Spatial statistics and image analysis. Lecture 2

Mats Rudemo

March 24, 2020

Background

Read LN Section 15.1, note the concept of *positive-definite matrix*.

Read LN Section 15.5, note the concepts of *covariance matrix* and the *d-dimensional normal distribution*.

Read LN Section 15.7, note the concepts of *stationary* and *normal* stochastic processes on the real line.

Spatial random processes

A spatial random process $X = (X_s, s \in S)$ may be characterized by its mean value function,

$$m_s = \mathbf{E}X_s \tag{1}$$

and its covariance function

$$C(s,t) = \mathbf{E}(X_s - m_s)(X_t - m_t). \tag{2}$$

A spatial random process $X = (X_s, s \in S)$ is Gaussian if the joint distribution of $(X_{s_1}, \ldots, X_{s_n})$ is an n-dimensional normal distribution for any choice of coordinates $s_1, \ldots, s_n$ in S.

A Gaussian random process is completely specified by its mean value and covariance functions.

Correlation functions

A spatial random process $(X_s, s \in S)$ is stationary if its distribution is invariant under a translation to $(X_{s+t}, s+t \in S)$ with $t \in \mathbb{R}^2$, *isotropic* if its distribution is also invariant under rotation of S. The covariance function can then be written on the form

$$C(s,t) = \sigma^2 \rho(|s-t|), \tag{3}$$

where $\rho = \rho(r), r \geq 0$ with $\rho(0) = 1$, is called the correlation function.

Examples of correlation functions ρ are: the *exponential correlation function*

$$\rho(r, \theta) = \exp(-r/\theta), \tag{4}$$

the *Gaussian correlation function*

$$\rho(r, \theta) = \exp(-(r/\theta)^2), \tag{5}$$

the *linear correlation function*

$$\rho(r, \theta) = (1 - r/\theta)1(r < \theta), \tag{6}$$

and the *spherical correlation function*

$$\rho(r, \theta) = (1 - \frac{2}{3}(r/\theta) + \frac{1}{2}(r/\theta)^3)1(r < \theta). \tag{7}$$

Simulation of a Gaussian spatial random process

We wish to simulate $(X_s, s \in S)$, where S has n_1 rows and n_2 columns, with mean matrix m and covariance matrix C.

Step 1 Reorder X into a $n{\times}1$ column vector $\tilde{X}$ with $n = n_1{\times}n_2$,

$$\tilde{X} = \mathcal{T}X \quad \text{and} \quad X = \mathcal{T}^{-1}\tilde{X}$$

Step 2 Let $\tilde{C}$ denote the $n \times n$ covariance matrix of $\tilde{X}$

Step 3 Let $\tilde{C} = R^T R$ be a Cholesky factorization of $\tilde{C}$

Step 4 Let Z be a $n{\times}1$ vector with i.i.d. $N(0,1)$ components, and put $\tilde{X} = R^T Z$

Step 5 Put $X = m + \mathcal{T}^{-1}\tilde{X}$

Matérn's correlation function

$$\rho(r) = \rho(r; \nu, \theta) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\theta}\right)^{\nu} K_{\nu}\left(\frac{r}{\theta}\right), \tag{8}$$

$\nu > 0$ and $\theta > 0$ are smoothness and scale parameters,
K_ν is a modified Bessel function of the second kind, which may be expressed as an integral

$$K_{\nu}(x) = \frac{2^{\nu}\Gamma(\nu + 1/2)}{\sqrt{\pi}x^{\nu}} \int_0^{\infty} \frac{\cos xt}{(t^2 + 1)^{\nu+1/2}}\, dt \tag{9}$$

Some special cases of Matérn's correlation function are given in Table 1 and Figure 1

Table 1: Special cases of Matérn's correlation function

Smoothness parameter ν	Matérn's correlation function $\rho(r)$ for scale parameter $\theta = 1$
$\nu = 1/2$	$\rho(r) = \exp(-r)$
$\nu = 3/2$	$\rho(r) = (1 + r)\exp(-r)$
$\nu = 5/2$	$\rho(r) = (1 + r + r^2/3)\exp(-r)$

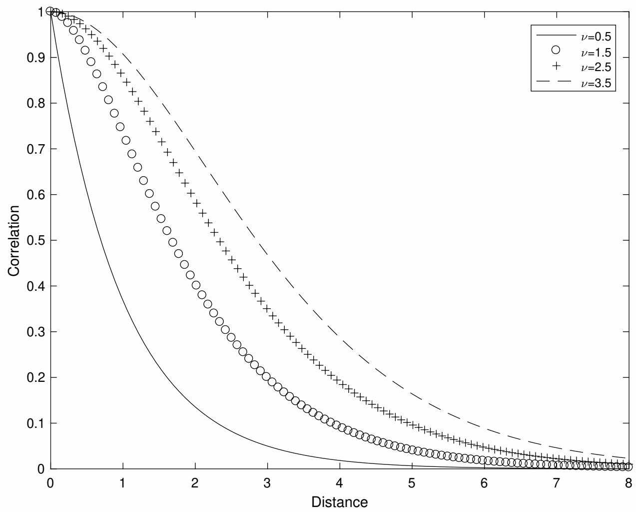

Figure 1: Four examples of Matérn correlation functions $\rho(r; \nu, \theta)$ from (8), plotted against distance r, with varying smoothness parameters ν and with constant scale parameter $\theta = 1$.

