Lecture 14

Interpolation of Spatial Data I

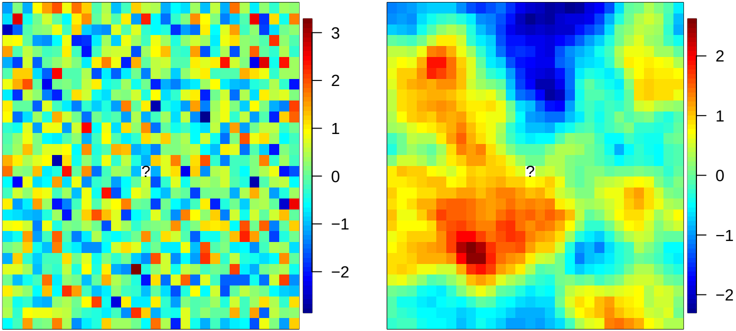
DSA 8020 Statistical Methods II
April 18-22, 2022

Whitney Huang
Clemson University

Agenda

Toy Examples of Spatial Interpolation

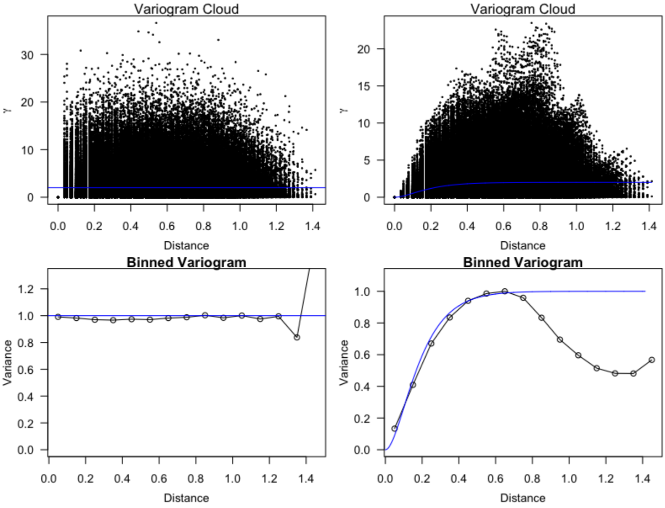
Let's consider two spatial images, each with a missing pixel



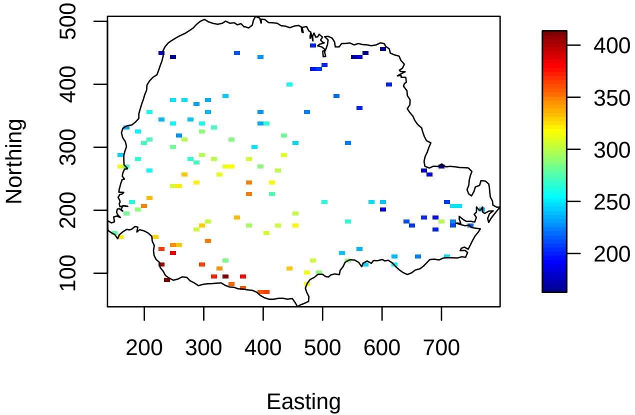
Question: What is your best guess of the value of the missing pixel, denoted as $Y(s_0)$, for each case?

Visualizing Spatial Dependence Structure

Similar to time series analysis, we can compute the covariance between data points in space *to examine the degree of spatial dependence*



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**

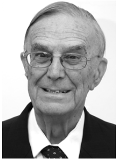
$$Y(\boldsymbol{s}_0), \qquad \boldsymbol{s}_0 \in \mathcal{S}$$

Applications

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

Some History of Spatial Statistics

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



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


Outline

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

Next, we will introduce a class of spatial model where the distribution is fully determined by its mean and covariance

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!

A covariance function is positive definite (p.d.) if

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

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

Question: what is the consequence if a covariance function is NOT p.d.? $\Rightarrow$ We can get a negative variance

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

- Using a parametric covariance function (see some examples in next slide)
- Using Bochner's Theorem to construct a valid covariance function

Some Commonly Used Covariance Functions

- Powered exponential:

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

- Spherical:

