GEODESY AND CARTOGRAPHY
Vol. 59, No 2, 2010, pp. 69-81
DOI No 10.2478/v10277-012-0002-0

Simple spatial prediction – least squares prediction, simple kriging, and conditional expectation of normal vector

Marcin Ligas, Marek Kulczycki

Department of Geomatics
AGH University of Science and Technology in Krakow
30 Mickiewicza Al., 30-059 Krakow, Poland
e-mail: marcin.ligas@agh.edu.pl, marek.kulczycki@agh.edu.pl

Received: 30 August 2010/Accepted: 26 November 2010

Abstract: The aim of the paper is the comparison of the least squares prediction presented by Heiskanen and Moritz (1967) in the classical handbook "Physical Geodesy" with the geostatistical method of simple kriging as well as in case of Gaussian random fields their equivalence to conditional expectation. The paper contains also short notes on the extension of simple kriging to ordinary kriging by dropping the assumption of known mean value of a random field as well as some necessary information on random fields, covariance function and semivariogram function. The semivariogram is emphasized in the paper, for two reasons. Firstly, the semivariogram describes broader class of phenomena, and for the second order stationary processes it is equivalent to the covariance function. Secondly, the analysis of different kinds of phenomena in terms of covariance is more common. Thus, it is worth introducing another function describing spatial continuity and variability.

For the ease of presentation all the considerations were limited to the Euclidean space (thus, for limited areas) although with some extra effort they can be extended to manifolds like sphere, ellipsoid, etc.

Keywords: least squares prediction, kriging, semivariogram, covariance function, random field

1. Introduction

Least squares prediction formula presented by Heiskanen and Moritz (1967) rooted in Wiener – Kolmogorov prediction theory (stochastic processes) was originally applied to prediction of gravity anomalies, but like with all spatial (also temporal) prediction methods, in their general case, they can be used to any random field (of course fulfilling some stationarity conditions). The same holds with kriging methods, originally developed for mining purposes, but with passing time they were bravely applied in many different fields.

As it will be shown in the paper, least squares prediction and geostatistical method of simple kriging are equivalent; going further, in case of Gaussian random fields, both are nothing but conditional expectation, and thus the best prediction.

The extensive overview and evolution of methods that contributed to the optimal spatial prediction can be found in literature (e.g. Cressie, 1990).

Original notation used by Heiskanen and Moritz (1967) was preserved with exception that all uncomfortable summation formulae have been changed into compact matrix form.

2. Brief on Random Fields

The extensive treatment of random fields' theory (also random functions, spatial stochastic processes) can be found in literature (e.g. Adler, 1981; VanMarcke, 1988; Christakos, 1992; Stein, 1999) and also in (Lauritzen, 1973) and (Krarup, 1969) who introduced second order stationary random functions and concepts of functional analysis with the use of Hilbert spaces, on the ground of physical geodesy. The presentation here is limited only to what is necessary to enable statistical inference on partial realization of random field, like it usually is the case in Earth sciences as well as construction of optimal predictors in the sense of minimum mean square error.

A random field is a set of random variables parameterised by some set $D \subset \Re^n$(in case $\Re^1$ one obtains stochastic processes, e.g. time series). The simplest form in which a random field can be introduced is as follows (Cressie, 1993; Schabenberger and Gotway, 2005):

$$\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \Re^n\} \tag{1}$$

where $Z(\mathbf{s})$ – random field; $\mathbf{s}$ – spatial coordinates; D – spatial domain; $\Re^n$ – n-dimensional Euclidean space.

The feasibility of statistical inference on single and partial realization of a random field as well as construction of optimal predictors is based on a notion of some form of stationarity. Assumptions of stationarity allow to treat the values at different places as though they are different realizations of the property (Webster and Olivier, 2007).

A random field is called second order stationary if the following assumptions hold:

$$\begin{aligned} &\forall_{\mathbf{s}\in D} \quad \mathrm{E}\left[Z\left(\mathbf{s}\right)\right] = \mu \\ &\forall_{\mathbf{s}_1,\mathbf{s}_2\in D} \quad \mathrm{Cov}\left[Z\left(\mathbf{s}_1\right), Z\left(\mathbf{s}_2\right)\right] = \mathrm{E}\left\{\left[Z\left(\mathbf{s}_1\right) - \mu\right]\left[Z\left(\mathbf{s}_2\right) - \mu\right]\right\} = \mathrm{C}\left(\mathbf{s}_1 - \mathbf{s}_2\right) \end{aligned} \tag{2}$$

where E – the expected value operator; Cov – covariance operator; $\mathbf{h} = \mathbf{s}_2 - \mathbf{s}_1$ is the separation vector between $Z(\mathbf{s}_1)$ and $Z(\mathbf{s}_2)$. The covariance function (2) can be expressed as follows:

$$\begin{aligned} \mathrm{Cov}\left[Z\left(\mathbf{s}_1\right), Z\left(\mathbf{s}_2\right)\right] &= \mathrm{Cov}\left[Z\left(\mathbf{s}_1\right), Z\left(\mathbf{s}_1 + \mathbf{h}\right)\right] = \mathrm{C}\left(\mathbf{h}\right) = \\ &= \mathrm{E}\left\{\left[Z\left(\mathbf{s}_1\right) - \mu\right]\left[Z\left(\mathbf{s}_1 + \mathbf{h}\right) - \mu\right]\right\} = \mathrm{E}\left[Z\left(\mathbf{s}_1\right) Z\left(\mathbf{s}_1 + \mathbf{h}\right)\right] - \mu^2 \end{aligned} \tag{3}$$

or, when the expected value of a random field is constant and equal to zero, i.e.

