Lecture 9: Introduction to Kriging

Math 586

Beginning remarks

Kriging is a commonly used method of interpolation (prediction) for spatial data. The data are a set of observations of some variable(s) of interest, with some spatial correlation present.

Usually, the result of kriging is the expected value ("kriging mean") and variance ("kriging variance") computed for every point within a region. Practically, this is done on a fine enough grid.

Illustration: suppose we observe some variable Z along 1-dim space (X). There are 5 measurements made. We might ask ourselves, knowing the probabilistic behavior of the random field being observed, what are possible trajectories (realizations) of the random field, that agree with the data? This is answered by *conditional simulation*. In Fig. 1, you see two sets of 5 conditional simula-

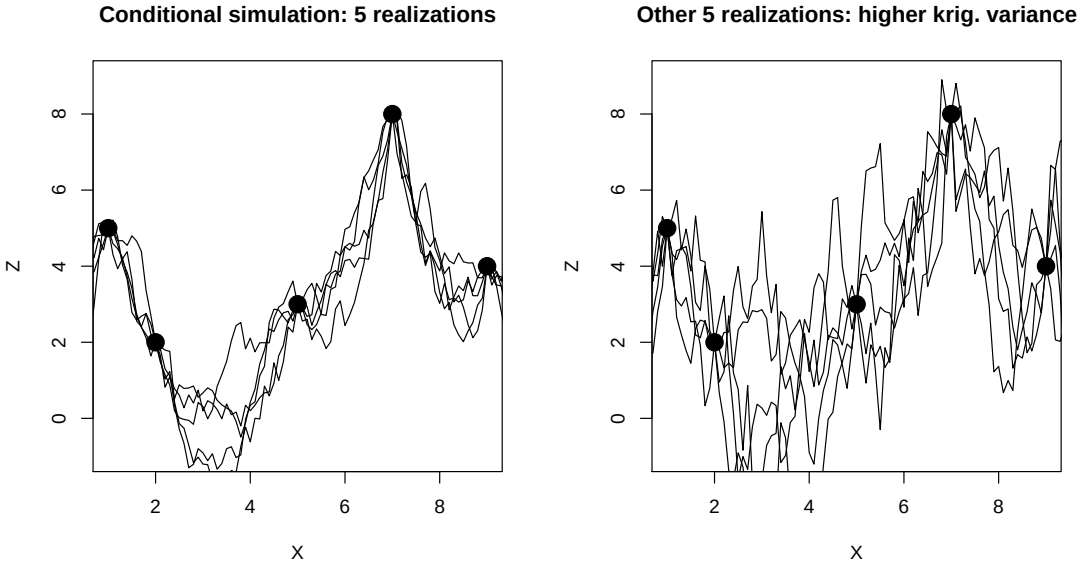


Figure 1: Conditional simulations, data given by dots

tions for the same data. The one on the left is for one value of σ^2= variance

of the random field, the one on the right is for another, higher, value. All the realizations are faithful to the data, but also faithful to the statistical model for the random field (i.e. mean and variogram) that we selected.

Kriging mean for every location can be thought of as the average of the whole *ensemble* of possible realizations, conditioned on data. Kriging variance is the variance of that ensemble. (The ensemble is of course infinite, we only show 5 of its representatives.)

You may see that the trajectories tend to diverge away from the observed data, that is, the kriging variance increases. Also, for the random field with higher variance (the one on the right in Fig. 1)

We will observe similar qualitative behavior frequently in the future.

Simple kriging (SK)

Let us observe some stationary (WSS) random field $V(\mathbf{x})$ at some points $\mathbf{x}_j$, $j = 1, ..., n$. First, assume that the mean m and covariance function $C(\cdot)$ of this process are known. The case of prediction with the known mean is often called *simple kriging*.

In order to better understand what happens and to devise an extendable approach, let's attack the question by directly trying to minimize MSE.

We will seek an estimate $\hat{V}_0$ of the value of V at the point $\mathbf{x}_0$. For simplicity, denote $V_j := V(\mathbf{x}_j)$. Also, denote $C(i,j) = Cov(V_i, V_j)$.

