Geostatistics and Analysis of Spatial Data

Allan A. Nielsen

Abstract—**This note deals with geostatistical measures for spatial correlation, namely the auto-covariance function and the semi-variogram, as well as deterministic and geostatistical methods for spatial interpolation, namely inverse distance weighting and kriging. Some semi-variogram models are mentioned, specifically the spherical, the exponential and the Gaussian models. Equations to carry out simple and ordinary kriging are deduced. Other types of kriging are mentioned, and references to international literature, Internet addresses and state-of-the-art software in the field are given. A very simple example to illustrate the computations and a more realistic example with height data from an area near Slagelse, Denmark, are given. Finally, a series of attractive characteristics of kriging are mentioned, and a simple sampling strategic consideration is given based on the dependence of the kriging variance of distance and direction to the nearest observations.**

I. Introduction

OFTEN we need to be able to integrate point attribute information with vector and raster data which we may already have stored in a Geographical Information System (GIS). This can be done by linking the point information to a geographical coordinate in the data base. If we have lots of point data, a tempting alternative will be to generate an interpolated map so that from our point data we calculate raster data which can be analysed along with other sources of raster data.

This note deals with geostatistical methods for description of spatial correlation between point measurements as well as deterministic and geostatical methods for spatial interpolation.

The basic idea in geostatistics consists of considering observed values of geochemical, geophysical or other natural variables as realisations of a stochastic process in the 2-D plane or in 3-D space. For each position $\boldsymbol{r}$ in a domain $\mathcal{D}$ which is a part of Euclidian space, a measureable quantity $z(\boldsymbol{r})$ termed a *regionalised variable* exists. $z(\boldsymbol{r})$ is considered as a realisation of a *stochastic variable* $Z(\boldsymbol{r})$. The set of stochastic variables $\{Z(\boldsymbol{r}) \,|\, \boldsymbol{r} \in \mathcal{D}\}$ is termed a *stochastic function*. $Z(\boldsymbol{r})$ has mean value or expectation value $\mathrm{E}\{Z(\boldsymbol{r})\} = \mu(\boldsymbol{r})$ and auto-covariance function $\mathrm{Cov}\{Z(\boldsymbol{r}), Z(\boldsymbol{r}+\boldsymbol{h})\} = C(\boldsymbol{r},\boldsymbol{h})$, where $\boldsymbol{h}$ is termed the displacement vector. If $\mu(\boldsymbol{r})$ is constant over $\mathcal{D}$, i.e., $\mu(\boldsymbol{r}) = \mu$, Z is said to be first order stationary. If also $C(\boldsymbol{r},\boldsymbol{h})$ is constant over $\mathcal{D}$, i.e., $C(\boldsymbol{r},\boldsymbol{h}) = C(\boldsymbol{h})$, Z is said to be second order stationary.

This statistical view is inspired by work carried out by Georges Matheron in 1962-1963. It is described in for example [1], [2]. [3] gives a good practical and data analytically oriented introduction to geostatistics. [4] is a chapter in a collection of articles which describe many different techniques and their application within the geosciences. [5] deals with geostatistics and other relevant subjects in the context of analysis of spatial data. Geostatistical expositions in a GIS context can be found in [6], [7]. [8] deals with multivariate geostatistics, i.e., studies of the joint spatial co-variation of more variables. *The International Association for Mathematical Geology (IAMG)* publishes i.a. the periodical *Mathematical Geology* where many results on geostatistical research are published. *State-of-the-art software* may be found in *GSLIB*, [9], and *Variowin*, [10]. Other easily obtainable softwares are *Geo-EAS* and *Geostatistical Toolbox*. All these packages can be found at http://www-sst.unil.ch/research/variowin/ (or via a search engine). Also commercial geostatistical software exists.

This note is inspired by [11], see also [12].

Allan Aasbjerg Nielsen, Danish National Space Center, Technical University of Denmark, Building 321, DK-2800 Lyngby, Telephone +45 4525 3425, Telefax +45 4588 1397, E-mail: aa@imm.dtu.dk, Internet http://www.imm.dtu.dk/∼aa/.

II. Spatial Correlation

This section mentions methods for description of similarity between measurements of natural variables in the 2-D plane or in 3-D space. Specifically the auto-covariance function and the semi-variogram are introduced. Also a relation between the two is given.

A. The Semi-Variogram

Consider two scalar quantities $z(\boldsymbol{r})$ and $z(\boldsymbol{r}+\boldsymbol{h})$ measured at two points in the plane or in space $\boldsymbol{r}$ og $\boldsymbol{r}+\boldsymbol{h}$ separated by the displacement vector $\boldsymbol{h}$. We consider z as a realisation of a stochastic variable Z. The variability may be described b.m.o. the *auto-covariance function* (assuming or enforcing first order stationarity, i.e., the mean value is position independent)

$$C(\boldsymbol{r},\boldsymbol{h}) \quad = \quad \mathrm{E}\{[Z(\boldsymbol{r})-\mu][Z(\boldsymbol{r}+\boldsymbol{h})-\mu]\}.$$

