Spatial statistics and image analysis (TMS016/MSA301)

Kriging: estimation

2021-04-12

Practical information

Student representatives

- Muhammad Abdullah (MPDSC)
- Axel Andersson (MPENM)
- Anastasios Tzinieris (MPDSC)
- Bharath Venkatesh Poojary (MPDSC)

Project groups

- Could those who are alone in a group think of working with somebody else?

Kriging: estimation

We have measurements $y_i, i = 1, ...n$ at spatial locations $s_1, ..., s_n$ and we assume that

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

where

- $B_1, ..., B_k$ are exploratory variables and $\beta_1, ..., \beta_K$ unknown parameters (mean)
- $X = (X(s_i), s \in S)$ is a zero mean Gaussian random field
- $\epsilon_1, ..., \epsilon_n$ are mutually independent zero mean normal random variables with variance σ_ϵ^2 and independent of X

Kriging prediction

For column vectors X_1 and X_2 with a joint Gaussian distribution,

$$\left(\begin{array}{c} X_1 \\ X_2 \end{array}\right) \sim N\left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right), \left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right)\right)$$

If X_2 represents a random field at some unobserved locations and X_1 represent the observations, the conditional mean

$$\mathbb{E}[X_2|X_1] = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1).$$

is called the kriging predictor at the unobserved locations.

Kriging

Different types of kriging

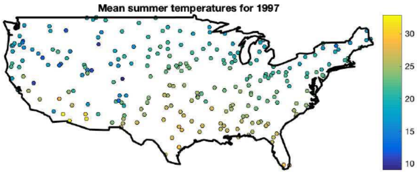
- Simple kriging: $\mu(s) = B(s)\beta$ is known
- Ordinary kriging: $\mu(s) = \beta$ is unknown but constant (no covariates)
- Universal kriging: $\mu(s) = B(s)\beta$ is unknown

We have to estimate the mean parameters β and the covariance parameters Θ before we can compute any predictions. Therefore, we

- estimate the model parameters β, Θ and σ_ϵ^2.
- given the parameter estimates, compute the kriging prediction.

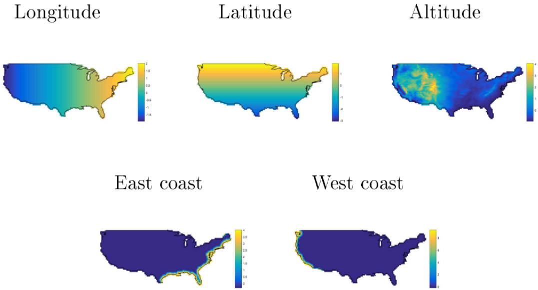
Example: US temperatures

- Mean summer (June-August) temperatures in the continental US in 1997 recorded at 250 (n) weather stations
- We would like to estimate temperatures in the whole country during this time based on the data.



Example: covariates

We have five covariates: longitude, latitude, altitude, east coast, and west coast.



Example: linear regression

First, we use linear regression and interpolate the data using only some covariates, i.e.

$$Y(s) = \sum_{k=0}^{5} B_k(s)\beta_k + \epsilon_s,$$

where ϵ_s are iid $N(0, \sigma_\epsilon^2)$ and β_0 is the intercept for which we set $B_0(s) = 1$.

The model can also be written in a matrix form as

$$Y = B\beta + \epsilon,$$

where $\epsilon \sim N(0, \sigma_\epsilon^2 \mathbb{I})$ and $\mathbb{I}$ is the identity matrix.

Estimation: Ordinary least square (OLS) estimates

To estimate the parameters in β, we minimize the sum of squared residuals

$$(Y - B\beta)^T(Y - B\beta)$$

with respect to β. This gives us the estimates

$$\hat{\beta} = (B^T B)^{-1} B^T Y.$$

A prediction of the mean temperature at location s is then

$$\hat{Y}(s) = \sum_{k=0}^{K} B_k(s)\hat{\beta}_k$$

or (for the set of locations)

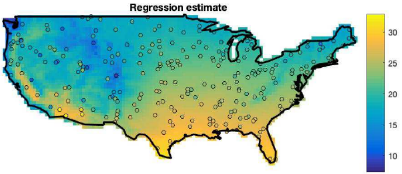
$$\hat{Y}_{\text{OLS}} = B\hat{\beta}_{\text{OLS}},$$

where $\hat{\beta}_{\text{OLS}}$ is estimated parameter vector.

