Lecture 3: Kriging and parameter estimation

Spatial Statistics and Image Analysis

David Bolin
University of Gothenburg

Gothenburg
April 1, 2019

Semivariograms

- In geostatistics, it is common do describe random fields in terms of semivariograms instead of covariance functions.
- For a random field $X(\mathbf{s})$, the semivariogram is defined as

$$\gamma(\mathbf{s}, \mathbf{t}) = \frac{1}{2}\mathsf{V}(X(\mathbf{s}) - X(\mathbf{t}))$$

  and the variogram is $\mathsf{V}(X(\mathbf{s}) - X(\mathbf{t}))$.
- For an isotropic random field with covariance $r(h)$, the semivariogram is

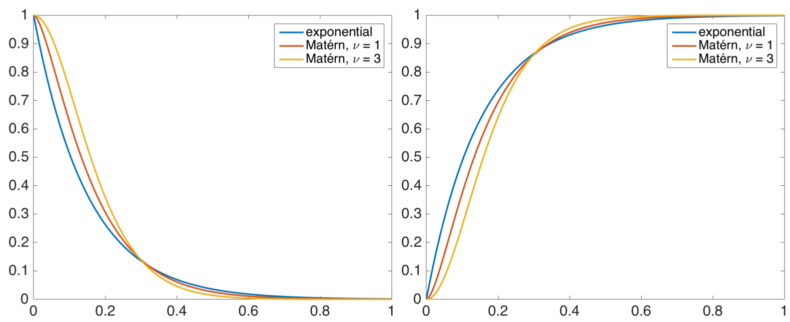
$$\gamma(h) = r(0) - r(h)$$

  (exersice!)

Gaussian random fields

- A random field $X(\mathbf{s})$ is Gaussian if $(X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))^T$ has a multivariate Gaussian distribution for each choice of $\mathbf{s}_1, \ldots, \mathbf{s}_n$.
- $X(\mathbf{s})$ is uniquely specified by
  1. The mean value function $\mu(\mathbf{s}) = \mathsf{E}(X(\mathbf{s}))$, and
  2. The covariance function $r(\mathbf{s}_1, \mathbf{s}_2) = \mathsf{C}(X(\mathbf{s}_1), X(\mathbf{s}_2))$.
- $X(\mathbf{s})$ is
  1. **stationary** if $\mu(\mathbf{s}) \equiv \mu$ and if $r(\mathbf{s}_1, \mathbf{s}_2)$ depends only on the separation between the locations, $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$.
  2. **isotropic** if $\mu(\mathbf{s}) \equiv \mu$ and if $r(\mathbf{s}_1, \mathbf{s}_2)$ only depends on the distance between the locations, $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$.
- Examples of isotropic covariance functions:
  1. Matérn: $r(h) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa h)^\nu K_\nu(\kappa h)$
  2. Exponential: $r(h) = \sigma^2 \exp(-\kappa h)$
  3. Spherical: $r(h) = \sigma^2(1 - \frac{3}{2}\frac{h}{\theta} + \frac{1}{2}\frac{h^3}{\theta^3})$, if $h \leq \theta$ and $r(h) = 0$ otherwise.

Matérn variograms



Some terminology

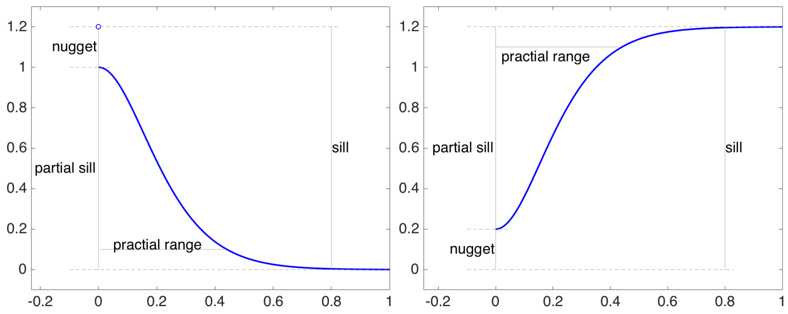
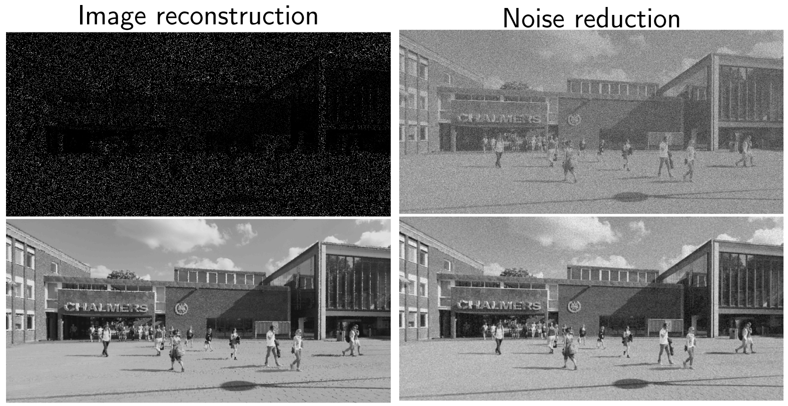