The *variogram*, 2γ, is defined as

$$2\gamma(\boldsymbol{r},\boldsymbol{h}) \quad = \quad \mathrm{E}\{[Z(\boldsymbol{r})-Z(\boldsymbol{r}+\boldsymbol{h})]^2\},$$

which is a measure for the average, squared difference between measurement values as a function of distance and direction between observations. In general the variogram will depend on on the displacement vector $\boldsymbol{h}$ as well as on the position vector $\boldsymbol{r}$. The *intrinsic hypothesis* of geostatistics says that the *semi-variogram*, γ, is independent of the position vector and that it depends only on the displacement vector, i.e.,

$$\gamma(\boldsymbol{r},\boldsymbol{h}) \quad = \quad \gamma(\boldsymbol{h}).$$

If $Z(\boldsymbol{r})$ is second order stationary (i.e., its auto-covariance function is position independent), the intrinsic hypothesis is valid whereas the opposite is not necessarily true.

If we assume or enforce second order stationarity the following relation between the auto-covariance function and the semi-variogram is valid

$$\gamma(\boldsymbol{h}) \quad = \quad C(\boldsymbol{0}) - C(\boldsymbol{h}). \tag{1}$$

Note, that $C(\boldsymbol{0}) = \sigma^2$, the variance of the stochastic variable.

Given a set of point measurements the semi-variogram may be calculated b.m.o. the following estimator, which calculates (half) the mean value of the squared differences between all pairs of measurements $z(\boldsymbol{r}_k)$ and $z(\boldsymbol{r}_k+\boldsymbol{h})$ separated by the displacement vector $\boldsymbol{h}$

$$\hat{\gamma}(\boldsymbol{h}) \quad = \quad \frac{1}{2N(\boldsymbol{h})}\sum_{k=1}^{N(h)}[z(\boldsymbol{r}_k)-z(\boldsymbol{r}_k+\boldsymbol{h})]^2.$$

$N(\boldsymbol{h})$ is the number of point pairs separated by $\boldsymbol{h}$. $\hat{\gamma}$ is termed the *experimental semi-variogram*. Often we calculate mean values of $\hat{\gamma}$ over intervals $\boldsymbol{h} \pm \Delta\boldsymbol{h}$ for both length (magnitude) and direction (argument) of $\boldsymbol{h}$. Mean values for the magnitude of $\boldsymbol{h}$ ($h \pm \Delta h$) are calculated in order to get a sufficiently high $N(\boldsymbol{h})$ to obtain a low estimation variance for the semi-variogram value. Mean values over intervals af the argument of

$\boldsymbol{h}$ are calculated to check for possible anisotropy. Anisotropy refers to the characteristic that the auto-covariance function and the semi-variogram do not behave similarly for all directions of the displacement vector between observations. This possible anisotropy may also be checked by calculating 2-D semi-variograms also known as variogram maps, [3], [13], [11], [10], [9].

B. Semi-Variogram Models

In order to be able to define its characteristics we parameterise the semi-variogram b.m.o. different semi-variogram models. An often used model, γ^*, is the *spherical model* (here we assume or impose isotropy, i.e., the semi-variogram depends only on distance and not on direction between observations, and we denote by h the magnitude of $\boldsymbol{h}$)

$$\gamma^*(h) = \begin{cases} 0 & h = 0 \\ C_0 + C_1 \left[\frac{3}{2}\frac{h}{R} - \frac{1}{2}\frac{h^3}{R^3} \right] & 0 < h < R \\ C_0 + C_1 & h \geq R \,, \end{cases}$$

where C_0 is the so-called *nugget effect* and R is termed the *range of influence* or just the *range*; $C_0/(C_0+C_1)$ is the relative *nugget effect* and C_0+C_1 is termed the *sill* ($= \sigma^2$). The parameters C_0 and C_1 are not to be confused with the auto-covariance function $C(\boldsymbol{h})$. The *nugget effect* is a discontinuity in the semi-variogram for $h = 0$, which is due to both measurement uncertainties and micro variability that cannot be revealed at the scale of sampling. The *range of influence* is the distance where covariation between samples ceases to exist; measurements taken further apart are uncorrelated.

Two other models often used are the exponential model (see Figure 4)

$$\gamma^*(h) = \begin{cases} 0 & h = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{3h}{R}\right)\right] & h > 0 \end{cases}$$

and the Gaussian model (see Figure 5)

$$\gamma^*(h) = \begin{cases} 0 & h = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{3h^2}{R^2}\right)\right] & h > 0 \,. \end{cases}$$

These latter two models never reach but approach the *sill* asymptotically. Due to its horisontal tangent for $h \rightarrow 0$ the Gaussian model is good for describing very continuous phenomena.

Also other semi-variogram models such as linear and power functions are some times applied. To allow for so-called *nested structures* where the semi-variogram has different structures depending on the magnitude and possibly the direction of the displacement vector between observations, combinations of models may be useful.

The model parameters may be estimated b.m.o. iterative, non-linear least squares methods. These minimise the squared differences between the experimental semi-variogram and the model considered as a function of the vector of parameters $\boldsymbol{\theta}$, here $\boldsymbol{\theta} = [C_0\ C_1\ R]^T$

$$\min_{\boldsymbol{\theta}} ||\hat{\gamma}(\boldsymbol{h}) - \gamma^*(\boldsymbol{\theta}, \boldsymbol{h})||^2.$$

For examples on an experimental semi-variogram and different models, see Figures 4 and 5.