$$\forall_{\mathbf{s}\in D} \quad \mu = \mathrm{E}\left[Z\left(\mathbf{s}\right)\right] = 0 \tag{4}$$

then

$$\mathrm{Cov}\left[Z\left(\mathbf{s}_1\right), Z\left(\mathbf{s}_2\right)\right] = \mathrm{Cov}\left[Z\left(\mathbf{s}_1\right), Z\left(\mathbf{s}_1 + \mathbf{h}\right)\right] = \mathrm{C}\left(\mathbf{h}\right) = \mathrm{E}\left[Z\left(\mathbf{s}_1\right) Z\left(\mathbf{s}_1 + \mathbf{h}\right)\right] \tag{5}$$

From the above, one can notice, that the mean value of a second order stationary random field is constant over the entire domain D, and furthermore, the covariance function does not depend on absolute locations $\mathbf{s}_1$ and $\mathbf{s}_2$, but on the separation vector $\mathbf{h}$. The existence of the covariance function implies the existence of finite variance $\mathrm{C}(\mathbf{0}) = \mathrm{Cov}[Z(\mathbf{s}),Z(\mathbf{s}+\mathbf{h})] = \mathrm{V}[Z(\mathbf{s})]$ for $\mathbf{h} = \mathbf{0}$.

For the processes for which the above does not hold, i.e. neither covariance function nor variance exist, another hypothesis is introduced – the intrinsic hypothesis, and a random field is then called intrinsic stationary if the following assumptions hold:

$$\begin{aligned} &\forall_{\mathbf{s}\in D}\ \mathrm{E}\left[Z\left(\mathbf{s}\right)\right] = \mu \quad \text{or} \quad \mathrm{E}\left[Z\left(\mathbf{s}_1\right) - Z\left(\mathbf{s}_2\right)\right] = 0 \\ &\forall_{\mathbf{s}_1,\mathbf{s}_2\in D}\ \mathrm{V}\left[Z\left(\mathbf{s}_1\right) - Z\left(\mathbf{s}_2\right)\right] = 2\gamma\left(\mathbf{s}_1 - \mathbf{s}_2\right) \end{aligned} \tag{6}$$

where V – variance operator; γ – semivariogram; 2γ – variogram.

Expressing (6) by means of expected value operator one gets

$$\begin{aligned} 2\gamma\left(\mathbf{h}\right) &= \mathrm{V}\left[Z\left(\mathbf{s}_1\right) - Z\left(\mathbf{s}_2\right)\right] = \mathrm{V}\left[Z\left(\mathbf{s}_1\right) - Z\left(\mathbf{s}_1 + \mathbf{h}\right)\right] = \\ &= \mathrm{E}\left\{\left[Z\left(\mathbf{s}_1 + \mathbf{h}\right) - Z\left(\mathbf{s}_1\right)\right]^2\right\} - \left\{\mathrm{E}\left[Z\left(\mathbf{s}_1 + \mathbf{h}\right) - Z\left(\mathbf{s}_1\right)\right]\right\}^2 = \\ &= \mathrm{E}\left\{\left[Z\left(\mathbf{s}_1 + \mathbf{h}\right) - Z\left(\mathbf{s}_1\right)\right]^2\right\} \end{aligned} \tag{7}$$

If additionally the covariance function $\mathrm{C}\left(\mathbf{s}_1 - \mathbf{s}_2\right) = \mathrm{C}\left(\mathbf{h}\right)$ or semivariogram (variogram) $\gamma\left(\mathbf{s}_1 - \mathbf{s}_2\right) = \gamma\left(\mathbf{h}\right)$ depend only on separation distance between $\mathbf{s}_1$ and $\mathbf{s}_2$, i.e. $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ then random field is called isotropic and both functions are denoted $\mathrm{C}(h)$ and $\gamma(h)$, respectively.

3. The covariance function and the semivariogram

As it was shown in the previous section, there are two fundamental functions (the generalized covariance function for intrinsic random functions of order k (e.g. Matheron, 1973; Kitanidis, 1997; Chiles and Delfiner, 1999) is not considered here) which describe the behaviour of spatial process, i.e. covariance function and semivariogram.

The semivariogram as a structure function of intrinsically stationary random field describes a broader class of phenomena (covariance may not exist). Besides, semivariogram is superior to covariance function because it does not require mean value of a random field to be known; it simply filters it, and therefore, it became the preferred function of geostatisticians. There is, however, a limitation in using semivariogram – not all linear combinations of random variables are authorized (the note on admissible linear combinations). In the case of an Intrinsic Random Function – IRF, without a covariance, only linear combinations with the sum of coefficients equal to zero can be used (Matheron, 1971).

In case of second order stationary spatial processes there is equivalence between covariance function and semivariogram which can be derived as follows:

$$\begin{aligned} \mathrm{V}\left[Z\left(\mathbf{s}+\mathbf{h}\right)-Z\left(\mathbf{s}\right)\right] &= \mathrm{V}\left[Z\left(\mathbf{s}\right)\right]+\mathrm{V}\left[Z\left(\mathbf{s}+\mathbf{h}\right)\right]-2\mathrm{Cov}\left[Z\left(\mathbf{s}\right),Z\left(\mathbf{s}+\mathbf{h}\right)\right]= \\ &= 2\mathrm{V}\left[Z\left(\mathbf{s}\right)\right]-2\mathrm{Cov}\left[Z\left(\mathbf{s}\right),Z\left(\mathbf{s}+\mathbf{h}\right)\right]= \\ &= 2\left[\mathrm{C}\left(\mathbf{0}\right)-\mathrm{C}\left(\mathbf{h}\right)\right]=2\gamma\left(\mathbf{h}\right) \end{aligned} \tag{8}$$

and thus

$$\gamma\left(\mathbf{h}\right)=\mathrm{C}\left(\mathbf{0}\right)-\mathrm{C}\left(\mathbf{h}\right) \tag{9}$$

The semivariogram is a measure of dissimilarity between pairs of observations $Z(\mathbf{s+h})$ and $Z(\mathbf{s})$ (as opposed to covariance function which describes similarity). As a function, semivariogram provides information on spatial continuity and variability of a random function. The inference on the shape of semivariogram is based on empirical semivariogram and *a priori* knowledge of the behaviour of a phenomenon. Three characteristic parameters of semivariogram for second order stationary processes, i.e. the so called nugget effect c_o, partial sill c, and range of influence a (radius of autocorrelation) are shown in Figure 1. The sum of nugget effect and partial sill, i.e. $c_o + c$ is called sill.

