Spatial statistics and image analysis. Lecture 3

Mats Rudemo

March 29, 2020

Background

Read LN Section 15.8, "Parameter estimation. Likelihood and least squares".

Read LN Section 15.9, "Linear and logistic regression", in particular equations (15.42)-(15.45). Thus if

$$Y = X\beta + \epsilon$$

the ordinary least squares estimate of β is

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

and this is what we use to compute the OLS estimate

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

in LN equ. (5.28), corresponding to equ. (4) below in this lecture.

Statistical models for observations of random fields

Measurements $Y_i, i = 1, \ldots, n$, at spatial locations $s_1, \ldots, s_n$

$B_1, \ldots, B_K$ explanatory variables

Model

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

Here $X = (X(s), s \in S)$ is a Gaussian random field and $\epsilon_1, \ldots, \epsilon_n$ are $N(0, \sigma_\epsilon^2)$ independent mutually and of X

Example: mean summer temperatures in continental US recorded at 250 weather stations 1997, and five explanatory variables, covariates.

Figure 1 shows the mean summer temperatures

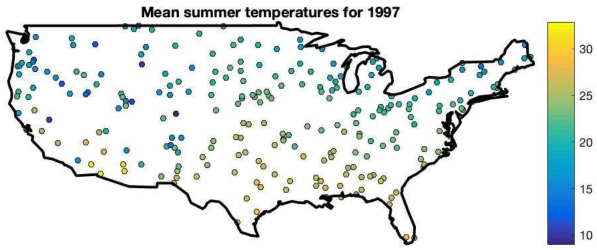

Figure 1: Mean summer temperatures for 1997 recorded at 250 weather stations in continental US.

The covariates we use are Longitude, Latitude, Altitude, East coast and West coast, see Figure 2.

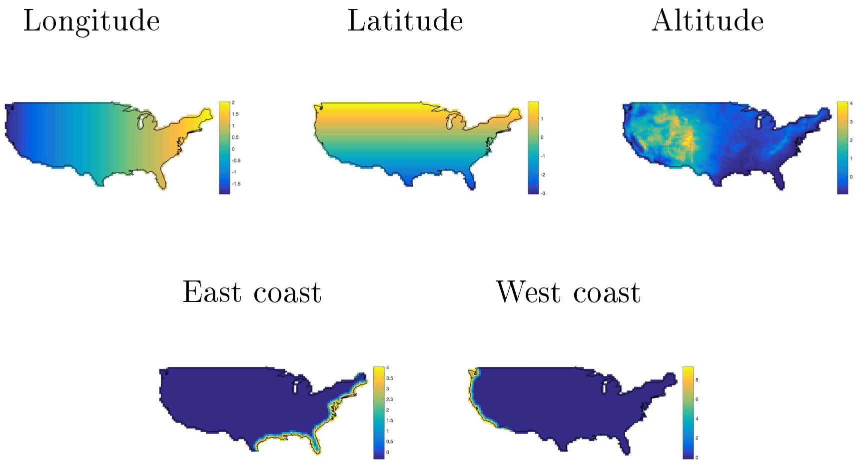

Figure 2: Five covariates used in the analysis of summer temperature in contintental US.

Ordinary least squares (OLS) analysis

First approach: use ordinary least squares with five covariates but without the random field X, that is, use the model

$$Y_i = \sum_{k=0}^{K} B_k(s_i)\beta_k + \epsilon_i \tag{2}$$

Here we also have included an intercept β_0 and correspondingly we put $B_0(s_i) = 1$. The model can also be written

$$Y = B\beta + \epsilon\,, \quad \epsilon \sim N(0, \sigma_\epsilon^2 I)\,. \tag{3}$$

Look at estimation of the parameters in (3).

Table 1: OLS (Ordinary Least Squares) analysis of US continental summer temperatures 1997 together with t-values (with 244 degrees of freedom) for test of the hypothesis that the corresponding parameter is zero. The residual standard deviation estimate is $\hat{\sigma}_\epsilon = 1.808$.

