

Lecture 27

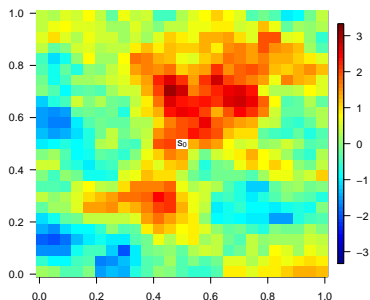
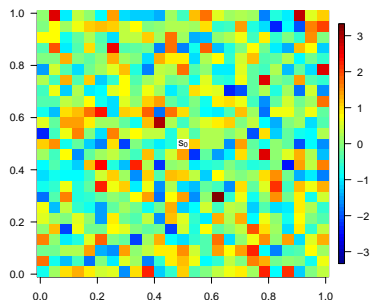
An Overview of Spatial Interpolation

STAT 8020 Statistical Methods II

December 3, 2020

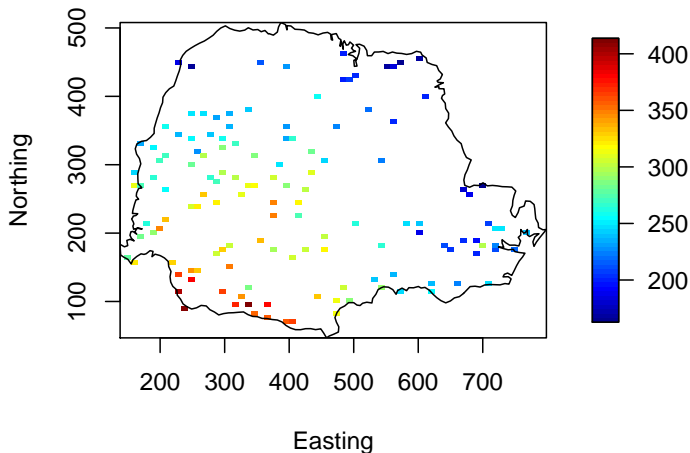
Whitney Huang
Clemson University

Toy Examples of Spatial Interpolation



Question: What is your best guess of the value of the missing pixel, denoted as $Y(s_0)$, for each case?

Interpolating Paraná State Precipitation Data



Goal: To interpolate the values in the spatial domain

The Spatial Interpolation Problem

Given observations of a spatially varying quantity Y at n spatial locations

$$y(\mathbf{s}_1), y(\mathbf{s}_2), \dots, y(\mathbf{s}_n), \quad \mathbf{s}_i \in \mathcal{S}, i = 1, \dots, n$$

We want to estimate this quantity at any **unobserved location**

$$Y(\mathbf{s}_0), \quad \mathbf{s}_0 \in \mathcal{S}$$

Applications

- Mining: ore grade
- Climate: temperature, precipitation, ...
- Remote Sensing: CO₂ retrievals
- Environmental Science: air pollution levels, ...

Some History

- Mining (Krige 1951)
Matheron (1960s),
Forestry (Matérn
1960)
- More recent work:
Cressie (1993) Stein
(1999)



1 Gaussian Process Spatial Model

2 Spatial Interpolation

3 Parameter estimation

The best guess (in a statistical sense) should be based on the conditional distribution $[Y(s_0) | \mathbf{Y} = \mathbf{y}]$ where

$$\mathbf{y} = (y(s_1), \dots, y(s_n))^T$$

- Calculating this conditional distribution can be difficult
- Instead we use a **linear predictor**:

$$\hat{Y}(s_0) = \lambda_0 + \sum_{i=1}^n \lambda_i y(s_i)$$

- The best linear predictor is completely determined by the **mean** and **covariance** of $\{Y(s), s \in \mathcal{S}\}$, and the observations \mathbf{y}

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

Model:

$$Y(\mathbf{s}) = m(\mathbf{s}) + \epsilon(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d$$

where

- Mean function:

$$m(\mathbf{s}) = \mathbb{E}[Y(\mathbf{s})] = \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}$$

- Covariance function:

$$\{\epsilon(\mathbf{s})\}_{\mathbf{s} \in \mathcal{S}} \sim \text{GP}(0, K(\cdot, \cdot)), \quad K(\mathbf{s}_1, \mathbf{s}_2) = \text{Cov}(\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2))$$

In practice, the covariance must be estimated from the data $(y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))^T$. We need to impose some structural assumptions

- **Stationarity:**

$$\begin{aligned}K(\mathbf{s}_1, \mathbf{s}_2) &= \text{Cov}(\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2)) = C(\mathbf{s}_1 - \mathbf{s}_2) \\ &= \text{Cov}(\epsilon(\mathbf{s}_1 + \mathbf{h}), \epsilon(\mathbf{s}_2 + \mathbf{h}))\end{aligned}$$

- **Isotropy:**

$$K(\mathbf{s}_1, \mathbf{s}_2) = \text{Cov}(\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2)) = C(\|\mathbf{s}_1 - \mathbf{s}_2\|)$$

A Valid Covariance Function Must Be Positive Definite (p.d.)!