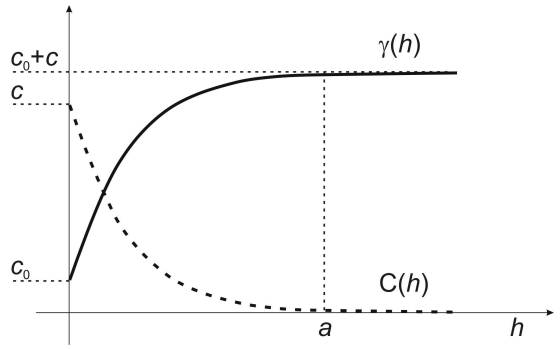

Fig. 1. The relation between the covariance function C(h) and semivariogram $\gamma(h)$ for second order stationary random fields

The most important issue whilst modelling the semivariogram is its behaviour near the point of origin which is strongly related to spatial continuity and regularity of a random field. Semivariogram approaching the origin may be: 1) quadratic (parabolic), what indicates a continuous and highly regular process or may be associated with the presence of a drift (trend); 2) linear, that indicates also continuous process but less regular than the previous one; or 3) discontinuous at the origin, which reveals highly irregular process at short distances (Armstrong, 1998; Journel and Huijbregts, 2003). The charts in Figure 2 illustrate the exemplary shape of 1D random field being characterized by different covariance functions (also semivariograms) with varying behaviour at short distances.

Both, semivariogram and covariance function have their special properties (Cressie, 1993; Matheron, 1971; Schabenberger and Gotway, 2005):

Covariance function	Semivariogram
positive definite function: $\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\lambda} \geq 0$, for any set of $\lambda_1, \lambda_2, ..., \lambda_n$ *even function*: $\mathrm{C}(\mathbf{h}) = \mathrm{C}(-\mathbf{h})$ *Schwarz's inequality*: $\mathrm{C}(\mathbf{0}) \geqslant \lvert\mathrm{C}(\mathbf{h})\rvert$ $\mathrm{V}[Z(\mathbf{s})] = \mathrm{Cov}[Z(\mathbf{s}), Z(\mathbf{s})] = \mathrm{C}(\mathbf{0})$	*conditionally negative definite function*: $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\lambda} \leq 0$, for any set of $\lambda_1, \lambda_2, ..., \lambda_n$ *even function*: $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$ $\gamma(0) = 0$, from definition, $\mathrm{V}[Z(\mathbf{s}) - Z(\mathbf{s})] = 0$ *in isotropic case when semivariogram is a function of distance only then* $\lim_{h\to\infty} \frac{\gamma(h)}{h^2} = 0$ (*semivariogram increases slower then* h^2)

where $\mathbf{C}$ – matrix of covariances; $\boldsymbol{\Gamma}$ – matrix of semivariances; $\boldsymbol{\lambda}$ – vector of coefficients.

For second order stationary random fields, from Schwarz's inequality

$$\gamma(h) \leq 2\mathrm{C}(0)$$

thus semivariogram for these processes is necessarily bounded.

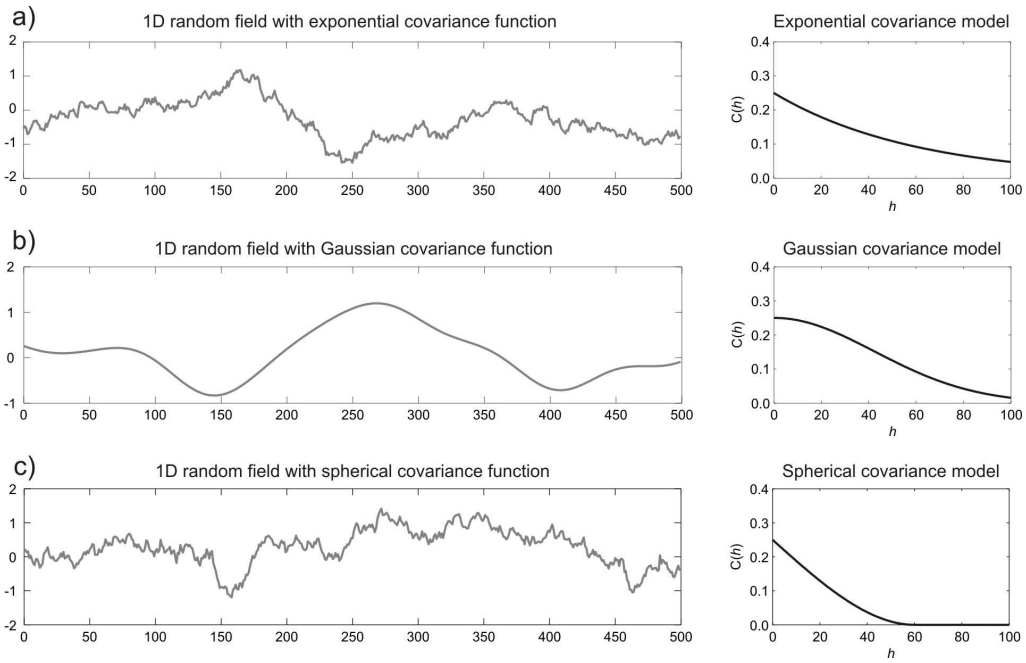

Fig. 2. 1-dimensional random fields characterized by different covariance functions: exponential covariance function a); Gaussian covariance function b); spherical covariance function c)