Example: OLS estimates

Covariate	$\hat{\beta}$ (OLS)
Intercept	21.63*
Longitude	-1.29^*
Latitude	-2.70^*
Altitude	-2.67^*
East coast	-0.10
West coast	-1.31^*

The parameter estimates that are significantly different from zero are indicated by $*$.

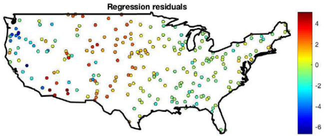


Residuals

To check the goodness-of-fit of the model, we can look at the residuals

$$Y(s) - \hat{Y}(s)$$

at the measured locations. These should be independent and identically distributed.



Residuals at locations close together seem to be highly correlated.
$\rightarrow$ Model could be improved.

Estimation: Generalized least square (GLS) estimates

To improve the model, we can add dependent errors, i.e.

$$Y = B\beta + \epsilon,$$

where $\epsilon \sim N(0, \Sigma)$, where Σ is a (positive definite) covariance matrix.

The resulting generalized least squares estimates are given by

$$\hat{\beta}_{\text{GLS}} = (B^T \Sigma^{-1} B)^{-1} B^T \Sigma^{-1} Y$$

and the estimates at the unknown locations by

$$\hat{Y}_{\text{GLS}} = B\hat{\beta}_{\text{GLS}}.$$

How to estimate the covariance function?

We can start by looking at the OLS residuals

$$\hat{\epsilon}_i = y_i - \sum_{k=1}^{K} B_k(s_i)\hat{\beta}_k$$

that can be computed at every measured location $s_i, \; i = 1, ..., n$.

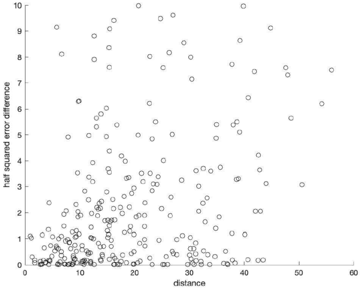
The half squared residual differences

$$v_{ij} = 0.5(\hat{\epsilon}_i - \hat{\epsilon}_j)^2$$

show how the error residuals vary with the distance $r_{ij} = |s_i - s_j|$ between the locations s_i and s_j. .

Example: Residual plot

The half squared residual differences $v_{ij} = 0.5(\hat{\epsilon}_i - \hat{\epsilon}_j)^2$ plotted against the distances r_{ij}. (Only 1% of the $250 \times 249/2 = 31125$ values are plotted and values with v_{ij} larger than 10 are omitted.)



v_{ij} tends to increase with increasing r_{ij}.

Example: Binned residuals

The increasing trend can be better seen if we bin the values: The distance values are divided into subintervals $I_l, \ l = 1, ..., L$ of equal length.

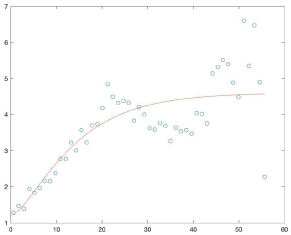
Let H_l denote the set of distance pairs r_{ij} in the interval I_l and $|H_l|$ the number of v_{ij}'s in the lth bin H_l. Then, we plot the averages of the half squared distances in the subintervals

$$\bar{v}_l = \frac{1}{|H_l|} \sum_{r_{ij} \in H_l} v_{ij}, \quad l = 1, ..., L,$$

against the midpoints of the bins.

Example: Binned residuals with an estimated semivariogam

The Matérn semivariogram is fitted to the binned residuals.