Image analysis applications



Statistical models including random fields

- We have measurements $y_i, \dots, y_n$ taken at some spatial locations $s_1, \dots, s_n$.
- Given that we also have some explanatory variables $B_1, \dots, B_K$, we use a model

$$Y_i = \sum_{k=1}^{K} B_k(s_i)\beta_k + X(s_i) + \varepsilon_i\,.$$

  where $X(s)$ is a mean-zero Gaussian random field.
- Questions:
  1. How do we estimate the parameters of the model?
  2. How can we perform prediction for an unobserved location s_0?

Conditional distributions

For a multivariate Gaussian variable

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

we have that

$$\mathbf{X}_2|\mathbf{X}_1 \sim \mathsf{N}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$$

If $\mathbf{X}_2$ represents a random field at some unobserved locations, and $\mathbf{X}_1$ the observations, the conditional mean

$$\mathsf{E}(\mathbf{X}_2|\mathbf{X}_1) = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1)$$

is often called the Kriging predictor.

Kriging prediction

Traditionally, one has separated between three cases

- Simple kriging: $\mu(\mathbf{s}) = \mathbf{B}(\mathbf{s})\boldsymbol{\beta}$ is known.
- Ordinary kriging: $\mu(\mathbf{s}) = \beta$ is unknown but constant.
- Universal kriging: $\mu(\mathbf{s}) = \mathbf{B}(\mathbf{s})\boldsymbol{\beta}$ is unknown.

For ordinary and universal kriging, we have to estimate the mean-value together with the covariance parameters $\boldsymbol{\theta}$ before computing the prediction.

So we have to

- Estimate the model parameters $\{\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\theta}\}$.
- Given the parameters, compute the kriging prediction.

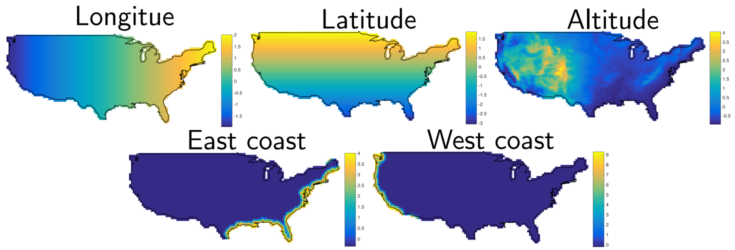
UNIVERSITY OF GOTHENBURG CHALMERS

Covariates

- A first idea is to use linear regression to interpolate the data:

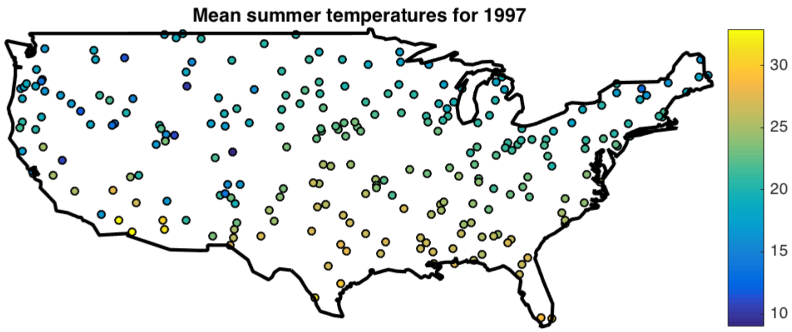
$$Y(\mathbf{s}) = \sum_{i=1}^{k} \beta_i B_i(\mathbf{s}) + \varepsilon_{\mathbf{s}}, \quad \text{where } \varepsilon_{\mathbf{s}} \text{ are iid } \mathsf{N}(0, \sigma^2)$$