Note on admissible linear combinations

Variance of any finite linear combination of random variables $Y = \boldsymbol{\lambda}^T \mathbf{Z}$ must be non negative. For second order stationary case (covariance function exists) elementary calculus shows that the variance of Y is a quadratic form with the vector of coefficients $\boldsymbol{\lambda}$

$$\mathrm{V}(Y) = \mathrm{V}\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}\right) = \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\lambda} \geq 0 \tag{10}$$

where $\mathbf{C}$ is a covariance matrix of the vector $\mathbf{Z}$. Thus, for the quadratic form (10) to be non negative for any vector $\boldsymbol{\lambda}$, the covariance matrix $\mathbf{C}$ must be at least positive semi-definite (or better strictly positive definite). As it was shown above, in case of stationary random fields of order two any finite linear combination of random variables can be used, without any constraints as to the coefficients in $\boldsymbol{\lambda}$.

In case of intrinsic but not second order processes (with unbounded semivariograms) one can rewrite (10) in terms of semivariogram taking into account (9), thus

$$\begin{aligned}\mathrm{V}(Y) &= \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{C}_0 - \boldsymbol{\Gamma})\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}_0\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\lambda} = \\ &= \mathrm{C}_0\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{U}\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\lambda} = \mathrm{C}_0\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u}\mathbf{u}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\lambda}\end{aligned} \tag{11}$$

where $\mathbf{U}$ – "unit" matrix (here matrix consisting of all 1s); $\mathbf{u}$ – "unit" vector (here vector consisting of all 1s).

Expression (11) contains the parameter $\mathrm{C}(0) = \mathrm{C}_0$ (variance of a random field) that does not exist for intrinsic random fields. It can be eliminated from (11) by restricting coefficients in $\boldsymbol{\lambda}$ sum to zero, i.e.

$$\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u} = 0 \tag{12}$$

and finally

$$\mathrm{V}(Y) = -\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\lambda} \geq 0 \tag{13}$$

The above expression indicates that, the semivariogram must be conditionally negative definite function and the variance of finite linear combination of random variables can still be expressed in terms of semivariogram only when coefficients in $\boldsymbol{\lambda}$ sum to zero.

Concluding, in case of intrinsic random fields there is a trade-off between a broader class of phenomena and restricted class of linear combinations, i.e. not every single linear predictor can be derived through semivariogram.

4. Least squares prediction from Heiskanen and Moritz

The remarkable thing is that the only function needed to derive optimum predictors in the mean square sense is a covariance function (Heiskanen and Moritz, 1967). In the isotropic case, the covariance function of only one variable, is the function of the distance between the two points in space. In the derivation of the least squares prediction formula the original notation (Heiskanen and Moritz, 1967) was preserved; the only difference is in the compact matrix formulation.

In (Heiskanen and Moritz, 1967) the observed gravity anomalies are considered as a realization of zero mean second order stationary random function. If this does not hold, i.e. the mean gravity anomaly is not zero, one can form a new random function by subtracting the true (or estimated) mean from the observed values and add it back at the end of the prediction process.

The calculated gravity anomaly Δg_P at an unobserved point P is represented through the linear combination of known gravity anomalies. The predictor of Δg_P is thus

$$\Delta \tilde{g}_P = \boldsymbol{\alpha}^{\mathrm{T}} \Delta \mathbf{g} \tag{14}$$

where $\Delta \tilde{g}_P$ – predictor of gravity anomaly Δg_P at point P; $\Delta \mathbf{g}$ – vector of observed gravity anomalies; $\boldsymbol{\alpha}$ – vector of coefficients.

The prediction error is

$$\varepsilon_P = \Delta g_P - \Delta \tilde{g}_P = \Delta g_P - \boldsymbol{\alpha}^{\mathrm{T}} \Delta \mathbf{g} \tag{15}$$

By squaring and taking expected value of (15) we obtain the formula for mean square error of prediction:

$$\begin{aligned} m_P^2 &= \mathrm{E}\left(\varepsilon_P\right)^2 = \mathrm{E}\left(\Delta g_P - \boldsymbol{\alpha}^{\mathrm{T}} \Delta \mathbf{g}\right)^2 = \\ &= \mathrm{V}\left(\Delta g_P\right) + \mathrm{V}\left(\boldsymbol{\alpha}^{\mathrm{T}} \Delta \mathbf{g}\right) - 2\mathrm{Cov}\left(\Delta g_P, \boldsymbol{\alpha}^{\mathrm{T}} \Delta \mathbf{g}\right) = \mathrm{C}_0 + \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{C} \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{c} \end{aligned} \tag{16}$$

where $\mathrm{C}_0 = \mathrm{C}(0) = \mathrm{V}(\Delta \mathrm{g})$ is the variance of random function (variance of the gravity anomalies); $\mathbf{C}$ – matrix of covariances between observed gravity anomalies (consisting of covariances as a function of distance between pairs of points P_i and P_j at which gravity anomalies Δg_i and Δg_j were observed); $\mathbf{c}$ – vector of covariances between observed and unobserved gravity anomalies (consisting of covariances as a function of distance between pairs of points P_{P} and P_i, i.e. at point being predicted and points with known (observed) gravity anomalies)