The final kriging estimates are

$$\mathbb{E}[Y(s)|Y] = \sum_{k=0}^{K} B_k(s)\hat{\beta}_k + C(\Sigma + \sigma_e^2 \mathbb{I})^{-1}(Y - B\hat{\beta}),$$

where C is a vector of values $C(s, s_i),\ s = 1, ..., n.$

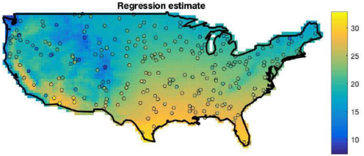
Example: GLS estimates

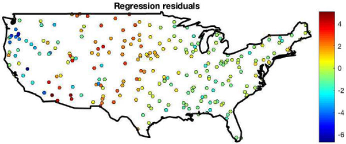
Covariate	$\hat{\beta}$ (OLS)	$\hat{\beta}$ (GLS)
Intercept	20.63^*	20.47^*
Longitude	-1.29^*	-1.00
Latitude	-2.70^*	-2.68^*
Altitude	-2.67^*	-4.22^*
East coast	-0.10	-0.01
West coast	-1.31^*	-1.01^*

The parameter estimates that are significantly different from zero are indicated by $*$.

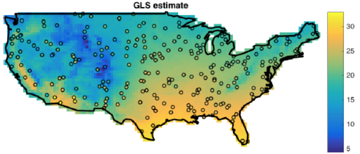
Example: OLS versus GLS

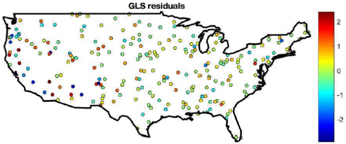
OLS estimates and residuals





GLS estimates and residuals





Estimation: Maximum likelihood (ML)

If Y is a Gaussian field, e.g. with Matérn covariance function, then

$$Y \sim N(B\beta, \Sigma(\Theta')),$$

where $\Theta' = (\sigma^2, \nu, \theta, \sigma_0^2, \sigma_\epsilon^2)$ and σ_0^2 is the nugget effect corresponding to the covariance function.

Therefore, we can write down the log-likelihood

$$\begin{aligned} l(Y; \beta, \Theta') &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma(\Theta')|) \\ &\quad -\frac{1}{2}(Y - B\beta)^T \Sigma(\Theta')^{-1}(Y - B\beta) \end{aligned}$$

and maximize it with respect to the parameters.

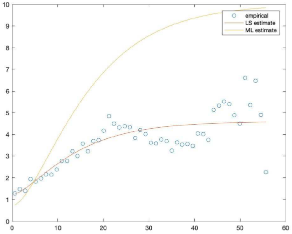
Profile likelihood

To make the computations easier, one can use profile likelihood:

- First, maximize the log-likelihood function with respect to β for given Θ'.
- Then, maximize the log-likelihood $l(Y; \hat{\beta}(\Theta'), \Theta')$ with respect to Θ'.

Example: Comparison of all the estimates

Covariate	$\hat{\beta}$ (OLS)	$\hat{\beta}$ (GLS)	ML
Intercept	20.63*	20.47*	19.80*
Longitude	−1.29*	−1.00	-0.53
Latitude	−2.70*	−2.68*	−2.64*
Altitude	−2.67*	−4.22*	−4.35*
East coast	−0.10	−0.01	0.02
West coast	−1.31*	−1.01*	−0.93*
$\hat{\sigma}$		1.84	3.05
$\hat{\nu}$		1.00	1.19
$\hat{\theta}$		9.38	10.20
$\hat{\sigma}_0$		1.09	0.81
$\hat{\sigma}_\epsilon$	1.81	1.10	0.85



ν and θ, and σ are the parameters of the Matérn covariance function, σ_0 the nugget effect, and σ_ϵ the residual standard deviation.

Comment on ML estimation

- ML estimators $(\hat{\beta}, \hat{\Theta}')$ may be biased, especially if the number of covariates, i.e. the number of parameters in β, is large.
- For example, the maximum likelihood estimate of the error variance is $\frac{1}{n} \sum e_i^2$ but the corresponding unbiased estimate is $\frac{1}{n-p} \sum e_i^2$, where p is the number of parameters in β.

  $\rightarrow$ restricted maximum likelihood (REML) (estimates the parameters by using $n - p$ linearly independent contrasts)