, j = 1, \dots, n$

for any fixed $\boldsymbol{\theta}_0 \in \Theta$ the unique value of $\boldsymbol{\beta}$ that maximizes ℓ_n is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0} \mathbf{y}$$

Then we obtain the profile log likelihood

$$\ell_n(\boldsymbol{\theta}; \mathbf{y}) \propto -\frac{1}{2} \log |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2} \mathbf{y}^T P(\boldsymbol{\theta}) \mathbf{y}$$

where

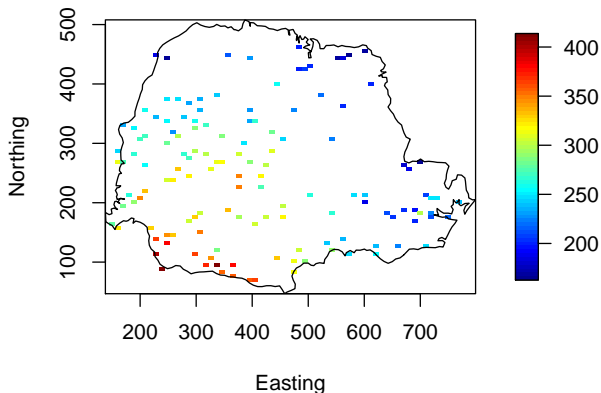
$$P(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$$

Solve the maximization problem above to get the MLE

- Maximizing $\ell_n(\boldsymbol{\theta}; \mathbf{y})$ is a constrained nonlinear optimization problem \Rightarrow ML estimates must be obtained numerically
- Restricted (or residual) maximum likelihood (REML) can be used instead

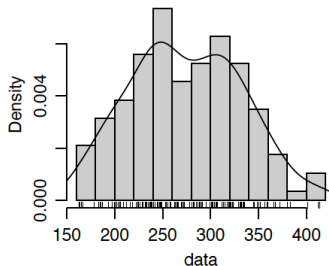
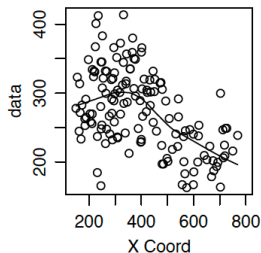
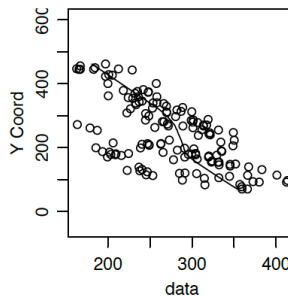
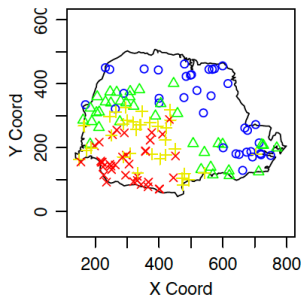
Paraná State Precipitation Data

We look at the average winter (May-June, dry season) rainfall at 143 locations throughout Paraná, Brazil

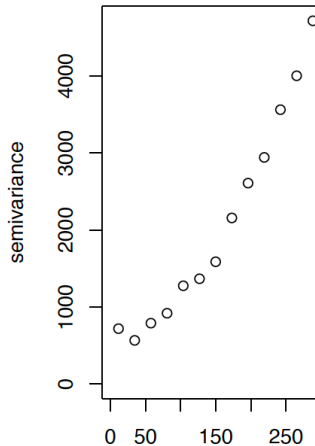
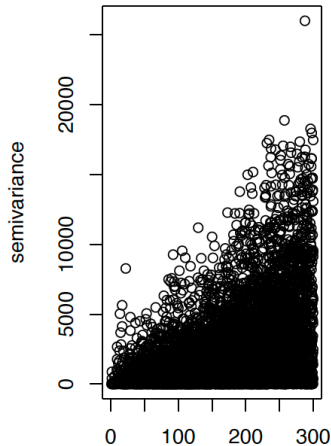


Goal: To interpolate the values in the spatial domain

Exploratory Data Analysis



Variogram Analysis



An increasing variogram suggests spatial dependence structure

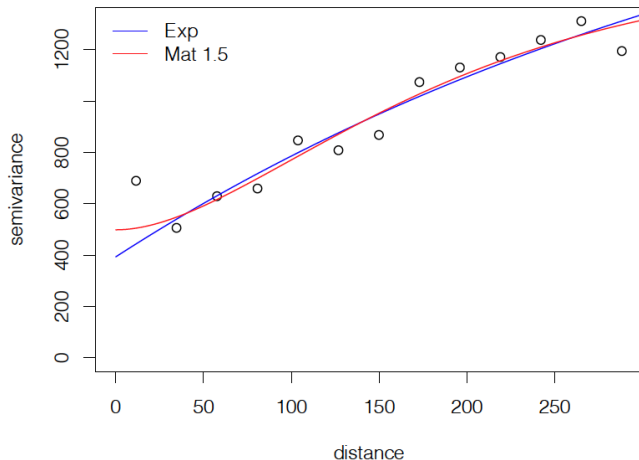
Estimating Spatial Covariance Via Varogram

```
parana.vtfit.exp <- variofit(parana.variot)  
parana.vtfit.mat1.5 <- variofit(parana.variot, kappa = 1.5)
```