Explaining variable	Estimate $\hat{\beta}_k$	Corresponding t-value
Intercept	21.63	189.17
Longitude	-1.29	-8.15
Latitude	-2.70	-22.72
Altitude	-2.67	-18.33
East coast	-0.10	-0.74
West coast	-1.31	-10.24

Table 1 shows the parameter estimates

$$\hat{\beta}_{\text{OLS}} = (B^T B)^{-1} B^T Y \tag{4}$$

of the OLS analysis. The residual degrees of freedom is 250-6= 244. The column of t-values, corresponds to tests that the corresponding parameter is zero. The OLS regression surface estimate

$$\hat{Y}_{\text{OLS}} = B\hat{\beta}_{\text{OLS}} \tag{5}$$

of the temperature surface is shown in Figure 3 and the OLS regression residuals

$$\text{res}_{\text{OLS}} = Y - B\hat{\beta}_{\text{OLS}} \tag{6}$$

are shown in Figure 4. From Figure 4 we see that residuals close in location seem highly correlated, which indicates that the model could be improved.

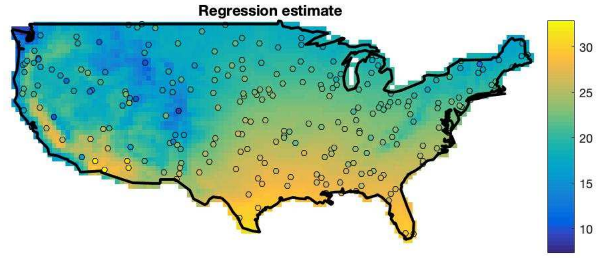

Figure 3: OLS regression temperature surface estimate

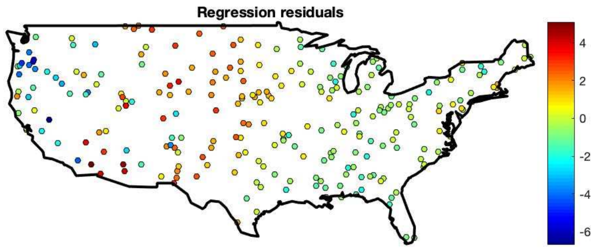

Figure 4: OLS regression temperature residuals

Generalized least squares (GLS) analysis

First improvement of the model (3): Assume that

$$Y = B\beta + \epsilon\,, \;\; \epsilon \sim N(0, \Sigma)\,, \tag{7}$$

where Σ is a general positive-definite covariance matrix. The least squares estimate of β is the GLS (Generalized Least Squares) estimate

$$\hat{\beta}_{\text{GLS}} = (B^T \Sigma^{-1} B)^{-1} B^T \Sigma^{-1} Y \tag{8}$$

with corresponding GLS regression surface estimate

$$\hat{Y}_{\text{GLS}} = B\hat{\beta}_{\text{GLS}} \tag{9}$$

and GLS regression residuals

$$\text{res}_{\text{GLS}} = Y - B\hat{\beta}_{\text{GLS}}\,. \tag{10}$$

One problem with GLS: Typically the covariance matrix Σ in (8) is unknown and has to be estimated.

Estimation of the covariance matrix Σ

Look at the OLS residuals

Let $s_i, i = 1, \ldots, n$, denote the locations for the measurements $Y_i, i = 1, \ldots, n$.

Let $\hat{\epsilon}_i, i = 1, \ldots, n$, denote the corresponding OLS residuals.

Let $r_{ij} = ||s_i - s_j||$ denote the distance between the locations s_i and s_j.

To see how OLS error residuals vary with the distance between the locations, plot half squared residual differences

$$v_{ij} = 0.5(\hat{\epsilon}_i - \hat{\epsilon}_j)^2 \tag{11}$$

against location distances r_{ij}. With $n = 250$ location we get $250 * 249/2 = 31\,125$ location pairs and plotting all would give a figure difficult to grasp.

In Figure 5 we have therefore randomly chosen and plotted 1% of all possible pairs (r_{ij}, v_{ij}).