Note, that $C(\mathbf{0})$ is the auto-covariance function for displacement vector $\boldsymbol{h} = \mathbf{0}$, and that C_0 is a parameter in the semi-variogram model.

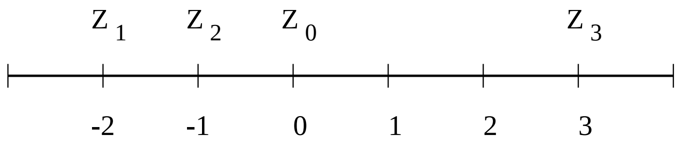

Fig. 1. Simple example with three observations.

C. Examples

To illustrate the calculations Figure 1 shows a very simple example with three observations, $z_1 = 1, z_2 = 3$ og $z_3 = 2$ with (1-D) coordinates –2, –1 og 3. The semi-variogram $\hat{\gamma}$ with $\Delta h = 1.5$ is calculated like this (*lags* are distance groups defined by $h \pm \Delta h$)

lag	h	N	$\hat{\gamma}$
0	$0 < h \leq 3$	1	$1/2(1-3)^2 = 2$
1	$3 < h \leq 6$	2	$1/4((1-2)^2 + (3-2)^2)$
			$= 1/2$

As another more realistic example Figure 2 shows a map with sample sites. Each circle is centered on a sample point and its radius is proportional to the quantity measured, in this case the height above the ground water in a 10 km $\times$ 10 km area near Slagelse, Denmark. Figure 3 shows a histogram for these data.

Figure 4 shows all possible pairwise squared differences as a function of distance between observations for the height data (assuming isotropy). Also an exponential variogram model estimated directly on this point cloud is shown. The *nugget effect* is 0 m^2, the effective *range* is 3,840 m and the *sill* is 840 m^2 (corresponding to 420 m^2 for the semi-variogram model).

Figure 5 shows the corresponding experimental semi-variogram. Δh is here 100 m and again we assume isotropy. Traditionally the first lag interval is half the size of the remaining lags, here 50 m. The experimental semi-variogram indicates that a Gaussian model may perform better than the exponential model in this case. Therefore a Gaussian model based on the experimental semi-variogram is shown also. The *nugget effect* is 18 m^2, the *range* is 1,890 m and the *sill* is 364 m^2.

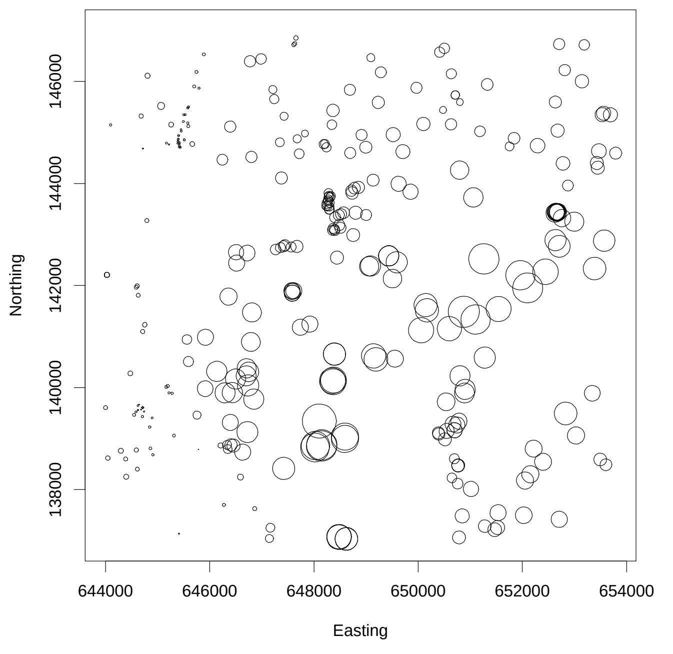

Fig. 2. Sample sites, each circle is centered on a sample point, radius is proportional to the quantity measured, in this case the height above the ground water in a 10 km $\times$ 10 km area near Slagelse, Denmark.

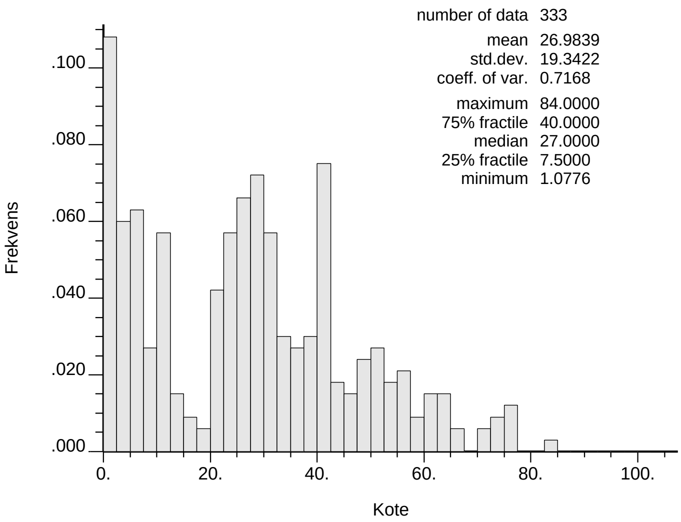

Fig. 3. Simple statistics and histogram for height data near Slagelse, Denmark.