Review: Spatial
Interpolation

Parameter estimation

A Case Study of
Paraná State
Precipitation Data



Maximum Likelihood Estimation of Paraná Rainfall

```
(parana.ml1 <- likfit(parana, trend = "1st", ini = c(1000, 50), nug = 100))
```

```
## -----  
## likfit: likelihood maximisation using the function optim.  
## likfit: Use control() to pass additional  
##      arguments for the maximisation function.  
##      For further details see documentation for optim.  
## likfit: It is highly advisable to run this function several  
##      times with different initial values for the parameters.  
## likfit: WARNING: This step can be time demanding!  
## -----  
## likfit: end of numerical maximisation.  
  
## likfit: estimated model parameters:  
##      beta0      beta1      beta2      tausq      sigmasq      phi  
## "416.4984" " -0.1375" " -0.3997" "385.5180" "785.6904" "184.3863"  
## Practical Range with cor=0.05 for asymptotic range: 552.3719  
##  
## likfit: maximised log-likelihood = -663.9
```

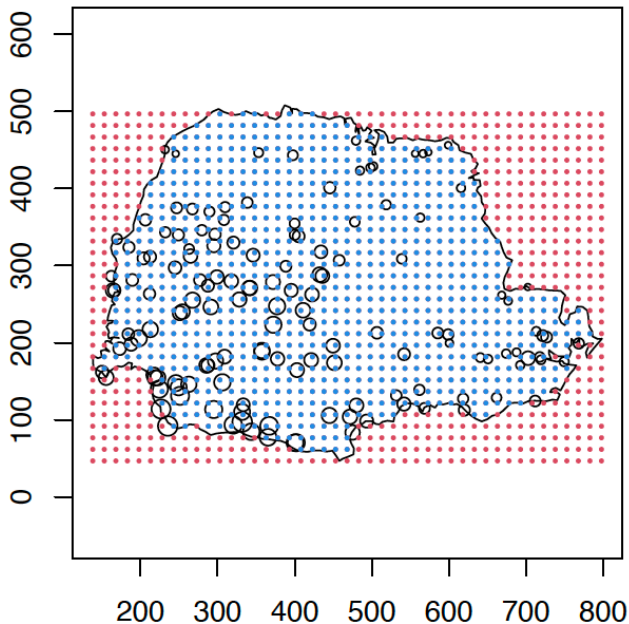
Review: Spatial
Interpolation

Parameter estimation

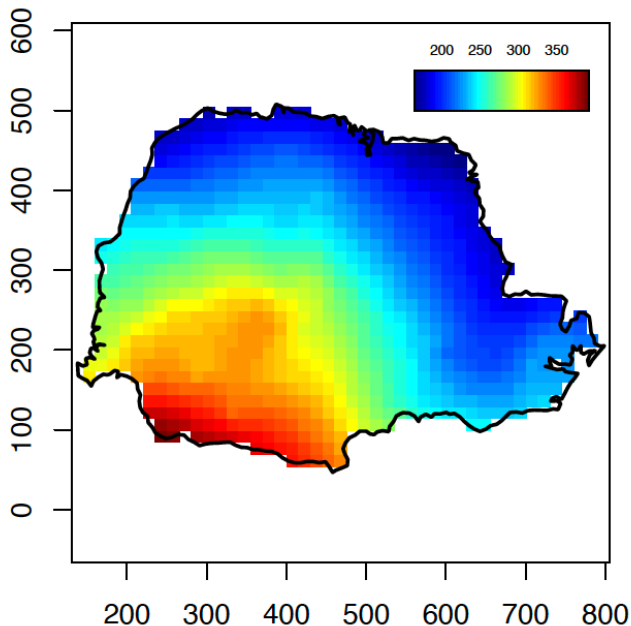
A Case Study of
Paraná State
Precipitation Data

Next, we will use these information to conduct spatial interpolation

Setting Up the Spatial Grids for Prediction



Spatial Predicted Map



Prediction Uncertainty Map