We may assume that the mean $m = 0$. Otherwise, subtract m from all of the V_j values, estimate V_0, then add the mean back.

Search for the estimate of the form

$$\hat{V}_0 = \sum_{j=1}^{n} \lambda_j V_j,$$

under the assumption of 0 mean, it is automatically unbiased. (Why?)

We will find the *kriging weights* λ_j that minimize MSE:

$$\texttt{MSE} = \mathbb{E}\left[\left(V_0 - \sum \lambda_j V_j\right)^2\right] = Var\left(V_0 - \sum \lambda_j V_j\right) = \qquad \text{(Why?)}$$

$$= Var(V_0) - 2\sum \lambda_j Cov(V_0, V_j) + \sum\sum \lambda_j \lambda_i Cov(V_i, V_j)$$

That is,

$$\texttt{MSE} = C(0,0) - 2\sum \lambda_j C(0,j) + \sum\sum \lambda_j \lambda_i C(i,j) \qquad (1)$$

where $C(i,j) = C_V(\mathbf{x}_i - \mathbf{x}_j)$ might be interpreted as the elements of covariance matrix $\mathbf{C}$. Note that the values $C(0,j)$ depend on the location $\mathbf{x}_0$.

To minimize MSE, we take the derivatives with respect to λ_k and equate to 0:

$$\frac{\partial \texttt{MSE}}{\partial \lambda_k} = -2C(0,k) + 2\lambda_k C(k,k) + 2\sum_{j \neq k} \lambda_j C(k,j) = 0, \quad k = 1, ..., n$$

That is, solve the system of equations[1]

$$\sum_{j=1}^{n} \lambda_j C(k,j) = C(0,k), \quad k = 1, ..., n \tag{2}$$

In matrix form, the above equations look like

$$\mathbf{C}\boldsymbol{\lambda} = \mathbf{b} \qquad \text{same as} \qquad \boldsymbol{\lambda} = \mathbf{C}^{-1}\mathbf{b} \tag{3}$$

If we denote the covariance matrix of the vector $(V_0, V_1, ..., V_n)'$ as $\boldsymbol{\Sigma}$, then

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_0^2 & \mathbf{b}' \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \quad \text{and} \quad \texttt{MSE=}\ \sigma^2_{\texttt{SK}} = C(0,0) - 2\boldsymbol{\lambda}'\mathbf{b} + \boldsymbol{\lambda}'\mathbf{C}\boldsymbol{\lambda} = C(0,0) - \boldsymbol{\lambda}'\mathbf{b}$$

Now compare the equation (3) with the formula for conditional mean in Lecture 6, p. 4. It turns out that the above minimization argument is directly related to the BLUE theory for multivariate Normal!

To avoid ambiguity, we will sometimes denote the optimal kriging weights $\lambda^{\texttt{SK}}$. From the equations (1) and (2), the optimal kriging variance is

$$\sigma^2_{\texttt{SK}} = C(0,0) - \sum \lambda_j^{\texttt{SK}} C(0,j) \tag{4}$$

Compare this to the conditional variance formula from Lecture 6.

What happens when you "predict" at one of the existing points $\mathbf{x}_j$? It is clear that the best choice is to just take V_j, in which case the kriging variance is 0! If we use the covariance function such that $C(\mathbf{h}) \to C(0)\sigma^2$ as $|\mathbf{h}| \to 0$, that is, dealing with a continuous random field (no nugget!), then we should obtain

$$\lim_{\mathbf{x}_0 \to \mathbf{x}_j} \hat{V}(\mathbf{x}_0) = V_j \qquad \text{and} \qquad \sigma^2_{\texttt{SK}} \to 0$$

[1]Note that we really did not use the assumption of WSS stationarity, since we allowed $C(i,j)$ to vary with location $\mathbf{x}_i$. We don't have to assume a constant mean m, either, and can replace it by varying values m_i, as long as these are known.