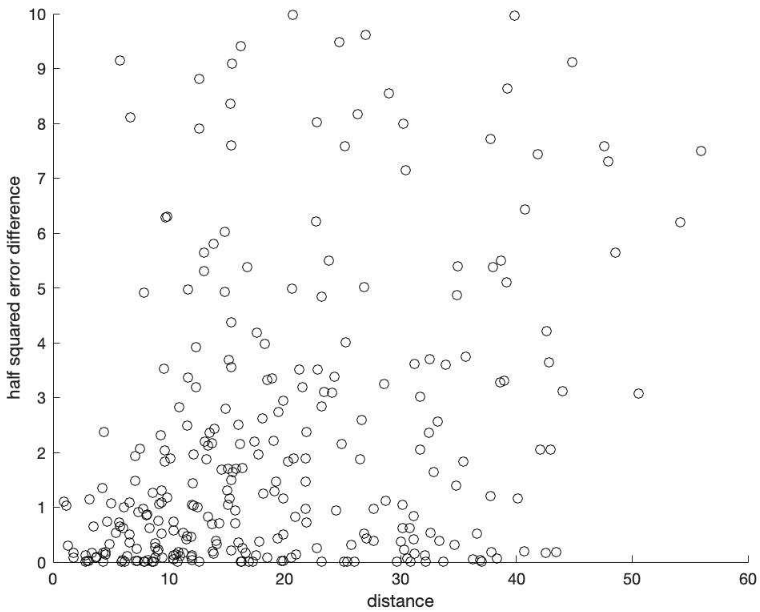

Figure 5: Half squared OLS residual differences v_{ij}, see (11), plotted against location distance differences r_{ij}.

In Figure 5 we see a tendency that the v_{ij}'s increase with r_{ij}.

To quantify, use a binning procedure: partition the interval of location differences into subintervals $I_k, k = 1, \ldots, K$, of equal length. Let

$$H_k = \{(i, j) : i < j, r_{ij} \in I_k\}$$

denote the set of distance pairs in I_k and put

$$\overline{v}_k = \frac{1}{|H_k|} \sum_{r_{ij} \in H_k} v_{ij} \;\; k = 1, \ldots, K \,. \tag{12}$$

The averages $\overline{v}_k$ are plotted as circles against subinterval midpoints $h_k, k = 1 \ldots, K$, in Figure 6.

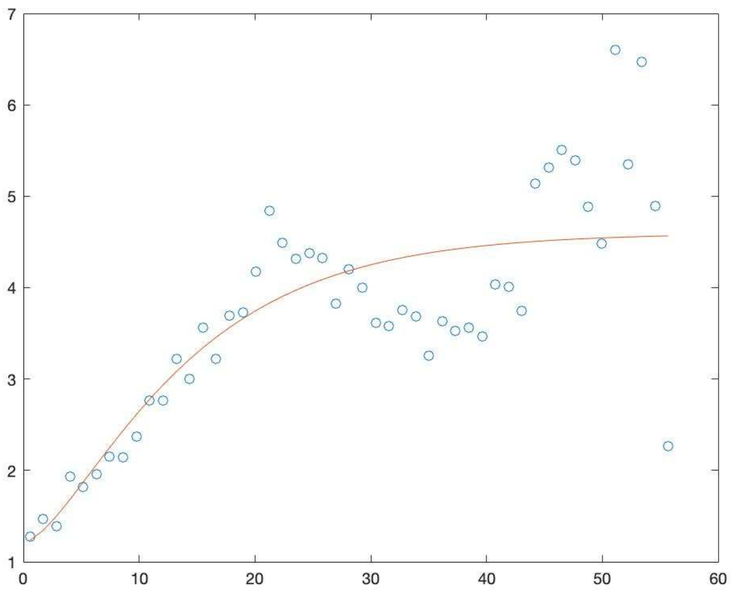

Figure 6: Average subinterval residual differences $\overline{v}_k, k = 1\ldots, K$, see (12), after binning plotted as circles against corresponding location difference subinterval midpoints. The figure also shows the fitted semivariogram for a Matérn covariance function.