One can show that the practical correlation range for Matérn's correlation function is

$$d_{\mathrm{range}} \approx \theta\sqrt{(8\nu)}\,. \tag{10}$$

With $\theta = 1/(2\sqrt{\nu})$ one gets $d_{\mathrm{range}} \approx \sqrt{2}$ and this choice is used in Figure 2.

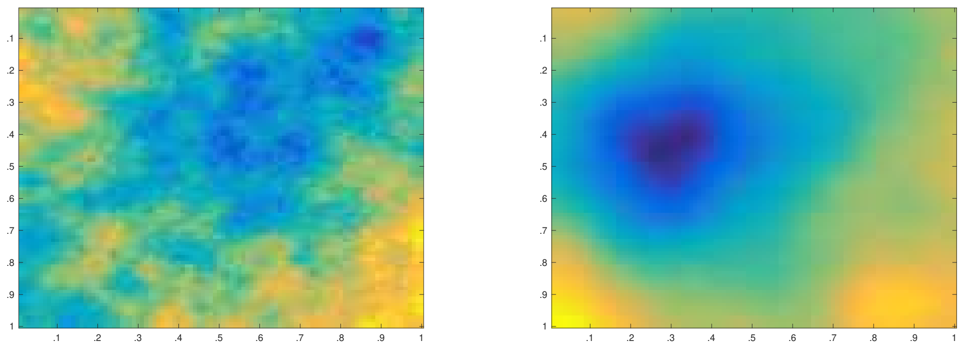

Figure 2: Two two-dimensional realizations with Matérn correlation functions $\rho(r;\nu,\theta)$ from (8) with $\nu = 0.5$ (left) and $\nu = 1.5$ (right). The simulation method with realizations obtained with the 5-step simulation scheme described above was used on a square two-dimensional set $S = [0,1] \times [0,1]$ with 100 pixels in both the horizontal and vertical directions. In both cases the scale parameter $\theta = 1/(2\sqrt{\nu})$.

Realizations from a Matérn process are continuous if $\nu > 0$

They are m times differentiable if $\nu > m$

Figure 3 shows one-dimensional realizations of Matérn processes with ν equal to $1/2$, $3/2$ and $5/2$, which are zero, one and two times differentiable.

With $\upsilon = 1/2$ we get in one dimension the *Ornstein-Uhlenbeck* process, a Markov process with correlation function

$$\rho(t) = \exp(-t)\,.$$

Similar to the Wiener process it is continuous but nowhere differentiable.

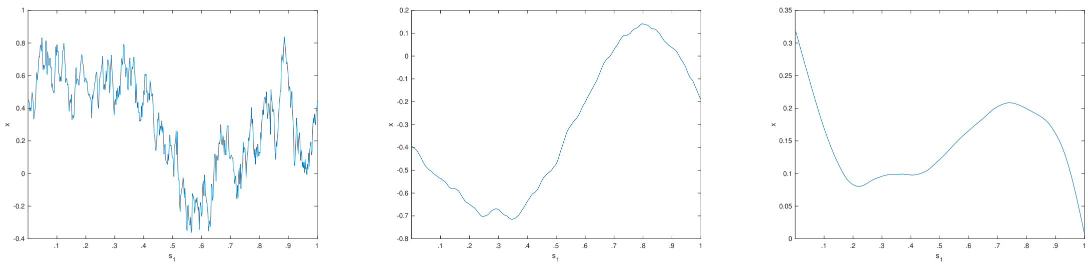

Figure 3: Three one-dimensional realizations with Matérn correlation functions $\rho(r; \nu, \theta)$ from (8) with $\nu = 0.5$ (left), $\nu = 1.5$ (center) and $\nu = 2.5$ (right). The processes are simulated on the interval $[0, 1]$ which is divided into 500 (one-dimensional) pixels.

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

$$Y_i = \sum_{k=1}^{K} B_k(s_i)\beta_k + X(s_i) + \epsilon_i \tag{11}$$

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

(i) How predict an observation at an unobserved location s_0 ?

(ii) How can we estimate parameters in the model (11)?

The prediction problem is sometimes called *kriging* after the South African mining engineer D. G. Krige.