A simple case:

Suppose that $C(i,j) = 0,\ i \neq j$, only $C(0,j) \neq 0,\ j = 1, ..., n$. In this case, the kriging equations (2) are solved by

$$\lambda_j = \frac{C(0,j)}{C(j,j)} = \rho_{0,j}\frac{\sigma_0}{\sigma_j},$$

where $\rho_{0,j}$ is the correlation coefficient. We obtain the prediction

$$\frac{\hat{V}_0}{\sigma_0} = \sum \rho_{0,j}\frac{V_j}{\sigma_j}$$

That is, the weights reflect the strength of correlation. This reminds us of simple linear regression.

Also,

$$\sigma^2_{\mathtt{SK}} = \sigma_0^2(1 - \sum \rho^2_{0,j})$$

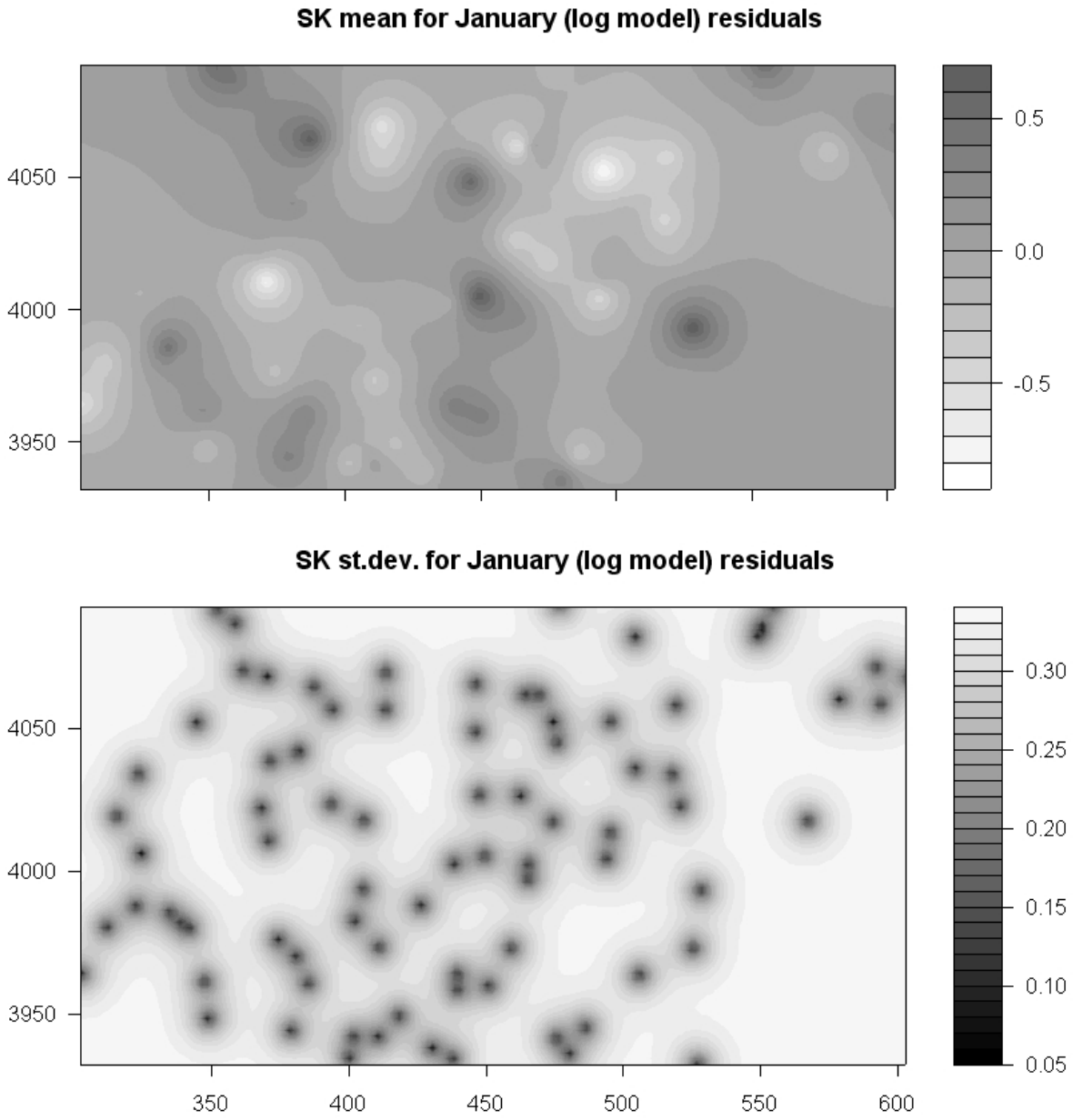
where we recognize the quantity $\sum \rho^2_{0,j}$ as coefficient of determination R^2! (Question: why $\sum \rho^2_{0,j} \leq 1$?)

