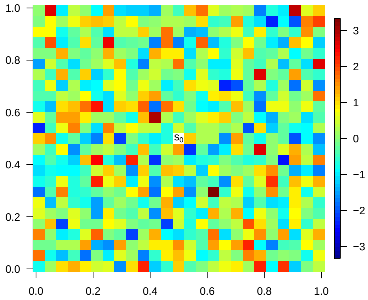
Lecture 27

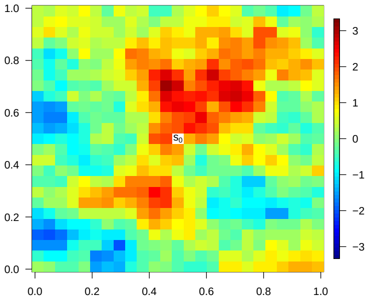
An Overview of Spatial Interpolation

STAT 8020 Statistical Methods II
December 3, 2020

Whitney Huang
Clemson University

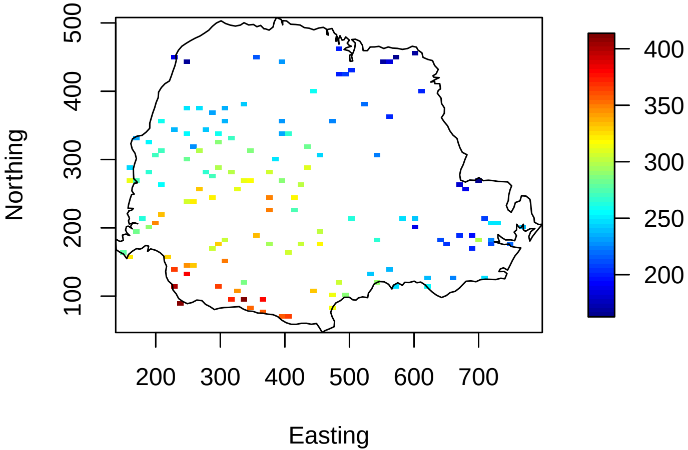
Toy Examples of Spatial Interpolation





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

$$y(\boldsymbol{s}_1), y(\boldsymbol{s}_2), \cdots, y(\boldsymbol{s}_n), \qquad \boldsymbol{s}_i \in \mathcal{S},\ i = 1, \cdots, n$$

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

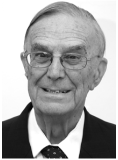
$$Y(\boldsymbol{s}_0), \qquad \boldsymbol{s}_0 \in \mathcal{S}$$

Applications

- Mining: ore grade
- Climate: temperature, precipitation, $\cdots$
- Remote Sensing: CO_2 retrievals
- Environmental Science: air pollution levels, $\cdots$

Some History

- Mining (Krige 1951) Matheron (1960s), Forestry (Matérn 1960)



- More recent work: Cressie (1993) Stein (1999)


Outline

1. **Gaussian Process Spatial Model**
2. **Spatial Interpolation**
3. **Parameter estimation**

Linear Interpolation

The best guess (in a statistical sense) should be based on the conditional distribution $[Y(\boldsymbol{s}_0)|\boldsymbol{Y}=\boldsymbol{y}]$ where

$$\boldsymbol{y} = (y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^{\mathrm{T}}$$

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

$$\hat{Y}(\boldsymbol{s}_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i y(\boldsymbol{s}_i)$$

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

Gaussian Process (GP) Spatial Model

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

Model:

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

where

- Mean function:

$$m(\boldsymbol{s}) = \mathrm{E}\left[Y(\boldsymbol{s})\right] = \boldsymbol{X}^T(\boldsymbol{s})\boldsymbol{\beta}$$

- Covariance function:

$$\{\epsilon(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}} \sim \mathrm{GP}\left(0, K\left(\cdot,\cdot\right)\right), \quad K(\boldsymbol{s}_1,\boldsymbol{s}_2) = \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1),\epsilon(\boldsymbol{s}_2)\right)$$

Assumptions on Covariance Function

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

- Stationarity:

$$K(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1), \epsilon(\boldsymbol{s}_2)\right) = C(\boldsymbol{s}_1 - \boldsymbol{s}_2)$$
$$= \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1 + \boldsymbol{h}), \epsilon(\boldsymbol{s}_2 + \boldsymbol{h})\right)$$

- Isotropy:

$$K(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1), \epsilon(\boldsymbol{s}_2)\right) = C(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|)$$

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

A covariance function is positive if

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

for any finite locations $s_1, \cdots, s_n$, and for any constants a_i, $i = 1, \cdots, n$