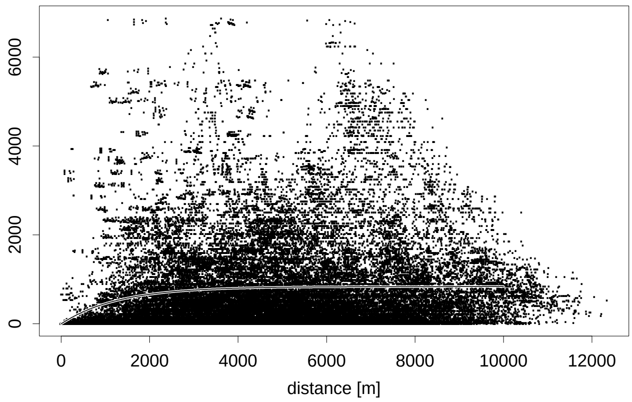

Fig. 4. All possible pairwise squared differences as a function of the magnitude of the displacement vector; exponential variogram model shown.

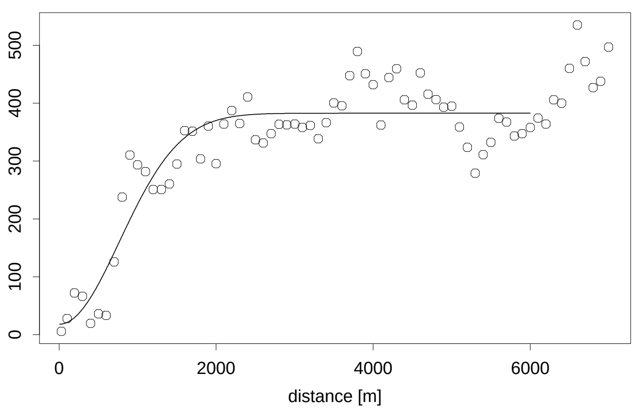

Fig. 5. Experimental semi-variogram as a function of the magnitude of the displacement vector; Gaussian semivariogram model shown.

III. Spatial Interpolation

This section deals with deterministic types of interpolation such as inverse distance weighting and statistical types known under the joint name of *kriging*. Specifically equations for simple and ordinary *kriging* are deduced.

A. Inverse Distance Weighting

Possibly the simplest conceivable way of carrying out interpolation consists of assigning the value of the nearest neighbour to a point where the value is unknown. An potential improvement consists of assigning higher weights to observations closer to the points to which we interpolate. An obvious way of doing this is to assign weights that are proportional to the inverse distance from the desired point to all N points entering into the interpolation. For the ith point we get the weight

$$w_i = \frac{1/d_i}{\sum_{j=1}^N 1/d_j},$$

where d_j is the distance from point j to the point to which we interpolate. This is readily extended to weighting with different powers, $p > 0$, of the inverse distance

$$w_i = \frac{1/d_i^p}{\sum_{j=1}^N 1/d_j^p}.$$

Other deterministic interpolation methods use (Delaunay) triangulation, regression analysis for determination of trend surfaces, minimum curvature etc., [15], [3].

A.1 Examples

We now wish to interpolate to Z_0 at position $r = 0$ in Figure 1 b.m.o. inverse distance weighting. d_i is the distance from point Z_i to Z_0. We readily calculate the following weights

r	d_i	$1/d_i$	$(1/d_i)/\sum(1/d_i)$
-2	2	1/2	3/11 $(= 0.2727)$
-1	1	1	6/11 $(= 0.5455)$
3	3	1/3	2/11 $(= 0.1818)$

For different powers of d_i we get the weights

r	d_i	$p = 0.1$	$p = 2.0$	$p = 10.0$
-2	2	0.3298	0.1837	0.0010
-1	1	0.3535	0.7347	0.9990
3	3	0.3167	0.0816	0.0000

We see that for low values of p the weights approach $1/p$ for all points used. For high values of p we get near a weight of one for the nearest neighbour.

B. Kriging

Kriging (after the South African mining engineer and professor Danie Krige) is a name for a family of methods for minimum error variance estimation. Consider a linear (or rather affine) estimate $\hat{z}_0 = \hat{z}(\boldsymbol{r}_0)$ at location $\boldsymbol{r}_0$ based on N measurements $\boldsymbol{z} = [z(\boldsymbol{r}_1), \ldots, z(\boldsymbol{r}_N)]^T = [z_1, \ldots, z_N]^T$

$$\hat{z}_0 \quad = \quad w_0 + \sum_{i=1}^N w_i z_i = w_0 + \boldsymbol{w}^T \boldsymbol{z},$$

where w_i are the weights applied to z_i and w_0 is a constant.