Summary:

- Kriging is an *exact* interpolator (it preserves the observations) when there's no nugget effect.
- Kriging weights λ_j depend only on locations and covariance function C, not on data.
- Kriging variance also depends on locations and C only.
- Kriging is a smooth predictor (if two points are close, then their kriging estimates are close).
- In MVN case, SK is *the* optimal predictor in the mean square sense (follows from MVN theory).
- Computationally, we only need to find the matrix $\mathbf{C}^{-1}$ once, then, as we vary the location, we will only vary vector $\mathbf{b}$ in (3)

Example

Northern NM precipitation, January log-model residuals



The covariance function used in kriging is $C(h) = 0.1144e^{-h/10.4451}$, with the practical range of about 30. One look at the kriging st. dev. map tells us that we cannot reliably predict the precipitation (at least for January) in-between the stations.

Ordinary kriging (OK)

The next issue will be estimating the mean m, which we assumed known in simple kriging.
Option 1: first, estimate mean from the observations, then apply SK.
Option 2: do both at the same time (Ordinary kriging).

Option 1: We could try and estimate m by simply averaging the observations, but this does not take into account the correlations between V_j. For example, if two locations are close together, and there is spatial continuity, then we have effectively one observation instead of two.
This will require a *Generalized least square* approach, discussed later.

Option 2: To keep the estimate unbiased, we need

$$m = \mathbb{E}[V_0] = \sum \lambda_j \mathbb{E}[V_j] = \sum \lambda_j m$$

Thus, we need $\sum \lambda_j = 1$.

MSE is still expressed by (1), and we need to minimize it.
We will apply *Lagrange multiplier* method of constrained optimization: minimize the function

$$F(\boldsymbol{\lambda}, \mu) = \mathtt{MSE} - 2\mu(\sum \lambda_j - 1)$$

where μ is the Lagrange multiplier. Taking partials with respect to λ_k and μ and equating them to 0, we will get a system of equations

$$\begin{cases} \sum \lambda_j^{\mathtt{OK}} C(k,j) - \mu = C(0,k), & k = 1, ..., n \\ \sum \lambda_j^{\mathtt{OK}} = 1 \end{cases}$$

In matrix form:

$$\begin{bmatrix} C(1,1) & \dots & C(1,n) & -1 \\ C(2,1) & \dots & C(2,n) & -1 \\ \dots & \dots & \dots & \dots \\ C(n,1) & \dots & C(n,n) & -1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C(0,1) \\ C(0,2) \\ \dots \\ C(0,n) \\ 1 \end{pmatrix}$$

The kriging variance is, similarly to (4),

$$\sigma^2_{\mathtt{OK}} = C(0,0) - \sum \lambda_j^{\mathtt{OK}} C(0,j) + \mu$$

Overall, $\sigma^2_{\mathtt{OK}}$ is slightly higher than $\sigma^2_{\mathtt{SK}}$ because of higher uncertainty associated with estimating m.