To find the optimum set of coefficients $\boldsymbol{\alpha}$ that minimizes mean square error of prediction (16), an optimisation problem must be solved

$$\frac{\partial m_P^2}{\partial \boldsymbol{\alpha}} = 2\mathbf{C}\boldsymbol{\alpha} - 2\mathbf{c} = \mathbf{0} \tag{17}$$

from which a system of n equations with n unknowns is obtained

$$\mathbf{C}\boldsymbol{\alpha} = \mathbf{c} \rightarrow \boldsymbol{\alpha} = \mathbf{C}^{-1}\mathbf{c} \tag{18}$$

The solution of (18) gives the optimum, in mean square sense, set of coefficients $\boldsymbol{\alpha}$. By inserting $\boldsymbol{\alpha}$ to (14), the best prediction for unknown gravity anomaly is found

$$\Delta \tilde{g}_P = \boldsymbol{\alpha}^{\mathrm{T}} \Delta \mathbf{g} = \mathbf{c}^{\mathrm{T}} \mathbf{C}^{-1} \Delta \mathbf{g} \tag{19}$$

or, in case of a non-zero expected value, i.e. $\mathrm{E}(\Delta \mathrm{g}) = \mu \neq 0$

$$\Delta \tilde{g}_P = \mu + \boldsymbol{\alpha}^{\mathrm{T}} \Delta \mathbf{g} = \mu + \mathbf{c}^{\mathrm{T}} \mathbf{C}^{-1} \left(\Delta \mathbf{g} - \boldsymbol{\mu}\right) \tag{20}$$

Thus, the mean square error of prediction (prediction variance) is given by

$$\begin{aligned} m_P^2 &= \mathrm{E}\,(\varepsilon_P)^2 = \mathrm{C}_0 + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{C}\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{c} = \\ &= \mathrm{C}_0 + \mathbf{c}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}\mathbf{c} - 2\mathbf{c}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c} = \mathrm{C}_0 - \mathbf{c}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c} = \mathrm{C}_0 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{c} \end{aligned} \tag{21}$$

5. Simple kriging

There are two fundamental criteria which constitute the basis for obtaining optimum predictors (also estimators) in the field of geostatistics, these are unbiasedness and minimum mean square error of prediction (estimation).

Unbiasedness:

$$\mathrm{E}\left[\hat{Z}\,(\mathbf{s}_0) - Z\,(\mathbf{s}_0)\right] = \mathrm{E}\left[\mathrm{p}\,(\mathbf{Z}, \mathbf{s}_0) - Z\,(\mathbf{s}_0)\right] = 0 \tag{22}$$

Minimum mean square error of prediction:

$$\mathrm{E}\left[\hat{Z}\,(\mathbf{s}_0) - Z\,(\mathbf{s}_0)\right]^2 = \mathrm{E}\left[\mathrm{p}\,(\mathbf{Z}, \mathbf{s}_0) - Z\,(\mathbf{s}_0)\right]^2 \rightarrow \min \tag{23}$$

Theoretically, the simplest case of geostatistical prediction is simple kriging which can be introduced as heterogeneously linear predictor (Cressie, 1993; Rao and Toutenburg, 1999) of the form

$$\hat{Z}\,(\mathbf{s}_0) = \mathrm{p}\,(\mathbf{Z}, \mathbf{s}_0) = \lambda_0 + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}\,(\mathbf{s}) \tag{24}$$

Assume $Z(\mathbf{s})$ to be second order stationary random field with constant and known expected value (not necessarily zero), what may be written for observed data as

$$\mathbf{Z}\,(\mathbf{s}) = \boldsymbol{\mu} + \boldsymbol{\varepsilon}\,(\mathbf{s}) \tag{25}$$

where $\boldsymbol{\varepsilon}(\mathbf{s})$ – vector of errors (random vector with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{C}$); $\boldsymbol{\mu}$ – vector of constant and known mean values of a random field; and for unobserved to be predicted as

$$Z\,(\mathbf{s}_0) = \mu + \varepsilon\,(\mathbf{s}_0) \tag{26}$$

Predictor (24) is unbiased for the choice of λ_0 fulfilling the following condition:

$$\begin{aligned} \mathrm{E}\left[\hat{Z}\,(\mathbf{s}_0) - Z\,(\mathbf{s}_0)\right] &= \mathrm{E}\left[\mathrm{p}\,(\mathbf{Z}, \mathbf{s}_0) - Z\,(\mathbf{s}_0)\right] = \\ &= \mathrm{E}\left[\lambda_0 + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}\,(\mathbf{s}) - Z\,(\mathbf{s}_0)\right] = \lambda_0 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\mu} - \mu = 0 \end{aligned}$$

$$\lambda_0 = \mu - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\mu} \tag{27}$$

Inserting λ_0 from (27) to (24) the simple kriging predictor becomes

$$\begin{aligned} \hat{Z}\,(\mathbf{s}_0) = \mathrm{p}\,(\mathbf{Z}, \mathbf{s}_0) &= \lambda_0 + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}\,(\mathbf{s}) = \mu - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\mu} + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}\,(\mathbf{s}) = \\ &= \mu + \boldsymbol{\lambda}^{\mathrm{T}}\left(\mathbf{Z}\,(\mathbf{s}) - \boldsymbol{\mu}\right) = \mu + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\varepsilon}\,(\mathbf{s}) \end{aligned} \tag{28}$$

Thus, simple kriging predictor is unbiased regardless of the choice of weights. But as it was shown in the note on admissible linear combinations it limits the user to second order stationary processes, i.e. simple kriging predictor cannot in general be expressed in terms of the semivariogram.