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

  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

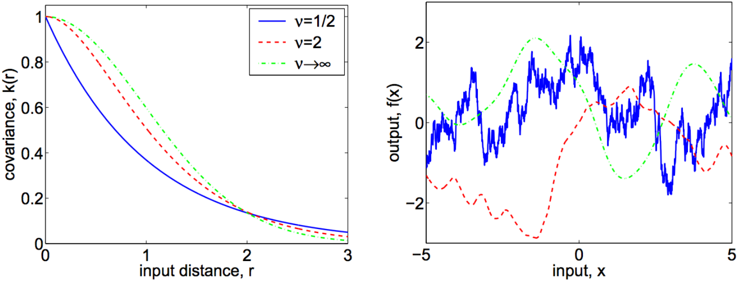
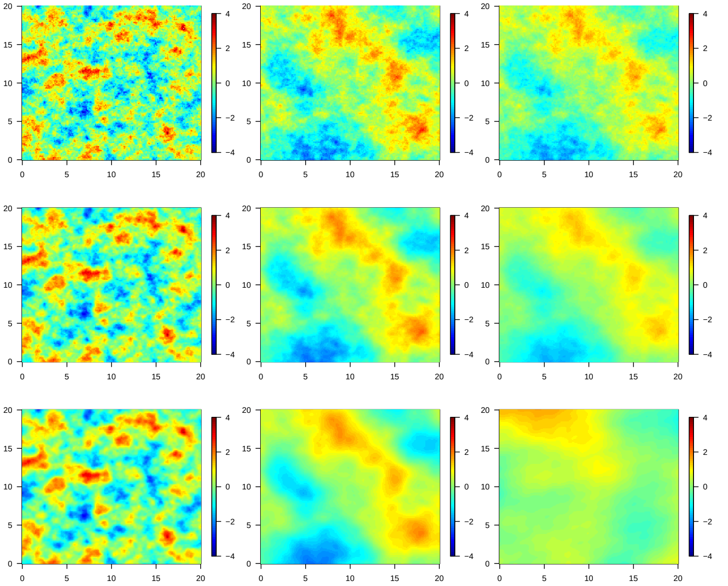


Figure: courtesy of Rasmussen & Williams 2006

The larger ν is, the smoother the process is

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

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

where

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

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

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

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

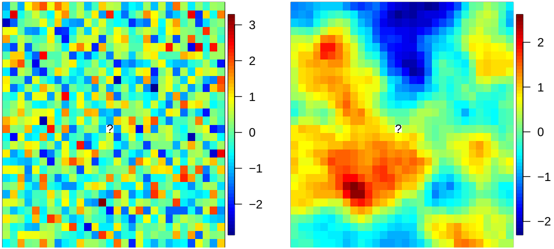
where

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

Next, we are going to revisit our toy examples

Toy Examples Revisited

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



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

Spatial uncorrelated field:

- $m_{Y_0|\boldsymbol{Y}} = 0$
- $\sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = \sigma_0^2$

Spatial correlated field:

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

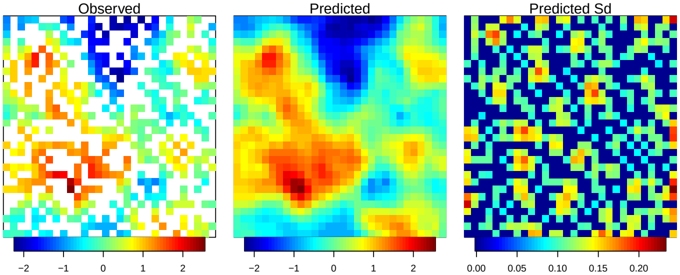
Interpolating Multiple Points in Space

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

$$[\boldsymbol{Y}_0|\boldsymbol{Y} = \boldsymbol{y}] \sim \mathrm{N}\left(\boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}}, \Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}}\right)$$

where

$$\boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} = \boldsymbol{m}_0 + \boldsymbol{k}^{\mathrm{T}}\Sigma^{-1}\left(\boldsymbol{y} - \boldsymbol{m}\right)$$
$$\Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} = \Sigma_0 - \boldsymbol{k}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{k}$$



GP-Based Spatial Interpolation: Kriging

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

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

We have

$$[\boldsymbol{Y}_0 | \boldsymbol{Y} = \boldsymbol{y}] \sim \mathrm{N}\left(\boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}}, \Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} \right)$$

where

$$\begin{aligned} \boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} &= \boldsymbol{m}_0 + \boldsymbol{k}^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{m}) \\ \Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} &= \Sigma_0 - \boldsymbol{k}^{\mathrm{T}} \Sigma^{-1} \boldsymbol{k} \end{aligned}$$

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

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

Bochner's Theorem

A complex-valued function C on $\mathbb{R}^{\mathrm{d}}$ is the covariance function for a weakly stationary mean square contituous complex-valued random process on $\mathbb{R}^d$ if and only if it can be represented as

$$C(\boldsymbol{h}) = \int_{\mathbb{R}^d} \exp(i\omega^{\mathrm{T}}\boldsymbol{h})F(d\boldsymbol{\omega}),$$

with F a positive finite measure. When F has a density with respect to Lebesgue measure, we have the spectral density f and

$$f(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}^d} \exp(-i\omega^{\mathrm{T}}\boldsymbol{h})C(\boldsymbol{h})\, d\boldsymbol{h}$$