A covariance function is positive if

$$\sum_{i,j=1}^n a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0$$

for any finite locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, and for any constants a_i ,
 $i = 1, \dots, n$

Question: what is the consequence if a covariance function is NOT p.d.? \Rightarrow weird things can happen

Question: How to guarantee a $C(\cdot)$ is p.d.?

- Using a **parametric covariance function**
- Using **Bochner's Theorem** to construct a valid covariance function

Some Commonly Used Covariance Functions

- **Powered exponential:**

$$C(h) = \sigma^2 \exp\left(-\left(\frac{h}{\rho}\right)^\alpha\right), \quad \sigma^2 > 0, \rho > 0, 0 < \alpha \leq 2$$

- **Spherical:**

$$C(h) = \sigma^2 \left(1 - 1.5\frac{h}{\rho} + 0.5\left(\frac{h}{\rho}\right)^3\right) \mathbb{1}_{\{h \leq \rho\}}, \quad \sigma^2, \rho > 0$$

Note: it is only valid for 1, 2, and 3 dimensional spatial domain.

- **Matérn:**

$$C(h) = \sigma^2 \frac{(\sqrt{2\nu}h/\rho)^\nu \mathcal{K}_\nu(\sqrt{2\nu}h/\rho)}{\Gamma(\nu)2^{\nu-1}}, \quad \sigma^2 > 0, \rho > 0, \nu > 0$$

“Use the Matérn model” – Stein (1999, pp. 14)

1-D Realizations from Matérn Model with Fixed σ^2, ρ

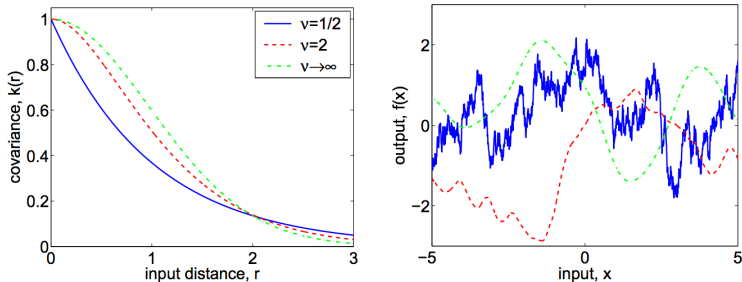
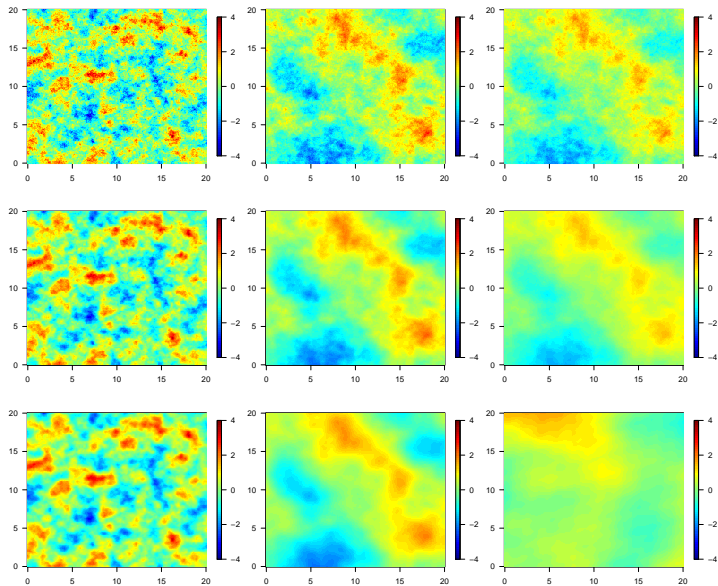


Figure: courtesy of Rasmussen & Williams 2006

2-D Realizations from Matérn Model with Fixed σ^2



1 Gaussian Process Spatial Model

2 Spatial Interpolation

3 Parameter estimation

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

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

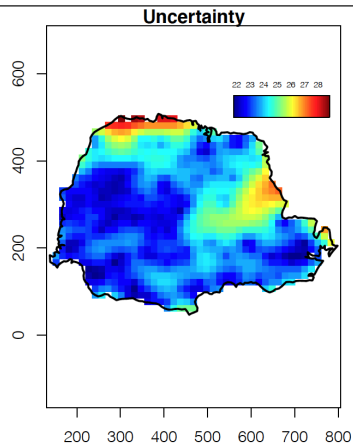
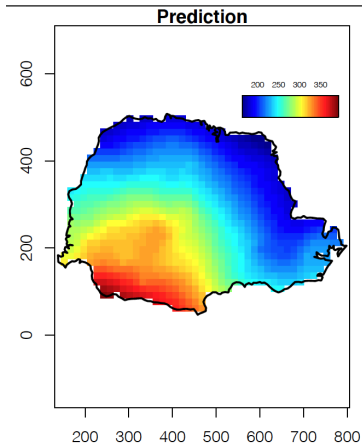
We have