Taking into account (26) and (28), the mean square error of prediction is expressed as

$$\begin{aligned} \mathrm{E}\left[\mathrm{p}\left(\mathbf{Z},\mathbf{s}_0\right)-Z\left(\mathbf{s}_0\right)\right]^2 &= \mathrm{E}\left[\mu+\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\mathbf{s}\right)-\mu-\varepsilon\left(\mathbf{s}_0\right)\right]^2 = \mathrm{E}\left[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\mathbf{s}\right)-\varepsilon\left(\mathbf{s}_0\right)\right]^2 = \\ &= \mathrm{V}\left[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\mathbf{s}\right)\right]+\mathrm{V}\left[\varepsilon\left(\mathbf{s}_0\right)\right]-2\mathrm{Cov}\left[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\mathbf{s}\right),\varepsilon\left(\mathbf{s}_0\right)\right] = \\ &= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\lambda}+\mathrm{C}_0-2\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c} \end{aligned} \tag{29}$$

Hence, on the basis of (29) the objective function $\Psi(\boldsymbol{\lambda})$ to be minimized is

$$\Psi\left(\boldsymbol{\lambda}\right)=\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\lambda}+\mathrm{C}_0-2\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c} \tag{30}$$

Taking partial derivatives with respect to the vector of coefficients $\boldsymbol{\lambda}$ and equating them to zero, one obtains simple kriging system of n equations with n unknowns

$$\frac{\partial\Psi\left(\boldsymbol{\lambda}\right)}{\partial\boldsymbol{\lambda}}=2\mathbf{C}\boldsymbol{\lambda}-2\mathbf{c}=\mathbf{0}\rightarrow\mathbf{C}\boldsymbol{\lambda}=\mathbf{c}\rightarrow\boldsymbol{\lambda}=\mathbf{C}^{-1}\mathbf{c} \tag{31}$$

The solution of (31) gives the optimum, in the mean square sense, set of kriging weights $\boldsymbol{\lambda}$. By inserting $\boldsymbol{\lambda}$ to (28), simple kriging predictor is obtained

$$\hat{Z}\left(\mathbf{s}_0\right)=\mathrm{p}\left(\mathbf{Z},\mathbf{s}_0\right)=\mu+\mathbf{c}^{\mathrm{T}}\mathbf{C}^{-1}\left[\mathbf{Z}\left(\mathbf{s}\right)-\boldsymbol{\mu}\right] \tag{32}$$

as well as simple kriging variance

$$\begin{aligned} \mathrm{V}\left(\mathbf{s}_0\right)=\mathrm{E}\left[\mathrm{p}\left(\mathbf{Z},\mathbf{s}_0\right)-Z\left(\mathbf{s}_0\right)\right]^2 &= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\lambda}+\mathrm{C}_0-2\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c} = \\ &= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c}+\mathrm{C}_0-2\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c}=\mathrm{C}_0-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c}=\mathrm{C}_0-\mathbf{c}^{\mathrm{T}}\mathbf{C}\mathbf{c} \end{aligned} \tag{33}$$

Note on a step further beyond simple kriging

Simple kriging is not as simple as its name states. There is nothing simple in unrealistic assumption of knowing *a priori* expected value of a random field (Schabenberger and Gotway, 2005). Chiles and Delfiner (1999) call it even "a wonderful case of a known mean". This can be the case in controlled field trials, with transformed variables or in "best linear unbiased prediction of residuals from regression fit provided that the model has been specified correctly" (Schabenberger and Gotway, 2005).

To weaken this assumption one can introduce ordinary kriging predictor which assumes constant but unknown mean value of a random field and can be expressed as homogenously linear predictor of the form

$$\hat{Z}\left(\mathbf{s}_0\right)=\mathrm{p}\left(\mathbf{Z},\mathbf{s}_0\right)=\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}\left(\mathbf{s}\right) \tag{34}$$

To fulfill a non-bias condition we put

$$\begin{aligned} \mathrm{E}\left[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right] &= \mathrm{E}\left[\mathrm{p}(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0)\right] = \mathrm{E}\left[\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}(\mathbf{s}) - Z(\mathbf{s}_0)\right] = \\ &= \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\mu} - \mu = \mu\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u} - 1\right) = 0 \end{aligned} \tag{35}$$

Hence, the non-bias condition requires that coefficients in $\boldsymbol{\lambda}$ sum to one, i.e.

$$\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u} = 1 \tag{36}$$

Mean square error of prediction for ordinary kriging is given by

$$\begin{aligned} \mathrm{E}\left[\mathrm{p}(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0)\right]^2 &= \mathrm{E}\left[\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}(\mathbf{s}) - Z(\mathbf{s}_0)\right]^2 = \\ &= \mathrm{V}\left[\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}(\mathbf{s})\right] + \mathrm{V}\left[Z(\mathbf{s}_0)\right] - 2\mathrm{Cov}\left[\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}(\mathbf{s}), Z(\mathbf{s}_0)\right] = \\ &= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\lambda} + \mathrm{C}_0 - 2\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c} \end{aligned} \tag{37}$$

Hence, the objective function to be minimized with the use of Lagrange multiplier can be expressed as

$$\begin{aligned} \Psi(\boldsymbol{\lambda}, \kappa) &= \mathrm{E}\left[\mathrm{p}(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0)\right]^2 - 2\kappa\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u} - 1\right) = \\ &= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\lambda} + \mathrm{C}_0 - 2\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c} - 2\kappa\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u} - 1\right) \rightarrow \min \end{aligned} \tag{38}$$

