Interpolation and 3D Visualization of Geodata

Marek KULCZYCKI and Marcin LIGAS, Poland

Key words: interpolation, kriging, real estate market analysis, property price index

ABSTRACT

Data derived from property markets have spatial character, no doubt about it, thus treating them with methods taking into account property location and so called "neighborhood influence" seem to be natural. In this paper we present a step by step algorithm for two of many geostatistical methods of spatial prediction which can find their application in property market analysis. As an exemplification of theoretical consideration, authors calculated particular problem and made it available on the website. It is worth noting that carrying out all the computations without dedicated software is very time consuming and might cause a headache. Authors implemented all the algorithms in one computer application and encourage for testing it simultaneously waiting for constructive critics.

Interpolation and 3D Visualization of Geodata

Marek KULCZYCKI and Marcin LIGAS, Poland

1. INTRODUCTION

Before describing a step by step algorithm it is necessary to introduce some terms which further part of the paper is based on, these are: random field and semivariogram.
Data on transactions from local property market make up a spatial data collection from the area called "spatial domain". Each transaction is determined by its location in a spatial domain, price and set of property characteristics. Such data representing values of distinguished characteristic under study (property price or one of its features) are presented on Figure 1 and described as $z(\mathbf{s}_i)$, where $\mathbf{s}_i$ for i=1…n stands for spatial location, $z(\mathbf{s}_0)$ stands for value being currently interpolated.

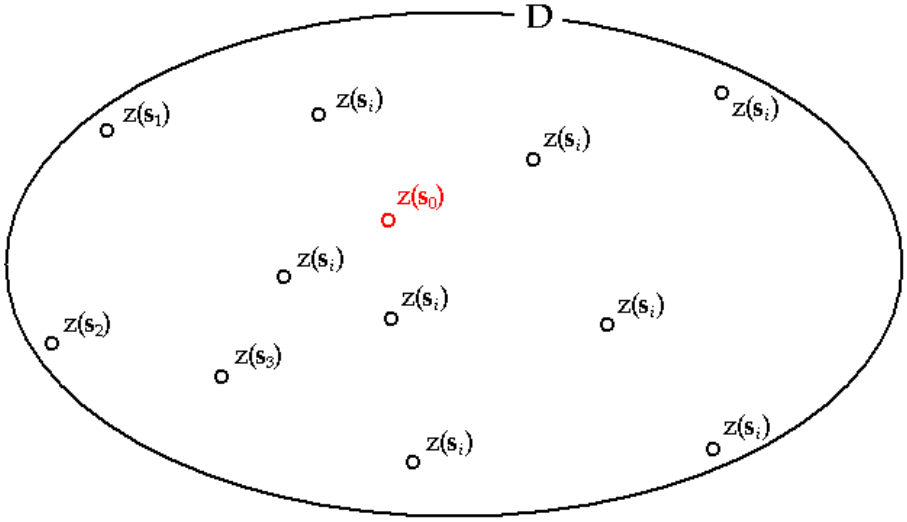

Figure 1. Image of the market, as a spatial domain „D"

We treat our data as a collection of random variables ordered by means of X and Y coordinates, determining location of real estate in geographical space. This collection makes up a single realization of random process also called a random field or spatial stochastic process and may be written as follows [SCHABENBERGER 2005, CRESSIE 1993]:

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

where:

s – vector of X, Y coordinates,
D – spatial domain,
Z(s) – random field, collection of random variables $z(s_i)$

On account of fact, that analyzing real estate market on particular date we have only single realization of random field there is a difficulty with statistical inferring on a sample of size "one". Elimination of this obstacle relies on introducing certain stationarity conditions of random fields.
Second order stationary random field is defined by its expected value and covariance function and can be described according to formula (2). Formula (2) shows that second order stationary random field has a constant expected value in the entire spatial domain "D" and covariance function depends only on a distance between observations.

$$\begin{aligned} E\big(Z(s)\big) &= \mu \\ Cov\big(Z(s), Z(s+h)\big) &= C(h) \\ V\big(\mathrm{Z(s)}\big) &= Cov\big(\mathrm{Z(s)}, \mathrm{Z(s)}\big) = C(0) \end{aligned} \tag{2}$$

where:

$E()$ – expected value operator,
$Cov()$ – covariance operator,
$V()$ – variance operator.