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

where

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

Spatial Prediction of Paraná State Rainfall



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

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

We have

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

where

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

Question: what if we don't know $m_0, \mathbf{m}, \sigma_0^2, \Sigma$?

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

1 Gaussian Process Spatial Model

2 Spatial Interpolation

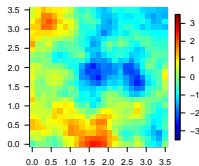
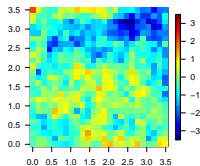
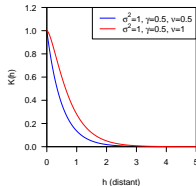
3 Parameter estimation

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

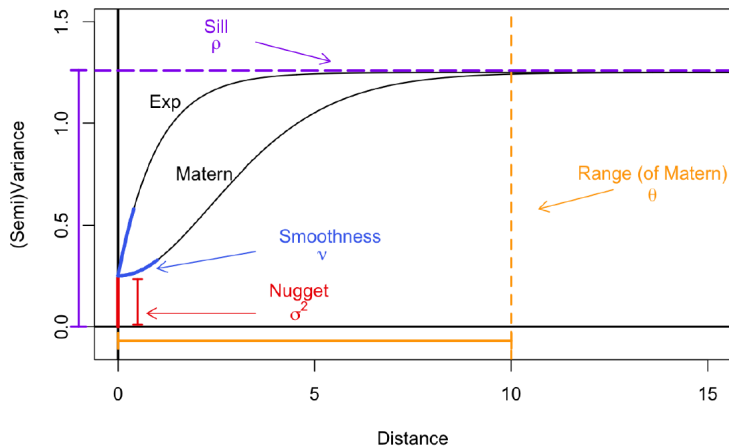
- Gaussian Processes GP $(m(\cdot), K(\cdot, \cdot))$ are widely used in modeling spatial stochastic processes

- Spatial statisticians often focus on the covariance function.

e.g.
$$K(h) = \sigma^2 \frac{(\sqrt{2\nu}h/\gamma)^\nu \mathcal{K}_\nu(\sqrt{2\nu}h/\gamma)}{\Gamma(\nu)2^{\nu-1}}$$



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



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

Under the stationary and isotropic assumptions

Variogram:

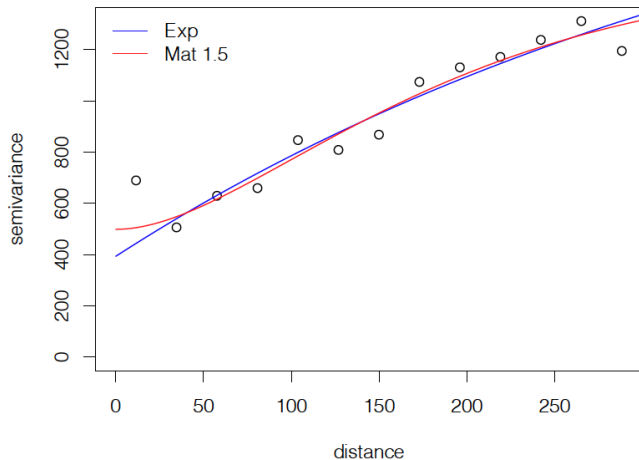
$$\begin{aligned}2\gamma(\mathbf{s}_i, \mathbf{s}_j) &= \text{Var}(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)) \\ &= \text{E} \left\{ \left((Y(\mathbf{s}_i) - \mu(\mathbf{s}_i)) - (Y(\mathbf{s}_j) - \mu(\mathbf{s}_j)) \right)^2 \right\} \\ &= \text{E} \left\{ (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2 \right\} \\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

Semivariogram and covariance function:

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

Estimation: Least Squares Method

$$\operatorname{argmin}_{\boldsymbol{\theta}} \sum_u \frac{n_u}{\gamma(h_u; \boldsymbol{\theta})^2} [\hat{\gamma}(h_u) - \gamma(h_u; \boldsymbol{\theta})]^2$$



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

Maximum Likelihood Estimation (MLE)

Log-likelihood:

Given data $\mathbf{y} = (y(s_1), \dots, y(s_n))^T$

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

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

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