Taking partial derivatives with respect to the vector of coefficients $\boldsymbol{\lambda}$ and the Lagrange multiplier κ, and setting them to zero provides the system of $n+1$ equations with $n+1$ unknowns of the form

$$\begin{cases} \dfrac{\partial \Psi(\boldsymbol{\lambda}, \kappa)}{\partial \boldsymbol{\lambda}} = 2\mathbf{C}\boldsymbol{\lambda} - 2\mathbf{c} - 2\kappa\mathbf{u} = \mathbf{0} \\ \dfrac{\partial \Psi(\boldsymbol{\lambda}, \kappa)}{\partial \kappa} = -2\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u} - 1\right) = 0 \end{cases} \tag{39}$$

or, finally

$$\begin{cases} \mathbf{C}\boldsymbol{\lambda} - \kappa\mathbf{u} = \mathbf{c} \\ \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{u} = 1 \end{cases} \tag{40}$$

Ordinary kriging system of equations in matrix formulation is given by

$$\begin{bmatrix} C_{11} & \cdots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\kappa \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \\ 1 \end{bmatrix}$$

and ordinary kriging variance can be expressed as follows

$$\mathrm{V}(\mathbf{s}_0) = \mathrm{E}\left[\mathrm{p}(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0)\right]^2 = \mathrm{C}_0 - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c} + \kappa \tag{41}$$

By restricting kriging weights $\boldsymbol{\lambda}$ to sum to one, ordinary kriging predictor can also be derived by means of semivariogram function, i.e. for intrinsic stationary processes.

6. Conditional expectation of normal vector

The well known fact from the theory of multivariate normal distribution states that for the jointly normally distributed vector $\mathbf{Z} \in \Re^n \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ partitioned into $\mathbf{Z}_1 \in \Re^{n_1}$ and $\mathbf{Z}_2 \in \Re^{n_2}$, $n = n_1 + n_2$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \tag{42}$$

the conditional distribution of $\mathbf{Z}_2$ given $\mathbf{Z}_1$ is multivariate normal with $\mathbf{Z}_2|\mathbf{Z}_1 \sim N(\boldsymbol{\mu}_{\mathbf{Z}_2|\mathbf{Z}_1}, \boldsymbol{\Sigma}_{\mathbf{Z}_2|\mathbf{Z}_1})$ (Morrison, 1990; Deutsch, 1969)

$$\boldsymbol{\mu}_{\mathbf{Z}_2|\mathbf{Z}_1} = \mathrm{E}\,(\mathbf{Z}_2|\mathbf{Z}_1) = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\,(\mathbf{Z}_1 - \boldsymbol{\mu}_1) \tag{43}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{Z}_2|\mathbf{Z}_1} = \mathrm{V}\,(\mathbf{Z}_2|\mathbf{Z}_1) = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \tag{44}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are expected values of random vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively; $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are covariance matrices for $\mathbf{Z}_1$ and $\mathbf{Z}_2$; $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^{\mathrm{T}}$ represent crosscovariances between the elements of the vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

Here, in comparison to the first part of the paper the authors maintain the traditional notation from the theory of multivariate normal distribution denoting parameters as $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ *which should not confuse the reader.*

Now, we have similar situation as in (42), given second order stationary Gaussian random field, we want to predict $Z(\mathbf{s}_0)$ on the basis of observations $\mathbf{Z}(\mathbf{s})$, thus random vector $\widehat{\mathbf{Z}}(\mathbf{s})$ can be partitioned in the same way as (42), i.e.

$$\widehat{\mathbf{Z}}(\mathbf{s}) = \begin{bmatrix} \mathbf{Z}\,(\mathbf{s}) \\ Z\,(\mathbf{s}_0) \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu} \\ \mu \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{ZZ}} & \sigma_{\mathbf{Z}Z_0} \\ \sigma_{Z_0\mathbf{Z}} & \sigma^2_{Z_0Z_0} \end{bmatrix} \right) \tag{45}$$

where $\boldsymbol{\mu}$, μ are vector and scalar of mean values of random field respectively; $\boldsymbol{\Sigma}_{\mathbf{ZZ}}$ is a covariance matrix for $\mathbf{Z}(\mathbf{s})$; $\boldsymbol{\sigma}_{\mathbf{Z}Z_0} = \boldsymbol{\sigma}_{Z_0\mathbf{Z}}^{\mathrm{T}}$ are vectors of covariances between $Z(\mathbf{s}_0)$ and $\mathbf{Z}(\mathbf{s})$; and $\sigma^2_{Z_0Z_0}$ is the variance for $Z(\mathbf{s}_0)$. Applying again results from theory of multivariate normal distribution, i.e. (43) and (44) to (45) we obtain best linear prediction (46) which is in fact conditional expectation of $Z(\mathbf{s}_0)$ given data $\mathbf{Z}(\mathbf{s})$

$$\mu_{Z(\mathbf{s}_0)|\mathbf{Z}(\mathbf{s})} = \mathrm{E}\,[Z(\mathbf{s}_0)|\mathbf{Z}\,(\mathbf{s})] = \mu + \sigma_{Z_0\mathbf{Z}}\boldsymbol{\Sigma}_{\mathbf{ZZ}}^{-1}\,[\mathbf{Z}(\mathbf{s}) - \boldsymbol{\mu}] \tag{46}$$

$$\sigma^2_{Z(\mathbf{s}_0)|\mathbf{Z}(\mathbf{s})} = \mathrm{V}\,[Z(\mathbf{s}_0)|\mathbf{Z}(\mathbf{s})] = \sigma^2_{Z_0Z_0} - \sigma_{Z_0\mathbf{Z}}\boldsymbol{\Sigma}_{\mathbf{ZZ}}^{-1}\sigma_{\mathbf{Z}Z_0} \tag{47}$$

Concluding, the formulae (46) and (47) as well as (19), (20), (21) and (32), (33) are identical, which prove that in case of Gaussian random fields least squares prediction and simple kriging are nothing but conditional expectation, thus, the best linear unbiased predictors among all linear or non-linear predictors.