- Possible covariates



UNIVERSITY OF GOTHENBURG CHALMERS

Example: US temperatures



- Mean summer temperatures (June-August) in the continental US 1997 recorded at 250 weather stations.
- We want to estimate all US temperatures based on the data.

UNIVERSITY OF GOTHENBURG CHALMERS

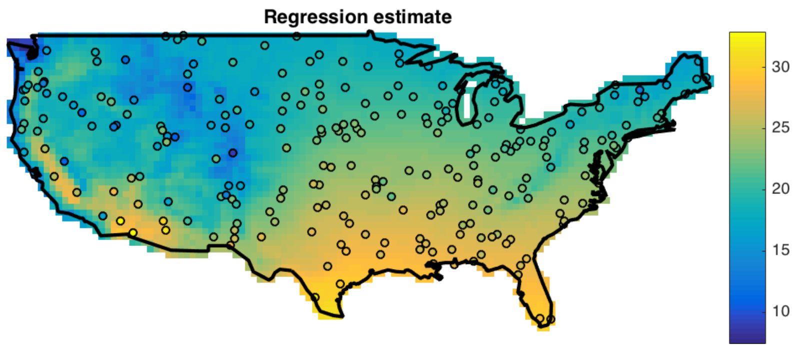
OLS estimate

- Estimate the parameters using ordinary least squares:

$$\hat{\boldsymbol{\beta}} = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{Y},$$

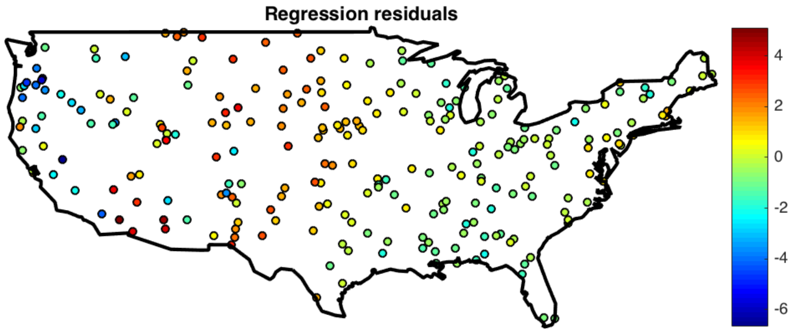
where $\mathbf{B}_{ij} = B_i(\mathbf{s}_j)$ and $\mathbf{Y}_i = Y(\mathbf{s}_i)$.

- Calculate the prediction $\hat{X}(\mathbf{s}) = \sum_{i=1}^{k} \hat{\beta}_i B_i(\mathbf{s})$.



Residudals

- How do we test whether the prediction is reasonable?
- If the model assumptions hold, the residuals $Y(\mathbf{s}) - \hat{X}(\mathbf{s})$ should be independent identically distributed.



UNIVERSITY OF GOTHENBURG CHALMERS

Regression parameters

Update regression parameters using GLS:

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{B}^\top \boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}^\top \boldsymbol{\Sigma}^{-1}\mathbf{Y},$$

Confidence interval for β_i:

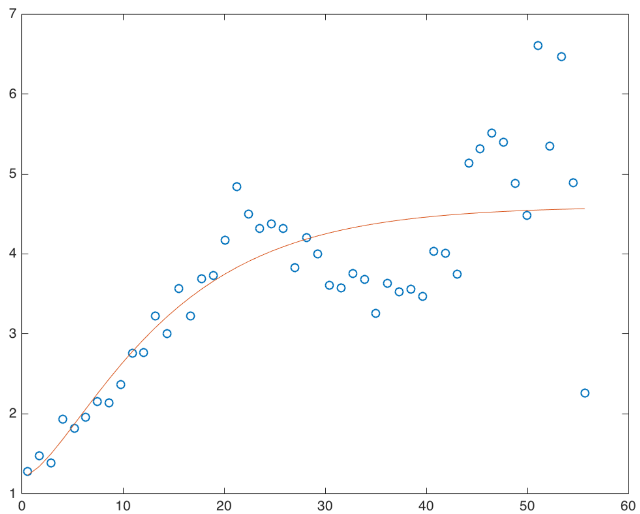
$$I_{\beta_i} = (\hat{\beta}_i \pm 1.96\sqrt{V_{ii}})$$