Second order stationarity condition is often too strong, thus intrinsic stationarity condition is introduced and defined as follows:

$$\begin{aligned} E\big(Z(s+h) - Z(s)\big) &= 0 \\ V\big(Z(s+h) - Z(s)\big) &= 2 \cdot \gamma(h) \end{aligned} \tag{3}$$

where:

$\gamma(h)$ – semivariogram (half of the variogram function),

As can be seen from (3), the semivariogram which is a fundamental function in geostatistics is defined by variance of increments of a random field. From practical point of view the relation between semivariogram function and covariance function for second order stationary processes is worth knowing. Relation afore mentioned for second order stationary processes can be written as follows [SCHABENBERGER 2005]:

$$\begin{gathered} V\big(Z(s+h) - Z(s)\big) = V\big(Z(s)\big) + V\big(Z(s+h)\big) - 2 \cdot Cov\big(Z(s), Z(s+h)\big) \\ \Downarrow \\ V\big(Z(s+h) - Z(s)\big) = 2 \cdot V\big(Z(s)\big) - 2 \cdot Cov\big(Z(s), Z(s+h)\big) \\ \Downarrow \\ V\big(Z(s+h) - Z(s)\big) = 2 \cdot \big(C(0) - C(h)\big) \\ \Downarrow \\ \gamma(h) = C(0) - C(h) \end{gathered} \tag{4}$$

Definition of random field as (1) gives no information about its structure, thus random field treated as mathematical representation of natural phenomena is presented as decomposition (data = structure + fluctuation).

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s}) \tag{5}$$

where, observed value of $Z(\mathbf{s})$ is decomposed into deterministic component $\mu(\mathbf{s})$ which describes global trend in the data and fluctuation $\varepsilon(\mathbf{s})$ which describes random deviation from the trend. In case of nonstationary random fields expected value changes over the spatial domain and is called large scale trend or drift. In case of stationary random fields expected value is constant over the spatial domain "D".

2. STEP 1 – GATHERING DATA AND INITIAL DATA ANALYSIS

The subject of the research is the data concerning purchases of properties. For each and every transaction, like with use of classical methods, the data like: date of sale, unit price and set of property characteristics are known. Moreover, for every single property the X and Y coordinates are known. Taking into consideration the limited level of trustworthiness of the transactions as well as the fact that the data are registered over time, the decision of starting geostatistical methods should be preceded by removing outliers and standardizing prices for single time point. These are carried out with the use of classical statistical methods thus they will not be the subject of the further divagation. The article presents no data nor intermediary results nor final results due to the volume of the data. To whom it might concern: the data and results are presented in extenso at http://geomatyka.agh.edu.pl/GSA.

3. STEP 2 – OPTIMAL STRUCTURE FUNCTION (SEMIVARIOGRAM)

Depending on which method of interpolation is chosen and what kind of data we have at hands, structure function of data might be estimated on the basis of row data or processed data – model's residuals.
We start with calculation of empirical semivariogram for example on the basis of classical semivariogram estimator (Matheron) or robust semivariogram estimator (Cressie, Hawkins). Calculations are based on starting parameters like: number of distance classes, direction, maximum lag distance. Both direction and lag distance are taken with tolerance what is presented on Fig (2).

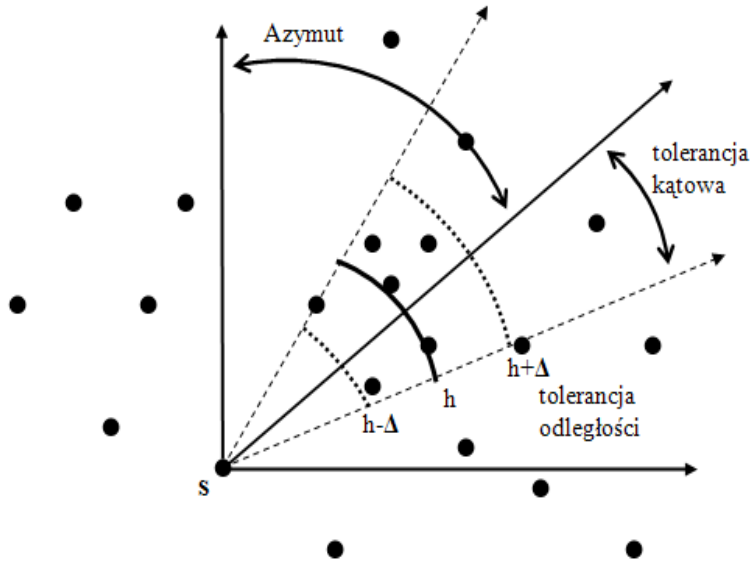

Figure 2. How empirical semivariogram is calculated.