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:

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

- Spherical:

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

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

- Matérn:

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

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

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

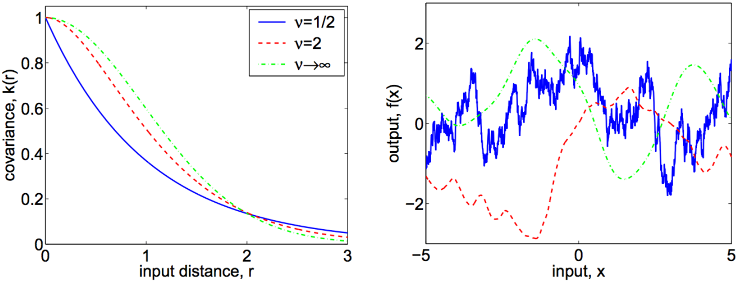
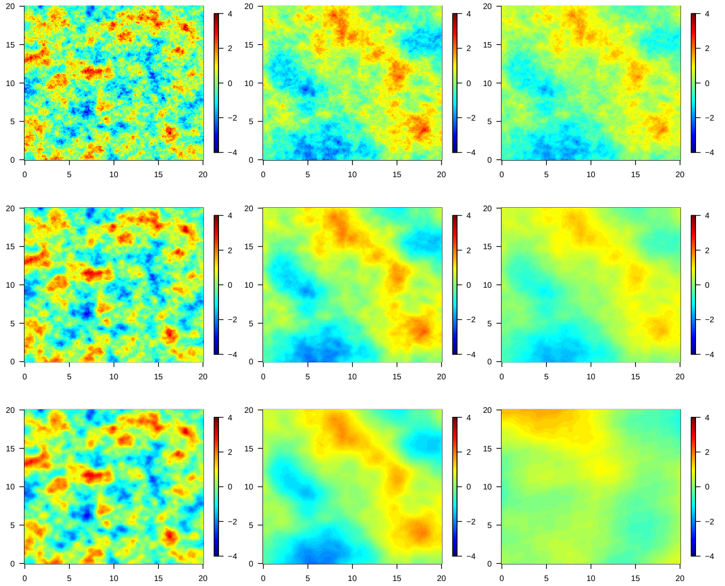


Figure: courtesy of Rasmussen & Williams 2006

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



Outline

Conditional Distribution of Multivariate Normal

If

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

Then

$$[\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2] \sim \mathrm{N}\left(\boldsymbol{\mu_{1|2}}, \Sigma_{1|2}\right)$$

where

$$\begin{aligned} \boldsymbol{\mu_{1|2}} &= \boldsymbol{\mu_1} + \Sigma_{12}\Sigma_{22}^{-1}\left(\boldsymbol{y}_2 - \boldsymbol{\mu_2}\right) \\ \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

GP-Based Spatial Interpolation: Kriging

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

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

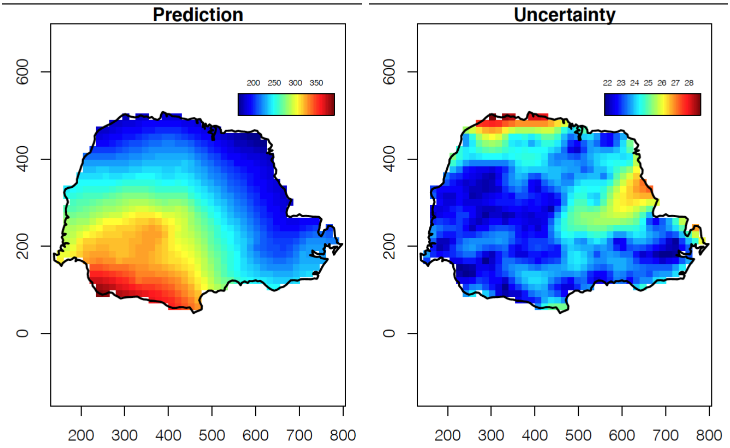
We have

$$[Y_0|\boldsymbol{Y} = \boldsymbol{y}] \sim \mathrm{N}\left(m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}}, \sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}}\right)$$

where

$$m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = m_0 + k^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$
$$\sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = \sigma_0^2 - k^{\mathrm{T}}\Sigma^{-1}k$$

Spatial Prediction of Paraná State Rainfall



GP-Based Spatial Interpolation: Kriging

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

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

We have

$$[Y_0|\boldsymbol{Y} = \boldsymbol{y}] \sim \mathrm{N}\left(m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}}, \sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}}\right)$$

where

$$m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = m_0 + k^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$
$$\sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = \sigma_0^2 - k^{\mathrm{T}}\Sigma^{-1}k$$