where $\mathbf{V} = (\mathbf{B}^\top \boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1}$.

	OLS	GLS
Intercept	$21.6317*$	$20.4688*$
Longitude	$-1.2897*$	-1.0022
Latitude	$-2.6959*$	$-2.6845*$
Altitude	$-2.6693*$	$-4.2177*$
East coast	-0.0952	-0.0096
West coast	$-1.3064*$	$-1.0139*$

UNIVERSITY OF GOTHENBURG CHALMERS

Variogram estimate



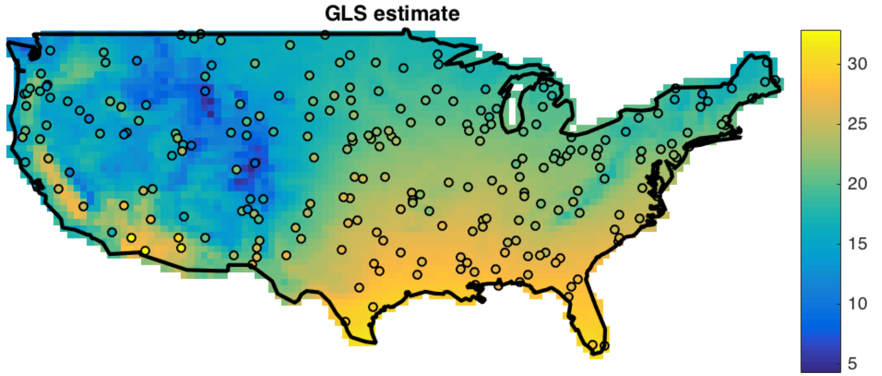
UNIVERSITY OF GOTHENBURG CHALMERS

Kriging estimation

The kriging estimator is

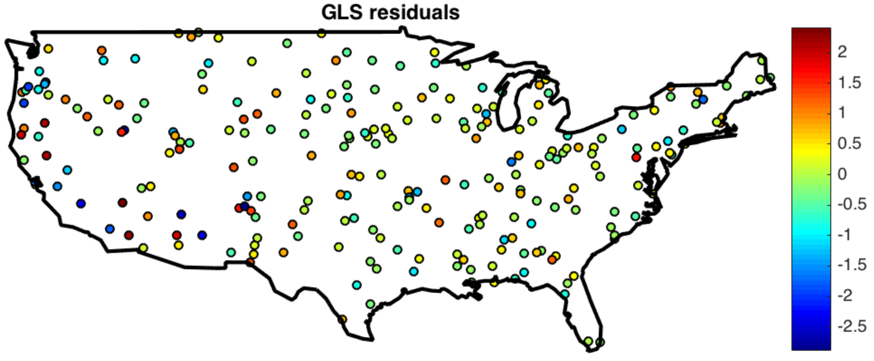
$$\mathsf{E}(X(\mathbf{s})|\mathbf{Y}, \hat{\boldsymbol{\theta}}) = \hat{\mu}(\mathbf{s}) + \mathbf{r}(\boldsymbol{\Sigma} + \sigma_e^2\mathbf{I})^{-1}(\mathbf{Y} - \mathbf{B}\hat{\boldsymbol{\beta}})$$

where $\boldsymbol{\Sigma}_{ij} = r(\mathbf{s}_i, \mathbf{s}_j)$, $\mathbf{r}_i = r(\mathbf{s}, \mathbf{s}_i)$, and $\hat{\boldsymbol{\mu}} = \sum_{k=1}^{K} B_k(\mathbf{s})\hat{\beta}_k$.



Kriging residuals

There is now less spatial structure in the residuals.



Empirical covariances of residuals