For isotropic case, it is enough to construct semivariogram in one direction with 90° angle tolerance, in turn, for anisotropic case it is necessary to calculate structure function in at least 4 main direction i.e. N-S, E-W, NW-SE, NE-SW and transform phenomena to isotropic case.
Estimation of empirical semivariogram may be based for example either on classical (Matheron) estimator or robust estimator (Cressie, Hawkins).
Classiscal or Matheron's estimator can be written as follows:

$$\hat{\gamma}(h) = \frac{1}{2 \cdot N(h)} \cdot \sum_{N(h)} \left(Z(\mathbf{s}+h) - Z(\mathbf{s}) \right)^2 \tag{6}$$

where:
$\hat{\gamma}(h)$ – estimated value of empirical semivariogram for lag „h", semivariance at lag „h",
$N(h)$ – number of pairs in distance class corresponding to lag „h",
$Z(\mathbf{s})$ – observed value at point **s**,
$Z(\mathbf{s}+h)$ – observed value at point „**s** + **h**".

This estimator is unbiased but has a drawback i.e. is very sensitive to outliers, thus should be used with caution. Robust version of semivariogram estimator was proposed by Cressie and Hawkins and can be written as:

$$2 \cdot \hat{\gamma}(h) = \frac{\left(\frac{1}{N(h)} \cdot \sum_{N(h)} \sqrt{\left| Z(\mathbf{s}+h) - Z(\mathbf{s}) \right|} \right)^4}{\left(0.457 + \frac{0.494}{N(h)} \right)} \tag{7}$$

Further, among permissible theoretical semivariogram models we search for best fit to our empirical semivariogram. Theoretical semivariogram may be represented by means of different functions like: linear, exponential, power, gaussian, cubic, sine hole effect and many

more. In order to get the best fit according to least squares method principle we chose the function and weighting scheme and perform some calculations according to formula:

$$\left[\hat{\boldsymbol{\gamma}}(h)-\boldsymbol{\gamma}(\boldsymbol{\theta},h)\right]^{\mathrm{T}}\cdot\mathbf{W}\cdot\left[\hat{\boldsymbol{\gamma}}(h)-\boldsymbol{\gamma}(\boldsymbol{\theta},h)\right]=\boldsymbol{\varepsilon}^{\mathrm{T}}\cdot\mathbf{W}\cdot\boldsymbol{\varepsilon}=\min \tag{8}$$

where:

$\hat{\boldsymbol{\gamma}}(h)$ – vector of empirical semivariogram values,
$\gamma(\boldsymbol{\theta},h)$ – theoretical model of semivariogram as a function of estimated parameters $\boldsymbol{\theta}$,
$\mathbf{W}$ – diagonal weighting matrix,
$\boldsymbol{\varepsilon}$ – residual vetor.

In afore mentioned formula we can apply either empirical or derived analytically weights like these of Cressie which can be written as:

$$\mathbf{W}=\frac{\mathbf{N}(h)}{\boldsymbol{\gamma}^{2}(\boldsymbol{\theta},h)} \tag{9}$$

where:

$\mathbf{N}(h)$ – number of pairs in classes vector,
$\boldsymbol{\gamma}(\boldsymbol{\theta},h)$ – vector of theoretical semivariogram values,

Step 2 is hard to automate, thus requires the most effort. It is often carried out by trial and error method until satisfactory results are reached. Final decision is based on the experience of the analyst. The quality of fit is measured, among others, by means of Akaike Information Criterion (AIC) according to formula:

$$AIC=(n+1)\cdot\ln\left(\frac{\boldsymbol{\varepsilon}^{\mathrm{T}}\cdot\mathbf{W}\cdot\boldsymbol{\varepsilon}}{n+1}\right)+2\cdot n_{\mathrm{p}} \tag{10}$$

where:

$n+1$ – number of distance classes with „0” class included,
n_p – number of theoretical semivariogram parameters.

4. STEP 3 - INTERPOLATION

Interpolation considered as spatial prediction of unknown value on the basis of neighoring observations is called kriging. The name kriging comes from the name of southafrican engineer, pioneer of geostatistical methods – Daniel G. Krige.
Nowadays, we know many kriging methods like for example: simple kriging, ordinary kriging, universal kriging, kriging the mean, co-kriging, disjunctive kriging and kriging methods with external drift [CRESSIE 1993, KITANIDIS 1997, WACKERNAGEL 2003].
Taking into account the fact, that the subject of our interest is the data coming from real estate market, we want to distinguish two methods which can be useful on this field. The first is the universal kriging method, which can be applied to general analysis of property prices distribution in some spatial domain. The second, regression – kriging, which is based on similar assumptions like universal kriging but takes into account auxiliary variables (in this case, property characteristics) and so can be used for property value prediction purposes.