To estimate the GLS covariance matrix Σ, fit with the semivariogram

$$\gamma(r;\Theta) = \sigma^2 \left[1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\theta}\right)^{\nu} K_\nu\left(\frac{r}{\theta}\right)\right] + \sigma_0^2,\ r = ||s-t||\,, \quad (13)$$

of a Matérn covariance function with a nugget effect σ_0^2 added.
Let Θ denoting the vector of all parameters.
Use weighted least squares and compute

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{1}{\operatorname{var}(\overline{v}_k)} (\overline{v}_k - \gamma(h_k;\Theta))^2 \approx \underset{\Theta}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{|H_k|}{(\gamma(h_k;\Theta))^2} (\overline{v}_k - \gamma(h_k;\Theta))^2\,, \quad (14)$$

Use the approximation in (14).

Parameter estimates are given in Tables 2 and 3, and the semivariogram in Figure 6.

The corresponding GLS regression surface estimate and residuals are shown in Figures 7 and 8.

Table 2: GLS Matérn semivariogram parameter estimates

Parameter	σ	ν	θ	σ_0	σ_ϵ
Estimate	1.839	1.004	9.381	1.087	1.102

Table 3: OLS (Ordinary Least Squares) and GLS (Generalized Least Square) parameter estimates for US continental summer temperatures 1997. Stars indicate that the corresponding parameter is significantly different from zero. The last row shows the residual standard deviation estimate $\hat{\sigma}_\epsilon$.

Explaining variable	OLS estimate	GLS estimate
Intercept	21.63***	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***
$\hat{\sigma}_\epsilon$	1.808	1.102

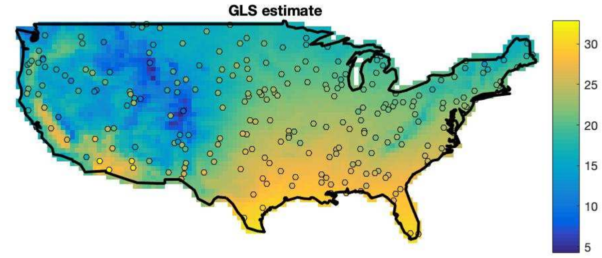

Figure 7: GLS regression temperature surface estimate

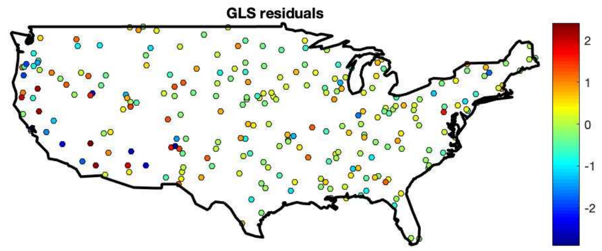

Figure 8: GLS regression temperature residuals

Maximum likelihood (ML) analysis

One step further: use maximum likelihood estimation for the model (1) assuming that X is a Gaussian spatial process with the Matérn covariance function. Let Θ be a parameter vector consisting of all parameters except the β-parameters. Then

$$Y \sim N(B\beta, \Sigma(\Theta)) \tag{15}$$

and the corresponding log-likelihood is

$$\ell(Y;\beta,\Theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma(\Theta)| - \frac{1}{2}(Y - B\beta)^T \Sigma(\Theta)^{-1}(Y - B\beta) \tag{16}$$

is optimized by numerical computations.

The parameter estimates are given in the fourth column in Table 4. We see here the successive improvements in going from OLS to GLS and then to ML. In Figure 9 we compare the semivariogram estimates from GLS and ML.

In the computation of the maximum likelihood estimates

$$(\hat{\beta}, \hat{\Theta}) = \underset{(\beta,\Theta)}{\operatorname{argmax}}\, \ell(Y;\beta,\Theta) \tag{17}$$

we use numerical optimization.

Profiling

Trick: for given Θ the log-likelihood $\ell(\beta, \Theta)$ is maximized for

$$\hat{\beta}(\Theta) = (B^T \Sigma(\Theta)^{-1} B)^{-1} B^T \Sigma(\Theta)^{-1} Y \,. \tag{18}$$

Then we compute by numerical computation

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}}\, \ell(Y; \hat{\beta}(\Theta), \Theta) \tag{19}$$

with $\hat{\beta}(\Theta)$ given (18). Note that this numerical computation is performed with optimization over $|\Theta|$ parameters while direct numerical optimization of (17) needs $|\beta| + |\Theta|$ parameters. After having obtained $\hat{\beta}(\Theta)$ we use (18) and find $\hat{\beta}$ as $\hat{\beta}(\hat{\Theta})$. This method of maximum likelihood computation with reduction of the number of parameters in the numerical optimization is often called *profiling* or *profile likelihood*.

