Spatial statistics and image analysis (TMS016/MSA301)

Spatial random processes

2021-03-24

Practical information

How to find a group?

- Discussion in Canvas can be used to find a group.
- As soon as you have a group, please, register it to Canvas in People/Groups (or inform me by email).
- If you want to be assigned to a group, please, send an email to me.

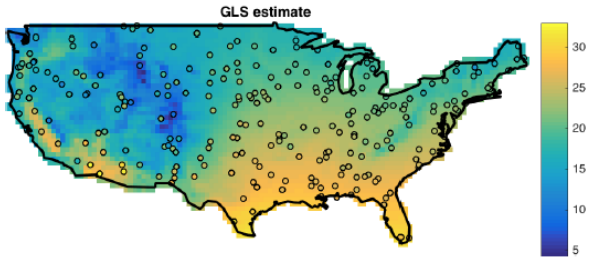
Programming language

- For project parts 1 and 2, Matlab is recommended since it has all the code you will need.
- For project part 3, you can use any program.

The book Image Analysis for the Biological Sciences by Glasbey and Horgan may disappear from the internet next month.

Spatial random processes

The variable of interest has some distribution, i.e. that the set of observations $\{x(s) : s \in S\}$ is a realization of a random process (random function, random field or stochastic process)



$$\{X(s) : s \in S\},$$

where S is a subset of $\mathbb{R}^2$ with positive 2-dimensional volume (area).

Observations are not an independent sample from the distribution due to spatial dependencies (observations that are located close to each other tend to be more similar than observations far away from each other).

Spatial random processes

A spatial random process can be characterized by its mean value

$$\mathbb{E}[X(s)] = m_s$$

and its covariance function

$$C(X(s), X(t)) = C(s,t) = \mathbb{E}[(X(s) - m_s)(X(t) - m_t)]$$

for all $s, t \in S$.

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

A Gaussian random process or Gaussian field is completely specified by its mean value and its covariance function.

Which functions are covariance functions?

Necessary and sufficient conditions for a function C to be a valid covariance function are that

- C is symmetric, i.e. $C(s,t) = C(t,s)$
- C is positive definite, i.e. it satisfies

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(s_i, s_j) \geq 0$$

for all n, $a_1, ..., a_n \in \mathbb{R}$ and $s_1, ..., s_n$.

Stationarity and isotropy

A spatial random process $(X(s), s \in S)$ is stationary if its distribution is invariant under a translation $t \in \mathbb{R}^2$, i.e. the distribution of $(X(s_1 + t), ..., X(s_n + t))$ does not depend on t for any n and $s_1, ..., s_n$.

It is isotropic if its distribution is invariant under rotations.

If the process is stationary and isotropic, the covariance function can be written as

$$C(s, t) = C(|s - t|) = \sigma^2 \rho(|s - t|),$$

where $|s - t|$ is the distance between s and t, and $\rho(r), \ r \geq 0$, with $\rho(0) = 1$, is called the correlation function.

Semivariogram

In geostatistics, a semivariogram is often used instead of the covariance function. It is defined as

$$\gamma(s,t) = \frac{1}{2}\,\mathrm{Var}(X(s) - X(t)).$$

(The function 2γ is called a variogram.)

For a stationary and isotropic process with variance σ^2 and correlation function ρ, we have

$$\gamma(s,t) = \sigma^2(1 - \rho(|s-t|)).$$

Nugget effect

In theory, $\gamma(s,s) = 0$ but in practise, this is not always the case. Instead,

$$\lim_{t \to s} \gamma(s,t) = c_0 > 0.$$

This is called a nugget effect (discontinuity at the origin). Then, we can define the semivariogram as

$$\gamma_0(s,t) = \begin{cases} c_0, & s = t \\ \gamma(s,t), & s \neq t \end{cases}.$$

The nugget effect can be explained by e.g. measurement error.

Some isotropic correlation functions

Often, a parametric model is chosen and fitted to the data, e.g.

1. Linear correlation function

$$\rho(r;\theta) = \begin{cases} 1 - \frac{r}{\theta}, & 0 \leq r < \theta \\ 0, & \text{otherwise} \end{cases}$$

2. Spherical correlation function

$$\rho(r;\theta) = \begin{cases} 1 - \frac{3}{2} \cdot \frac{r}{\theta} + \frac{1}{2} \left(\frac{r}{\theta}\right)^3, & 0 \leq r < \theta \\ 0, & \text{otherwise} \end{cases}$$

In both cases the parameter $\theta > 0$.