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

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

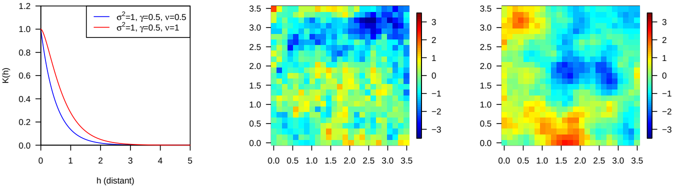
Outline

1. Gaussian Process Spatial Model
2. Spatial Interpolation
3. **Parameter estimation**

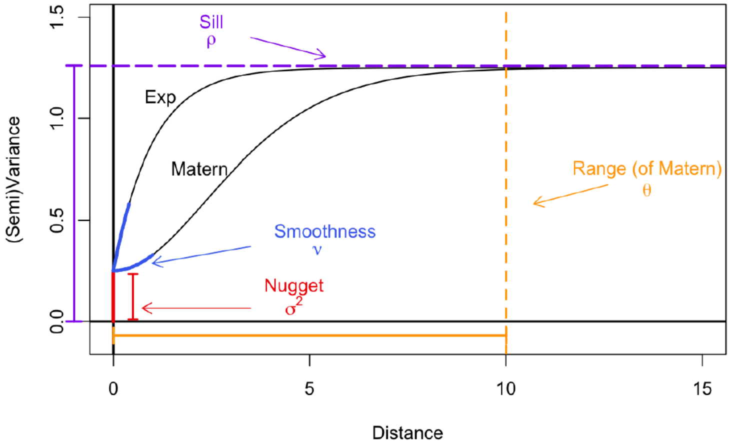
Gaussian Process

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

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



Semivariogram $\{\frac{1}{2}\mathrm{Var}\left(Y\left(\boldsymbol{s}_i, \boldsymbol{s}_j\right)\right)\}_{i,j}$



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

Variogram, Semivariogram, and Covariance Function

Under the stationary and isotropic assumptions

Variogram:

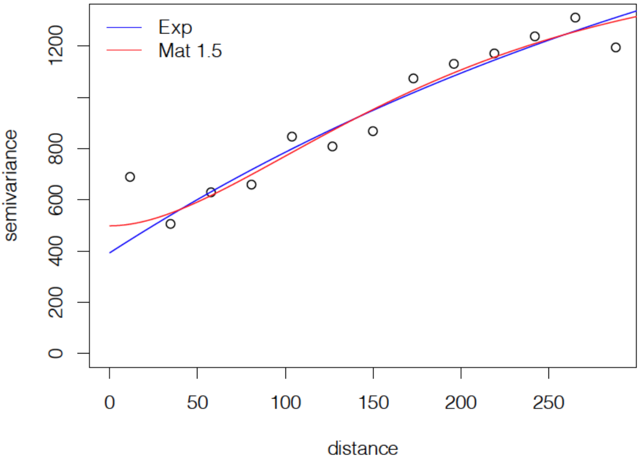
$$\begin{aligned} 2\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) &= \mathrm{Var}\left(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)\right) \\ &= \mathrm{E}\left\{\left((Y(\boldsymbol{s}_i) - \mu(\boldsymbol{s}_i)) - (Y(\boldsymbol{s}_j) - \mu(\boldsymbol{s}_j))\right)^2\right\} \\ &= \mathrm{E}\left\{\left(Y(\boldsymbol{s}_i)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

Semivariogram and covariance function:

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

Estimation: Least Squares Method

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



Maximum Likelihood Estimation (MLE)

Log-likelihood:

Given data $\boldsymbol{y} = (y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^{\mathrm{T}}$

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

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

Maximum Likelihood Estimation (MLE)

Log-likelihood:

Given data $\boldsymbol{y} = (y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^{\mathrm{T}}$

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

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

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

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

Then we obtain the profile log likelihood

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

where

$$P(\boldsymbol{\theta}) = \Sigma^{-1}_{\boldsymbol{\theta}} - \Sigma^{-1}_{\boldsymbol{\theta}} \boldsymbol{X} \left(\boldsymbol{X}^{\mathrm{T}} \Sigma^{-1}_{\boldsymbol{\theta}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{T}} \Sigma_{\boldsymbol{\theta}}$$

Solve the maximization problem above to get the MLE