REML

Problem with maximum likelihood estimation: parameter estimates may be biased particularly if the number p of parameters in β is large.

Think of the estimation of the error variance σ_ϵ^2 in the model (2). Let $e_i = Y_i - (B\hat{\beta})_i, i = 1, \ldots, n$, denote the residuals. Then the maximum likelihood estimate of the error variance is $(\sum_i e_i^2)/n$ which is biased compared to the unbiased estimate $(\sum_i e_i^2)/(n-p)$.

To avoid (or reduce) the bias problem one can use a method called *restricted maximum likelihood* (*REML*).

Basic idea in REML: for a model such as (15) with p parameters in β: consider $n-p$ linearly independent linear contrasts of type $a^T Y$ of the observations that have expectation zero for all β.

From the likelihood of these $n-p$ contrasts we can estimate the parameter Θ in a similar way as in profile likelihood. Then use (18) to estimate β.

Table 4: OLS (Ordinary Least Squares), GLS (Generalized Least Square) and ML parameter estimates for US continental summer temperatures 1997. Stars indicate that the corresponding parameter is significantly different from zero. The rows starting with $\hat{\sigma}$, $\hat{\nu}$, $\hat{\theta}$ and $\hat{\sigma}_0$ show estimates for the Matérn covariance function parameters. The last row shows the residual standard deviation estimate $\hat{\sigma}_\epsilon$.

Explaining variable	OLS estimate	GLS estimate	ML estimate
Intercept	21.63***	20.47***	19.80***
Longitude	-1.29***	-1.00	-0.53
Latitude	-2.70***	-2.68***	-2.64***
Altitude	-2.67***	-4.22***	-4.36***
East coast	-0.10	-0.01	0.02
West coast	-1.31***	-1.01***	-0.93***
$\hat{\sigma}$		1.839	3.048
$\hat{\nu}$		1.004	1.188
$\hat{\theta}$		9.381	10.204
$\hat{\sigma}_0$		1.087	0.811
$\hat{\sigma}_\epsilon$	1.808	1.102	0.848

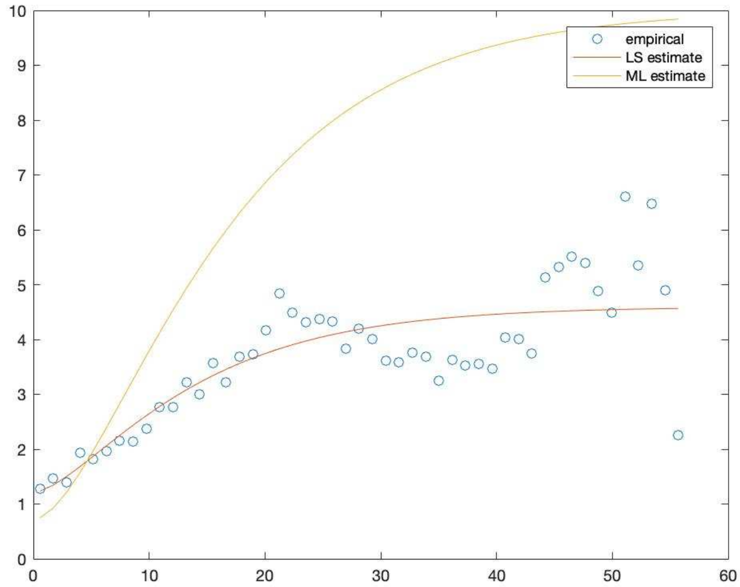

Figure 9: Average subinterval residual differences $\overline{v}_k, k = 1 \ldots, K$ after binning plotted as circles against corresponding location difference subinterval midpoints. The figure also shows fitted semivariograms for a Matérn covariance function from GLS and ML analyses.