Some isotropic correlation functions

3. Exponential correlation function

$$\rho(r;\theta) = \exp\left(-\frac{r}{\theta}\right), \ r \geq 0$$

4. Gaussian correlation function

$$\rho(r;\theta) = \exp\left(-\left(\frac{r}{\theta}\right)^2\right), \ r \geq 0$$

In both cases the parameter $\theta > 0$.

Matérn correlation function

Matérn correlation function is a flexible and often used correlation function. It is defined as

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

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

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

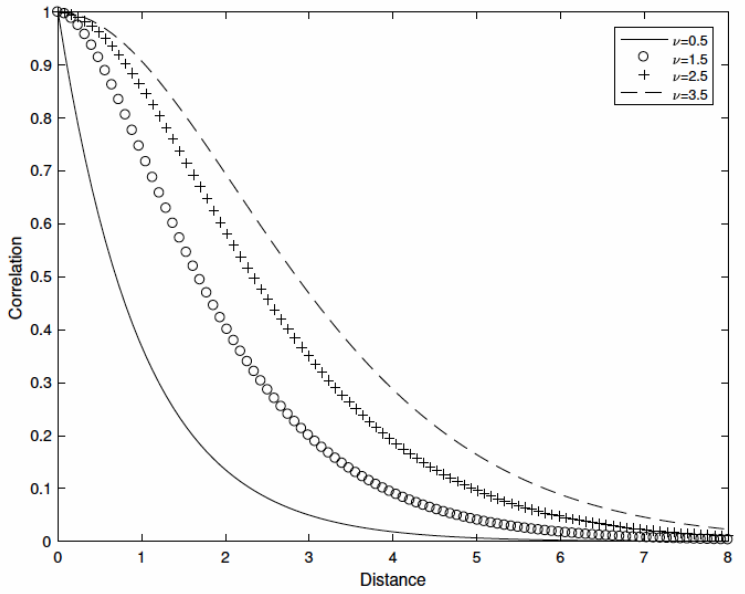
If $\nu = 0.5$ and $\theta = 1$, we get $\rho(r) = \exp(-r)$ (exponential), and if $\theta = 1/(2\sqrt{\nu})$ and $\nu \to \infty$, we get $\rho(r) = \exp(-r^2)$ (Gaussian).

Examples of Matérn correlation function

- $\theta = 1$ and $\nu = 0.5, 1.5, 2.5$ and 3.5
- The practical correlation range for Matérn correlation function is

$$d_{\text{range}} \approx \theta\sqrt{8\nu},$$

i.e. 2, 3.5, 4.5, and 5.3 for $\nu = 0.5, 1.5, 2.5$ and 3.5, respectively.



Simulation of a Gaussian field

The task is to simulate a realization from a Gaussian field $X(s),\ s \in S$, with mean matrix m and covariance matrix C.

Let S be a regular lattice with n_1 rows and n_2 columns and X a matrix of random variables that we want to simulate.

The values of X are put into a column vector $\tilde{X}$ of length $n = n_1 n_2$, e.g. by starting with the first column of X, then the second column of X and so on. We denote this transformation by $\mathcal{T}$ and its inverse by $\mathcal{T}^{-1}$. Therefore,

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

Simulation of a Gaussian spatial random process

A sequence of values in X can then be simulated using

$$\tilde{X} = R^T Z$$

where

- R is an upper triangular matrix with non-negative diagonal elements satisfyng $R^T R = C$
- Z is a vector of n independent standard normal random variables

and R can be computed by the so-called Cholesky decomposition.

Remarks

- Since C is positive definite and it can be factorized as above.
- R could be a lower triangular matrix as well. In this case, $\tilde{X} = RZ$ and $RR^T = C$