We consider z_i as realisations of stochastic variables Z_i, $\boldsymbol{Z} = [Z(\boldsymbol{r}_1), \ldots, Z(\boldsymbol{r}_N)]^T = [Z_1, \ldots, Z_N]^T$. We think of $Z(\boldsymbol{r})$ as consisting of a mean value and a residual $Z(\boldsymbol{r}) = \mu(\boldsymbol{r}) + \epsilon(\boldsymbol{r})$ with mean value zero and constant variance σ^2, $\mathrm{E}\{\epsilon\} = 0$ and $\mathrm{Var}\{\epsilon\} = \sigma^2$. For the linear estimator we get

$$\hat{Z}_0 = w_0 + \boldsymbol{w}^T \boldsymbol{Z}. \quad (2)$$

The estimation error $z_0 - \hat{z}_0$ is unknown. But for the expectation value of the estimation error we get

$$\begin{aligned} \mathrm{E}\{Z_0 - \hat{Z}_0\} &= \mathrm{E}\{Z_0 - w_0 - \boldsymbol{w}^T \boldsymbol{Z}\} \\ &= \mu_0 - w_0 - \boldsymbol{w}^T \boldsymbol{\mu}, \end{aligned} \quad (3)$$

where $\mu_0 = \mu(\boldsymbol{r}_0)$ is the expectation value of Z_0 and $\boldsymbol{\mu}$ is a vector of expectation values for $\boldsymbol{Z}$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu(\boldsymbol{r}_1) \\ \vdots \\ \mu(\boldsymbol{r}_N) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}.$$

We want our estimator to be unbiased or central, i.e., we demand $\mathrm{E}\{Z_0 - \hat{Z}_0\} = 0$ or

$$\mu_0 - w_0 - \boldsymbol{w}^T \boldsymbol{\mu} = 0. \quad (4)$$

The variance of the estimation error is

$$\begin{aligned} \sigma_E^2 &= \mathrm{Var}\{Z_0 - \hat{Z}_0\} \\ &= \mathrm{Var}\{Z_0\} + \mathrm{Var}\{w_0 + \boldsymbol{w}^T \boldsymbol{Z}\} \\ &\quad -2\mathrm{Cov}\{Z_0, w_0 + \boldsymbol{w}^T \boldsymbol{Z}\} \\ &= \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}), \end{aligned}$$

where $\boldsymbol{C}$ is the dispersion or variance-covariance matrix of the stochastic variables, $\boldsymbol{Z}$, entering into the estimation.

What is said in Section III-B so far is valid for all linear estimators. The idea in *kriging* is now to find the linear estimator which minimises the estimation variance.

B.1 Simple Kriging

In simple kriging (SK) we assume that $\mu(\boldsymbol{r})$ is known. From Equations 2 and 4 we get

$$\hat{Z}_0 - \mu_0 = \boldsymbol{w}^T(\boldsymbol{Z} - \boldsymbol{\mu}).$$

The weights w_i are found by minimising the estimation variance σ_E^2. This is done by setting the partial derivatives to zero

$$\frac{\partial \sigma_E^2}{\partial \boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\} = \boldsymbol{0},$$

which results in the SK system

$$\boldsymbol{C}\boldsymbol{w} = \mathrm{Cov}\{Z_0, \boldsymbol{Z}\}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} \\ \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \end{bmatrix},$$

where $C_{ij},\ i,j = 1,\ldots,N$ is the covariance between points i and j among the N points, which enter into the estimation of point 0. $C_{0j},\ j = 1,\ldots,N$ is the covariance between point j and point 0, the point to which we interpolate. We get these covariances from the semi-variogram model (remembering Equation 1, $\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h})$) as the sill minus the value of the semi-variogram model for the relevant distance (and possibly direction) between observations. (Alternatively, the kriging system may be formulated b.m.o. the semi-variogram; to avoid zeros on the diagonal of $\boldsymbol{C}$ we prefer the covariance formulation for numerical reasons.) Here C_{ij} must not be confused with the semi-variogram parameters C_0 and C_1.

The minimised squared estimation error termed the simple kriging variance is

$$\begin{aligned} \sigma_{\mathrm{SK}}^2 &= \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}) \\ &= \sigma^2 - \boldsymbol{w}^T \mathrm{Cov}\{Z_0, \boldsymbol{Z}\}. \end{aligned}$$

In SK the mean value $\mu(\boldsymbol{r})$ is known. In practice it is often assumed constant for the entire domain (or study area), or we must estimate it before the interpolation (or we must construct an interpolation algorithm which does not require knowledge of the mean field, see the next section.

B.2 Ordinary Kriging

In ordinary kriging (OK) we assume that the mean $\mu(\boldsymbol{r})$ is constant and equal to μ_0 for Z_0 and the N points that enter into the estimation of Z_0. From Equations 3 and 4 we get

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \mu_0(1 - \boldsymbol{w}^T \boldsymbol{1}) - w_0 = 0$$

for any μ_0. $\boldsymbol{1}$ is a vector of ones. This is possible only if $w_0 = 0$ and $\boldsymbol{w}^T \boldsymbol{1} = 1$.