Example – continued

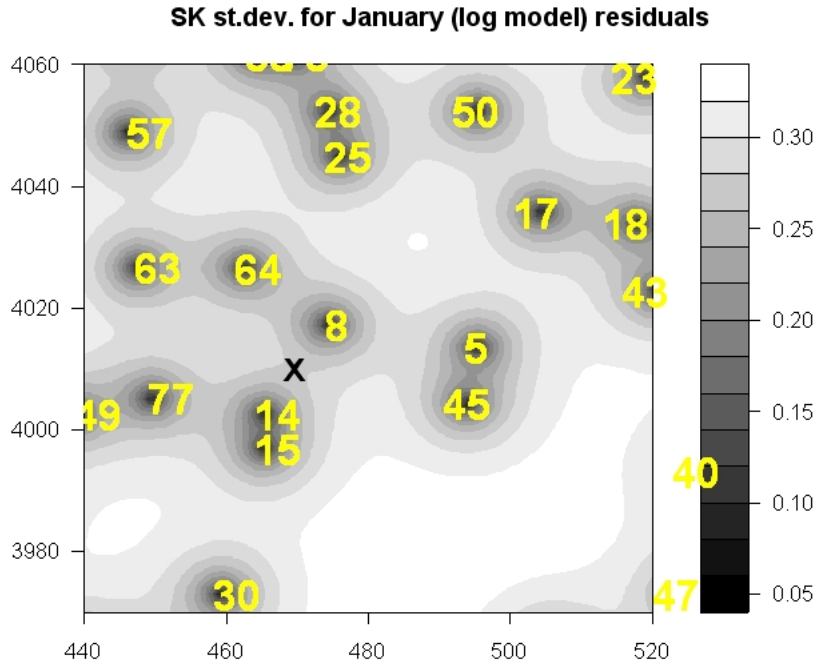
Let's zoom in on a portion of NM precipitation map: the locations marked are stations close to the point $x_0 = 470, y_0 = 4010$, marked by $\times$.

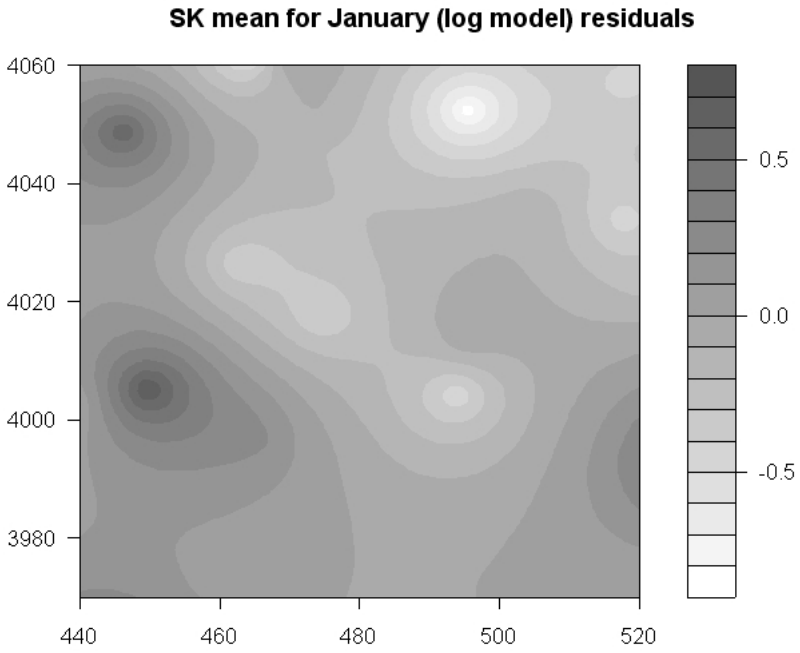
Comparing kriging st.dev. at that point, $\sigma_{\texttt{SK}} = 0.2778$, $\sigma_{\texttt{OK}} = 0.2779$, and maximum possible is $\sigma = \sqrt{C(0,0)} = 0.3382$.

The kriging weights λ_j were sorted according to their absolute values and the first 10 are:

```
    lambda_SK  lambda_OK  point no.
 [1,]  0.3551   0.3569      8
 [2,]  0.3443   0.3455     14
 [3,]  0.0527   0.0544     64
 [4,]  0.0283   0.0306     45
 [5,]  0.0279   0.0295     77
 [6,]  0.0117   0.0135     15
 [7,]  0.0103   0.0126     63
 [8,]  0.0090   0.0109      5
 [9,] -0.0036  -0.0020     25
[10,] -0.0025  -0.0005     49
```

Generally, closer points get higher weights. Notice, though, that 15 is masked by the presence of 14, and some stations get negative weights!




SK mean for January (log model) residuals
4060
4040
4020
4000
3980
440
460
480
500
520
0.5
0.0
-0.5