Start with the prediction (kriging) problem with a fully specified model. Assume that we have column vectors X_1 and X_2 with a joint multivariate normal distribution

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right). \tag{12}$$

Then the conditional distribution of X_2 given X_1 is

$$X_2 | X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)\,,\, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \tag{13}$$

For kriging, let X_1 be our observations and X_2 the variables that we want to predict

Look now at problem (ii) with parameter estimation. As an example we shall look at mean summer time (June – August) temperatures in continental US recorded at 250 weather stations 1997. The temperatures and a number of possible explanatory variables can be obtained from
http://www.image.ucar.edu/GSP/Data/US.monthly.met/

Figure 4 shows the mean summer temperatures

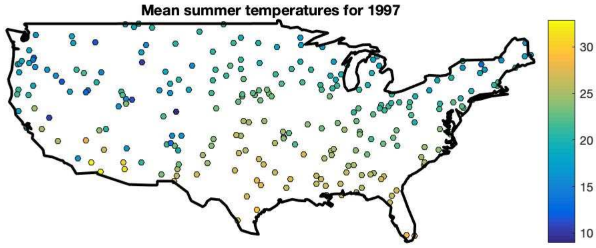

Figure 4: Mean summer temperatures for 1997 recorded at 250 weather stations in continental US.

The covariates we use are Longitude, Latitude, Altitude, East coast and West coast, see Figure 5.

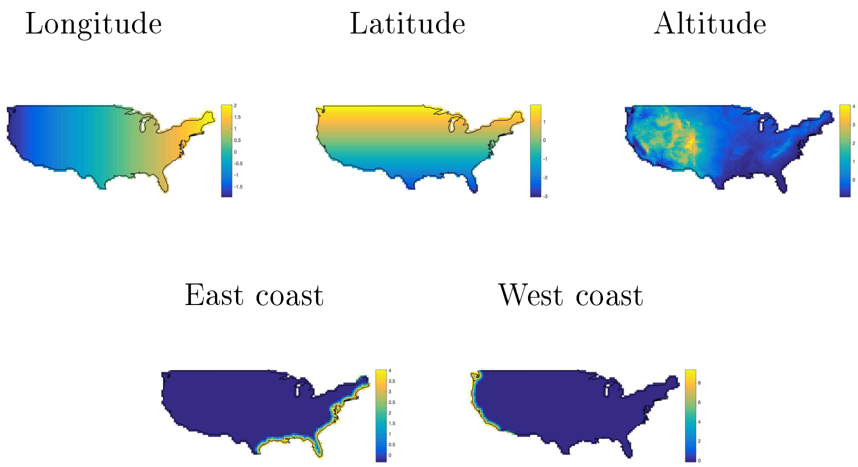

Figure 5: Five covariates used in the analysis of summer temperature in contintental US.

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

$$Y_i = \sum_{k=0}^{K} B_k(s_i)\beta_k + \epsilon_i \tag{14}$$

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

$$Y = B\beta + \epsilon\,, \ \ \epsilon \sim N(0, \sigma_\epsilon^2 I)\,. \tag{15}$$

Table 2: 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 2 shows the parameter estimates

$$\hat{\beta}_{\text{OLS}} = (B^T B)^{-1} B^T Y \tag{16}$$

of the OLS analysis. The residual degrees of freedom is 250-6= 244. From the column of t-values, corresponding to tests that the corresponding parameter is zero, we see that all the parameter estimates except one in Table 2 are highly significantly different from zero. The OLS regression surface estimate

$$\hat{Y}_{\text{OLS}} = B\hat{\beta}_{\text{OLS}} \tag{17}$$

of the temperature surface is shown in Figure 6 and the OLS regression residuals

$$\text{res}_{\text{OLS}} = Y - B\hat{\beta}_{\text{OLS}} \tag{18}$$

are shown in Figure 7. From Figure 7 we see that residuals close in location seem highly correlated, which indicates that the model could be improved.

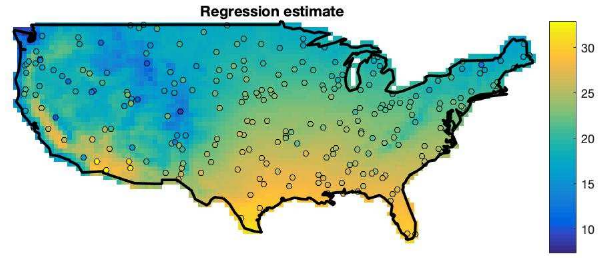

Figure 6: OLS regression temperature surface estimate

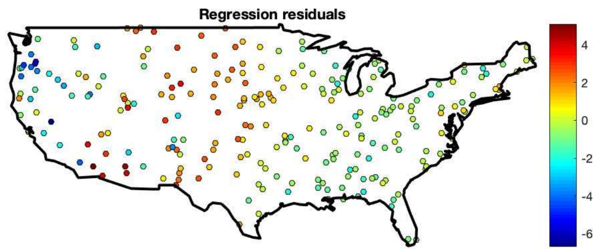

Figure 7: OLS regression temperature residuals

Current plan for lectures

1. Introduction and background
2. Gaussian random fields
3. Kriging and parameter estimation
4. Pattern recognition
5. Machine learning
6. Statistical image modelling
7. Point processes
8. Warping, Microarrays
9. Electrophoresis, Remote sensing
10. Diffusion
11. TEM images
12. Recapitulation
13. Seminars
14. Seminars