The weights w_i are found by minimising σ_E^2 under the constraint $\boldsymbol{w}^T \boldsymbol{1} = 1$. A standard technique for minimisation under under a constraint is introducing a function F with a so-called Lagrange multiplier (here -2λ) which we multiply by the constraint set to zero and then minimising

$$F = \sigma_E^2 + 2\lambda(\boldsymbol{w}^T \boldsymbol{1} - 1)$$

without constraints. Again the partial derivatives are set to zero

$$\begin{aligned} \frac{\partial F}{\partial \boldsymbol{w}} &= 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\} + 2\lambda \boldsymbol{1} = \boldsymbol{0} \\ \frac{\partial F}{\partial \lambda} &= 2(\boldsymbol{w}^T \boldsymbol{1} - 1) = 0, \end{aligned}$$

which results in the OK system

$$\begin{aligned} \boldsymbol{C}\boldsymbol{w} + \lambda \boldsymbol{1} &= \mathrm{Cov}\{Z_0, \boldsymbol{Z}\} \\ \boldsymbol{1}^T \boldsymbol{w} &= 1 \end{aligned}$$

eller

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{N1} & \cdots & C_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \lambda \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \\ 1 \end{bmatrix}.$$

The values requested for C_{ij} are found as described in the previous section on SK.

The minimised squared estimation error termed the ordinary kriging variance is

$$\begin{aligned} \sigma_{\mathrm{OK}}^2 &= \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{Z_0, \boldsymbol{Z}\}) \\ &= \sigma^2 - \boldsymbol{w}^T \mathrm{Cov}\{Z_0, \boldsymbol{Z}\} - \lambda. \end{aligned}$$

OK implies an implicit re-estimation of μ_0 for each new constellation of points. This is an attractive property making OK well suited for interpolation in situations where the mean is not constant (i.e., in the absence of first order stationarity).

B.3 Examples

Let us consider the data in Figure 1 again. We now wish to interpolate to the position $r = 0$ b.m.o. ordinary kriging. To carry out the calculations we use a stipulated semi-variogram based on the spherical model with $C_0 = 0, C_1 = 1$ and $R = 6$. Remembering Equation 1, $C(h) = C(0) - \gamma(h)$, this gives the auto-covariance function (in this case where $C_0 + C_1 = 1$ this is the same as the auto-correlation function)

h	$\hat{\gamma}(h)$	$C(h)$
0	0.0000	1.0000
1	0.2477	0.7523
2	0.4815	0.5185
3	0.6875	0.3125
4	0.8519	0.1481
5	0.9606	0.0394
6	1.0000	0.0000

Therefore the OK system looks like this

$$\begin{bmatrix} 1.0000 & 0.7523 & 0.0394 & 1 \\ 0.7523 & 1.0000 & 0.1481 & 1 \\ 0.0394 & 0.1481 & 1.0000 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0.5185 \\ 0.7523 \\ 0.3125 \\ 1 \end{bmatrix},$$

where the values for C_{ij} come from the $C(h)$ table. The solution is $w_1 = -0.0407$, $w_2 = 0.7955$, $w_3 = 0.2452$ and $\lambda = -0.0489$, which gives a kriging variance of 0.3949. We see that even though Z_1 is closer to Z_0 than Z_3, the weight on Z_1 is much smaller than the weight on Z_3. This is an attractive characteristic of kriging, which allows for possible clustering of the sampling locations. We say that Z_2 *screens* for Z_1. This *screening* effect becomes weaker for higher nugget effects and it disappears for pure nugget effect (i.e., $C_1 = 0$ for the models shown here), where all weights become equal.

In a more realistic example Figure 6 shows a kriged (OK) map over heights above the ground water for the area near Slagelse. The interpolation is based on the isotropic Gaussian model for the experimental semi-variogram in Figure 5 (nugget effect 18 m^2, range 1.890 m and sill 364 m^2). We have kriged to 100 by 100 points in a 100 m by 100 m grid using a moving window to include the points from which to interpolate. The search radius of the moving window was 2,000 m and the estimation for each point was based on minimum 4 and maximum 20 points. We see that the interpolated map shows a good correspondance with the map of sample sites in Figure 2. Figure 7 shows the corresponding OK variances. We see that the kriging variance is large where the distances to the nearest samples are large.

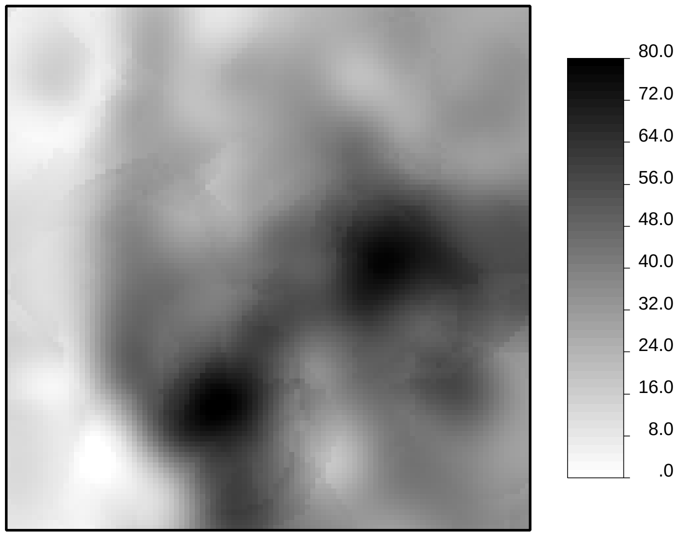

Fig. 6. Kriged map of heights above the ground water (unit is m).