Simulation of a Gaussian spatial random process

The covariance matrix of $\tilde{X} = R^T Z$ is then

$$\mathbb{E}[R^T Z Z^T R] = R^T R = C$$

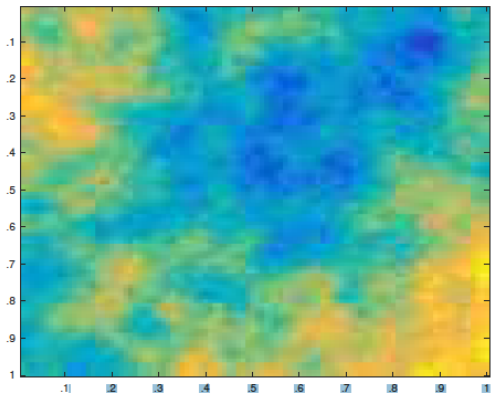
and therefore,

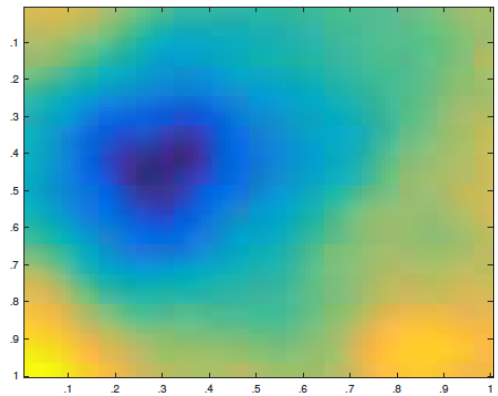
$$X = m + \mathcal{T}^{-1} \tilde{X}$$

has mean matrix m and covariance matrix C.

Properties of Matérn process

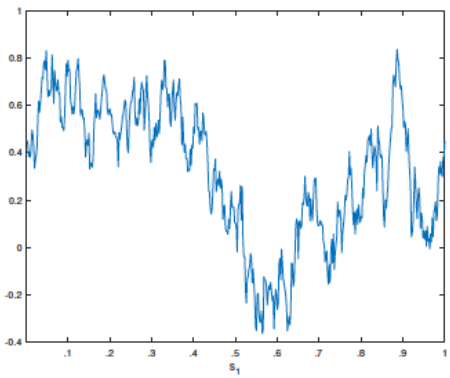
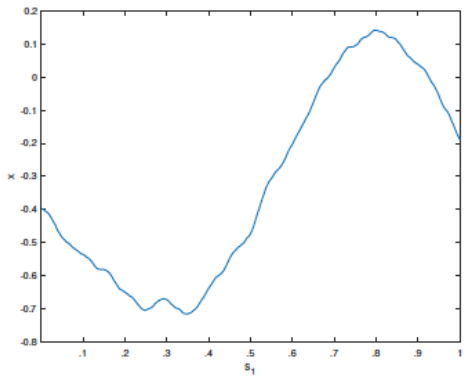
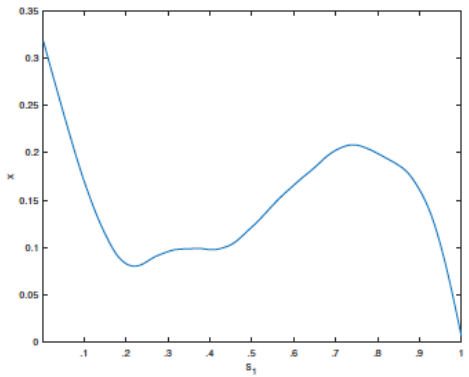
- Top: $\theta = 0.7,\ \ \nu = 0.5$
  Bottom: $\theta = 0.4,\ \nu = 1.5$
- Realizations of a Matérn process with $\nu > 0$ are continuous.
- Realizations of a Matérn process with $\nu > 0$ are m times differentiable if and only if $\nu > m$.
- The upper realization ($\nu = 0.5$) is continuous but not differentiable and the lower realization ($\nu = 1.5$) is continuous and differentiable once.





1D illustration of Matérn process

1D realizations of Matérn process with $\nu = 0.5, 1.5$ and 2.5 (and $\theta = 1/(2\sqrt{\nu})$) which are 0, 1, and 2 times differentiable, respectively.

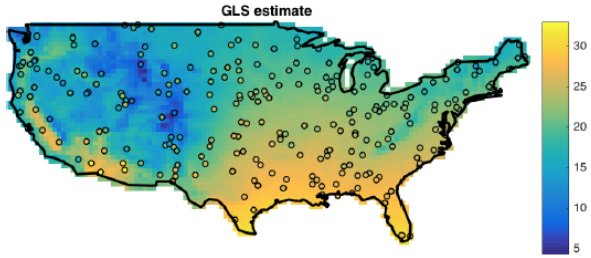




If $\nu = 0.5$, the process is called the Ornstein-Uhlenbeck process which is continuous but nowhere differentiable.

Statistical models including random fields

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^2_ϵ and independent of X

Questions

- How can we predict a measurement at an unorserved location s_0? (kriging)
- How can we estimate the parameters?

Prediction (kriging) with a fully specified model

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

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

we have that the conditional distribution of X_2 given X_1 is

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

If X_2 represents a random field at some unobserved locations and X_1 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 prediction

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 predictor.