An Introduction to Kriging Part II

Whitney Huang

CLEMSON UNIVERSITY

wkhuang@clemson.edu

Clemson ENVR Group, September 16, 2020

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:

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

- 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:

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

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

Question: what is the consequence if a covariance function is NOT p.d.? $\Rightarrow$ 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)

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

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

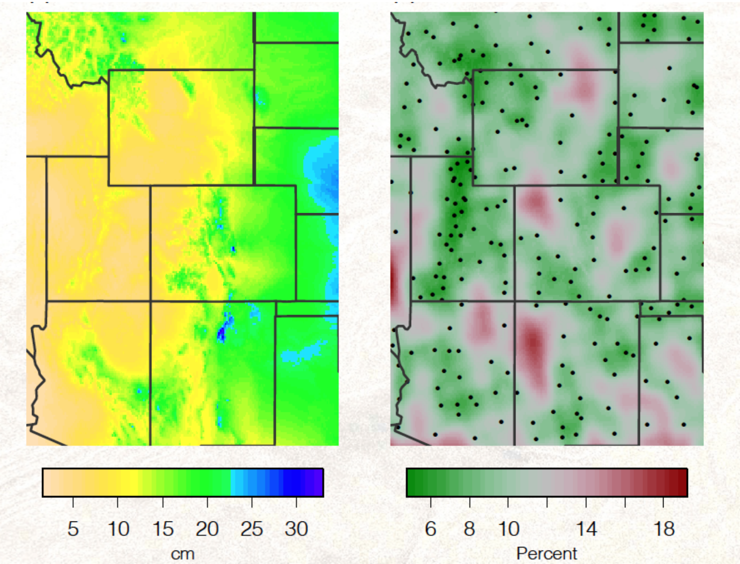
We have

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

where

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

Estimated "Summer" 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

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

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 $\boldsymbol{y}$.

Estimation: MLE

Log-likelihood:

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

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

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

for any fixed $\theta_0 \in \Theta$ the unique value of β that maximizes ℓ_n is given by

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

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

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

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

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

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

Asymptotics for spatial data

- MLE: (usually) consistency, asymptotic normality, efficient
- Two different asymptotic frameworks in spatial statistics: increasing-domain, fixed-domain

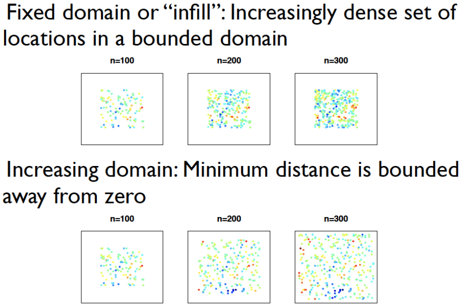
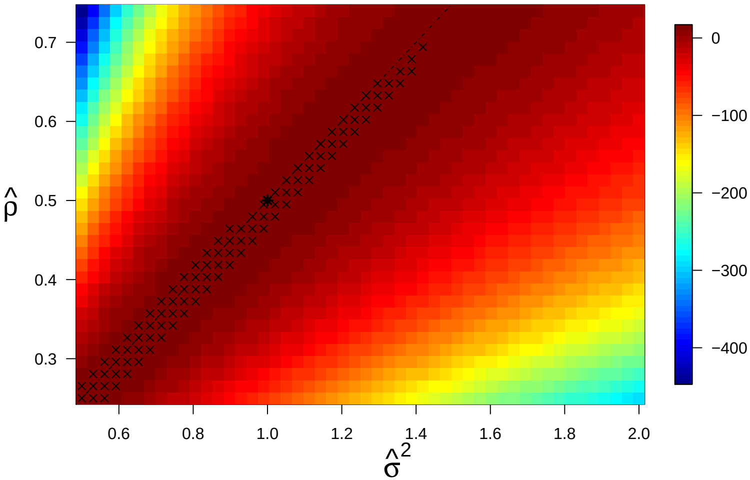


Figure: Figure courtesy of Cari Kaufman

- Inconsistent estimation and asymptotically equal interpolations in Model-Based Geostatistics (Zhang, 2004)

An Illustration of Inconsistent Estimation of GP Parameters



"Big n Problem"

- Modern environmental instruments have produced a wealth of space–time data $\Rightarrow$ n is big
- Evaluation of the likelihood function involves factorizing large covariance matrices that generally requires
  - $\mathcal{O}(n^3)$ operations
  - $\mathcal{O}(n^2)$ memory
- Modeling strategies are needed to deal with large spatial data set.
  - parameter estimation $\Rightarrow$ MLE, Bayesian
  - spatial interpolation $\Rightarrow$ Kriging
  - multivariate spatial data $(np \times np)$, spatio-temporal data $(nt \times nt)$