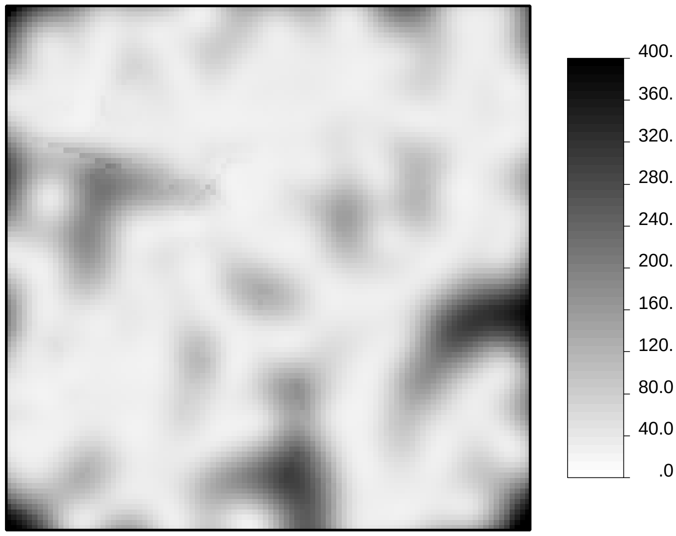

Fig. 7. Kriging variance corresponding to Figure 6 (unit is m^2).

B.4 Other Types of Kriging

If we wish to estimate average (also known as regularised) values over an area or a volume rather than point values, we may use *block kriging* which can be combined with several other forms of kriging.

If more variables are studied simultaneuously the above methods for description of spatial correlation may be extended to handle the spatial covariation between all pairs of variables in the form of cross semi-variograms or cross-covariance functions. Also more variables may be interpolated simultaneuously using *cokriging*. *Cokriging* is most useful when one variable is sampled on fewer locations than other correlated variables. *Universal kriging* is a method for the case where the mean value is described by linear combinations of known functions ideally determined by the physics of the problem at hand. Also methods for non-linear kriging exist such as *lognormal kriging*, *multi-Gaussian kriging*, *rank kriging*, *indicator kriging* and *disjunctive kriging*. References here are [1], [3], [9], [14].

IV. Conclusions

The above sections and examples demonstrate the following properties of kriging:

- Kriging is a type of interpolation that gives us both an estimate based on the spatial structure of the variable in question as expressed by the auto-covariance function (or the semi-variogram) as well as an estimation variance which is minimised.
- The kriging estimator is the best linear unbiased estimator (BLUE) in the sense that it minimises the estimation variance. Also it is exact, i.e., if we interpolate to a point which coincides with an existing sample point, kriging gives the same value as the one measured and the kriging variance is zero.
- The kriging system and the kriging variance depend on the auto-covariance function (or the semi-variogram) and the spatial layout of the sample locations only and not on the actual data values. If an auto-covariance function (or a semi-variogram)

is known (or assumed or imposed) this has important potential for minimising the estimation variance in experimental design (i.e., in the planning phase of the spatial lay-out of the sampling scheme).

• The solution of the kriging system implies a statistical distance weighting of the data values which enter into the interpolation. Also for OK, the weights are scaled so they add to one. Furthermore, possible redundancy in the form of clustering of the sample locations is accounted for; the above mentioned screening effect is due to this de-clustering.

• Because of the implicit re-estimation of the mean value for each new point constellation, OK is well suited for situations where the mean is not constant over the study region, i.e., where we don't have first order stationarity.

Further, the kriging system has a unique solution if and only if the covariance matric C (Section III-B) is positive definite; this also guarantees a non-negative kriging variance.

The strength of kriging may be attributted to a combination of the above characteristics.

If we choose to formulate the kriging system in terms of the auto-covariance function which is done in this note, we must assume (or impose) second order stationarity, i.e., the same auto-covariance function over the entire study area. The system may also be formulated in terms of the semi-variogram and in this case we must assume the intrinsic hypothesis, i.e., the same semi-variogram over the entire study area.

These assumptions may seem to be a drawback of kriging but if deterministic methods are applied, we implicitly make similar assumptions. It is hardly a drawback of geostatistical methods that we are forced to consider the timeliness of such assumptions.

V. Final Remarks

A sampling strategy may be based on the dependence of the kriging variance on the distance to the nearest samples. If the auto-covariance function (or the semi-variogram) and the sample locations are known, we can determine the kriging weights and the kriging variances before the actual sampling takes place. If the variances become too large in some regions of our study area we may modify the sample locations to obtain smaller variances. Also, to obtain a good estimate of the nugget effect which is an important parameter for the outcome of the kriging process, it may be an advantage to position some samples close to each other.

In multivariate studies where the joint behaviour of several variables is investigated, rather than interpolating the original variables we may interpolate combinations of them. For instance we may interpolate principal components of factors resulting from a factor analysis or a spatial factor analysis, [16], [17], [11], [18], [19]. Generel references to multivariate statistics are for example [20], [21]. [22] is written especially for geographers, [15] for geologists.

Lately there has been some interest in temporal aspects in connection with the application of data that vary in both space and time. Spatio-temporal semi-variograms and spatio-temporal kriging are dealt with in for example [23], [24]. A GIS for handling of temporal data is described in [25].