7. Conclusions

It was shown that the two methods; least squares prediction and simple kriging are equivalent. The striking thing is that they were developed at about the same time, in the sixties of the last century; also in the field of meteorology Lev Gandin introduced similar method which he called objective analysis. All the methods grew in different fields where the need of precise prediction was and still is of great importance. Later on, the methods evolved. In the field of physical geodesy blossomed as least squares collocation being an advanced method for prediction heterogeneous data. In geostatistics there is cokriging that is similar to collocation method, and variety of non-linear prediction methods like indicator kriging or disjunctive kriging.

Simple kriging as well as least squares prediction can easily be extended to the case of constant and unknown mean which is more robust method of prediction in respect of fluctuation in mean value – this extension in geostatistical parlance is called ordinary kriging. This kind of improvement broadens the range of phenomena under study to those which cannot be described by means of covariance function but for which the semivariogram exists and can be used instead of latter mentioned.

In case of Gaussian random fields, both, least squares prediction formula and simple kriging are nothing but conditional expectation, thus, the best linear unbiased predictor among all linear or non-linear predictors.

Acknowledgments

The paper is the result of research on prediction methods carried out within statutory research grant No 11.11.150.006 in the Department of Geomatics, University of Science and Technology, AGH, Krakow.

References

Adler R.J., (1981): *The geometry of Random Fields*, John Wiley & Sons, New York.
Armstrong M., (1998): *Basic linear geostatistics*, Springer, New York.
Chiles J.P., Delfiner P., (1999): *Geostatistics – modeling spatial uncertainty*, John Wiley & Sons, New York.
Christakos G., (1992): *Random field models in earth sciences*, Dover Publications Inc., New York.
Cressie N.A.C., (1990): *The origins of kriging*, Mathematical Geology, Vol. 22, No 3, pp. 239–252.
Cressie N.A.C., (1993): *Statistics for spatial data*, John Wiley & Sons, New York.
Deutsch R., (1969): *Estimation Theory*, Prentice – Hall, Inc. Englewood Cliffs, New York.
Heiskanen A.W., Moritz H., (1967): *Physical Geodesy*, W.H. Freeman and Company, San Francisco.
Journel A.G., Huijbregts Ch.J., (2003): *Mining geostatistics*, The Blackburn Press, New Jersey.
Kitanidis P.K., (1997): *Introduction to geostatistics – applications in hydrogeology*, Cambridge University Press, Cambridge.
Krarup T., (1969): *A contribution to the mathematical foundation of physical geodesy*, Geodaetisk Institut, Kobenhavn.
Lauritzen S.L., (1973): *The probabilistic background of some statistical methods in physical geodesy*, Geodaetisk Institut, Kobenhavn.

Matheron G., (1971): *The theory of regionalized variables and its application,* Ecole Nationale Superieure des Mines de Paris.

Matheron G., (1973): *The intrinsic random functions and their applications*, Adv. Appl. Prob., Vol. 5, pp. 439–468.

Morrison D.F., (1990): *Multivariate statistical methods*, McGraw-Hill, New York.

Rao C.R., Toutenburg H., (1999): *Linear models – least squares and alternatives*, Springer, New York.

Schabenberger O., Gotway C.A., (2005): *Statistical methods for spatial data analysis*, Chapman & Hall/CRC, New York.

Stein L.M., (1999): *Interpolation of spatial data – some theory for kriging*, Springer, New York.

VanMarcke E., (1988): *Random Fields*, MIT Press, Cambridge.

Webster R., Olivier M.A., (2007): *Geostatistics for environmental scientists*, John Wiley & Sons, New York.

Prosta predykcja przestrzenna – predykcja metodą najmniejszych kwadratów, kriging prosty oraz warunkowa wartość oczekiwana wektora normalnego

Marcin Ligas, Marek Kulczycki

Katedra Geomatyki
Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie
Al. Mickiewicza 30, 30-059 Kraków
e-mail: marcin.ligas@agh.edu.pl, marek.kulczycki@agh.edu.pl

Streszczenie

Motywację dla niniejszego artykułu stanowił rozdział „Metody statystyczne w geodezji fizycznej" pochodzący z uznawanej obecnie za klasykę gatunku pozycji „Geodezja fizyczna" (Heiskanen i Moritz, 1967), jak również zainteresowanie autorów metodami predykcji geostatystycznej. Celem artykułu jest studium porównawcze predykcji metodą najmniejszych kwadratów w ujęciu Heiskanena i Moritza z metodą krigingu prostego a w przypadku gaussowskich pól losowych ich równoważność z warunkową wartością oczekiwaną. Artykuł zawiera również rozszerzenie krigingu prostego do krigingu zwyczajnego poprzez odrzucenie założenia o znajomości wartości oczekiwanej pola losowego oraz podstawowe informacje na temat pól losowych, funkcji kowariancji i semiwariogramu. W treści artykułu większy nacisk położony został na funkcję semiwariogramu z dwóch powodów. Po pierwsze, semiwariogram opisuje bogatszą klasę zjawisk, a dla procesów stacjonarnych rzędu drugiego mamy zależność między funkcją kowariancji a semiwariogramem. Po drugie, analiza zjawisk za pomocą funkcji kowariancji jest powszechniejsza, zatem przedstawienie kolejnej funkcji wydaje się uzasadnione. Dla łatwości prezentacji wszystkie rozważania zostały ograniczone do przestrzeni euklidesowej (więc dla obszarów o ograniczonej powierzchni) jednakże z dodatkowym wysiłkiem rozważania mogą zostać uogólnione na przypadek rozmaitości takich jak sfera, elipsoida i inne.