References

[1] A. G. Journel and Ch. J. Huijbregts, *Mining Geostatistics*, Academic Press, London, 1978, 600 pp.

[2] Isobel Clark, *Practical Geostatistics*, Elsevier Applied Science, London, 1979, http://www.stokos.demon.co.uk/.

[3] Edward H. Isaaks and R. Mohan Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989, 561 pp.

[4] K. Conradsen, A. A. Nielsen, and K. Windfeld, "Analysis of geochemical data sampled on a regional scale," in *Statistics in the Environmental and Earth Sciences*, A. Walden and P. Guttorp, Eds., pp. 283–300. Griffin, 1992.

[5] Noel A. C. Cressie, *Statistics for Spatial Data*, revised edition, Wiley, New York, 1993.

[6] P. A. Burrough and R. A. MacDonnell, *Principles of Geographical Information Systems*, Oxford University Press, 1998.

[7] M. F. Goodchild, B. O. Parks, and L. T. Steyaert, *Environmental Modeling with GIS*, Oxford University Press, 1993.

[8] H. Wackernagel, *Multivariate Geostatistics*, third edition, Springer, 2003.

[9] C. V. Deutsch and A. G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, second edition, Oxford University Press, 1998, Internet http://www.gslib.com/.

[10] Y. Pannatier, *VARIOWIN: Software for Spatial Data Analysis in 2D*, Springer, 1996, Internet http://www-sst.unil.ch/research/variowin/.

[11] Allan A. Nielsen, *Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data*, Ph.D. thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 1994, Internet http://www.imm.dtu.dk/pubdb/p.php?296.

[12] Karsten Hartelius, *Analysis of Irregularly Distributed Points*, Ph.D. thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 1996. Internet http://www.imm.dtu.dk/pubdb/p.php?1204.

[13] Allan A. Nielsen, "2D semivariograms," in *Proceedings of the Fourth South African Workshop on Pattern Recognition*, Paul Cilliers, Ed., Simon's Town, South Africa, 25–26 November 1993, pp. 25–35.

[14] Allan A. Nielsen, "Kriging," lecture note, Technical University of Denmark, 2004. Internet http://www.imm.dtu.dk/pubdb/p.php?3479.

[15] J. C. Davis, *Statistics and Data Analysis in Geology*, second edition, John Wiley & Sons, 1986.

[16] E. C. Grunsky and F. P. Agterberg, "Spatial and multivariate analysis of geochemical data from metavolcanic rocks in the Ben Nevis area, Ontario," *Mathematical Geology*, vol. 20, no. 7, pp. 825–861, 1988.

[17] E. C. Grunsky and F. P. Agterberg, "SPFAC: a Fortran program for spatial factor analysis of multivariate data," *Computers & Geosciences*, vol. 17, no. 1, pp. 133–160, 1991.

[18] Allan A. Nielsen, Knut Conradsen, John L. Pedersen, and Agnete Steenfelt, "Spatial factor analysis of stream sediment geochemistry data from South Greenland," in *Proceedings of the Third Annual Conference of the International Association for Mathematical Geology*, Vera Pawlowsky-Glahn, Ed., Barcelona, Spain, September 1997, pp. 955–960.

[19] Allan A. Nielsen, Knut Conradsen, John L. Pedersen, and Agnete Steenfelt, "Maximum autocorrelation factorial kriging," in *Proceedings of the 6th International Geostatistics Congress (Geostats 2000)*, W. J. Kleingeld and D. G. Krige, Eds., Cape Town, South Africa, April 2000. Internet http://www.imm.dtu.dk/pubdb/p.php?3639.

[20] Knut Conradsen, *Introduktion til Statistik*, Informatics and Mathematical Modelling, Technical University of Denmark, 1984.

[21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, second edition, John Wiley, New York, 1984.

[22] D. A. Griffith and C. G. Amrhein, *Multivariate Statistical Analysis for Geographers*, Prentice Hall, 1997.

[23] Annette Kjær Ersbøll, "A comparison of two spatio-temporal semivariograms with use in agriculture," in *Lecture Notes in Statistics*, T. G. Gregoire et al., Ed., vol. 122, pp. 299–308. Springer-Verlag, 1997.

[24] Annette Kjær Ersbøll and Bjarne Kjær Ersbøll, "On spatio-temporal kriging," in *Proceedings of the Third Annual Conference of the International Association for Mathematical Geology (IAMG'97)*, Vera Pawlowsky-Glahn, Ed., Barcelona, Spain, September 1997, pp. 617–622.

[25] Thomas Knudsen, *Busstop - A Spatio-Temporal Information System*, Ph.D. thesis, Niels Bohr Institute, Department of Geophysics, University of Copenhagen, Denmark, 1997.

Index Terms

auto-correlation function	auto-covariance function
BLUE	data analysis
distance weighting	estimation variance
geostatistics	GIS
GSLIB	intrinsic hypothesis
covariance function	kriging
kriging variance	Lagrange multiplier
least squares method	multivariate statistics
nested structures	nugget effect
ordinary kriging	sampling strategy
range of influence	regionalised variable
screening effect	semi-variogram
semi-variogram model	sill
simple kriging	spatial interpolation
spatial correlation	stationarity
statistics	stochastic function
stochastic process	stochastic variable
variogram	Variowin