Universal kriging which can be also called kriging with an internal trend model is a generalization of methods like simple or ordinary kriging. Generalization relies on skipping the assumption on constant expected value in a spatial domain "D".
In universal kriging random field $\mathbf{Z}(\mathbf{s})$ is decomposed onto deterministic, known in entire spatial domain, part $\boldsymbol{\mu}(\mathbf{s})$ called a trend and random field $\boldsymbol{\varepsilon}(\mathbf{s})$ represented by random term. The most common form of a large scale trend $\boldsymbol{\mu}(\mathbf{s})$ is a polynomial form:

$$\boldsymbol{\mu}(\mathbf{s}) = \beta_0 + \sum_{i=1}^{n} \beta_i \cdot f_i(\mathbf{s}) = \mathbf{F}(\mathbf{s}) \cdot \boldsymbol{\beta} \tag{11}$$

The universal kriging model (for observed data) can be presented as (based on [SCHABENBERGER 2005], [OLEA 1999]) :

$$\mathbf{Z}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \boldsymbol{\varepsilon}(\mathbf{s}) = \mathbf{F}(\mathbf{s}) \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}(\mathbf{s}) \tag{12}$$

and for the value at s_0, $Z(s_0)$ like:

$$Z(\mathbf{s}_0) = \mu(\mathbf{s}_0) + \varepsilon(\mathbf{s}_0) = \mathbf{f}^{\mathrm{T}}(\mathbf{s}_0) \cdot \boldsymbol{\beta} + \varepsilon(\mathbf{s}_0) \tag{13}$$

The predictor $p(\mathbf{Z},\mathbf{s}_0)$ of $Z(\mathbf{s}_0)$ can be written as linear function of observed data:

$$p(\mathbf{Z},\mathbf{s}_0) = \boldsymbol{\lambda}^{\mathrm{T}} \cdot \mathbf{Z}(\mathbf{s}) \tag{14}$$

Universal kriging predictor must be unbiased, thus must fulfill the condition:

$$E\left(p(\mathbf{Z},\mathbf{s}_0) - Z(\mathbf{s}_0)\right) = 0 \tag{15}$$

Transforming (15) using formulas (12), (13) and (14):

$$\begin{aligned} &E\left(p(\mathbf{Z},\mathbf{s}_0) - Z(\mathbf{s}_0)\right) = \\ &= E\left(\boldsymbol{\lambda}^{\mathrm{T}} \cdot \mathbf{Z}(\mathbf{s}) - Z(\mathbf{s}_0)\right) = \\ &= E\left[\boldsymbol{\lambda}^{\mathrm{T}} \cdot \left(\mathbf{F}(\mathbf{s}) \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}(\mathbf{s})\right) - \left(\mathbf{f}^{\mathrm{T}}(\mathbf{s}_0) \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}(\mathbf{s}_0)\right)\right] = \\ &= \boldsymbol{\lambda}^{\mathrm{T}} \cdot \mathbf{F}(\mathbf{s}) \cdot \boldsymbol{\beta} - \mathbf{f}^{\mathrm{T}}(\mathbf{s}_0) \cdot \boldsymbol{\beta} = 0 \end{aligned} \tag{16}$$

Finally, we get a condition for unbiasedness of the predictor which can be written as:

$$\boldsymbol{\lambda}^{\mathrm{T}} \cdot \mathbf{F}(\mathbf{s}) = \mathbf{f}^{\mathrm{T}}(\mathbf{s}_0) \tag{17}$$

Mean square prediction error can be written as:

$$\begin{aligned} &E\left(p(\mathbf{Z},\mathbf{s}_0) - Z(\mathbf{s}_0)\right)^2 = \\ &= E\left(\boldsymbol{\lambda}^{\mathrm{T}} \cdot \mathbf{Z}(\mathbf{s}) - Z(\mathbf{s}_0)\right)^2 = \\ &= E\left(\boldsymbol{\lambda}^{\mathrm{T}} \cdot \boldsymbol{\varepsilon}(\mathbf{s}) - \boldsymbol{\varepsilon}(\mathbf{s}_0)\right)^2 = \\ &= V\left(\boldsymbol{\lambda}^{\mathrm{T}} \cdot \boldsymbol{\varepsilon}(\mathbf{s})\right) + V\left(\boldsymbol{\varepsilon}(\mathbf{s}_0)\right) - 2 \cdot Cov\left(\boldsymbol{\lambda}^{\mathrm{T}} \cdot \boldsymbol{\varepsilon}(\mathbf{s}), \boldsymbol{\varepsilon}(\mathbf{s}_0)\right) = \\ &= \boldsymbol{\lambda}^{\mathrm{T}} \cdot \mathbf{C}(h) \cdot \boldsymbol{\lambda} + \sigma_0 - 2 \cdot \boldsymbol{\lambda}^{\mathrm{T}} \cdot \boldsymbol{\sigma} \end{aligned} \tag{18}$$

Minimization of mean squared prediction error with an unbiasedness condition is carried out by means of Lagrange multipliers method, thus the minimization function can be written as:

$$\begin{aligned}\Psi(\boldsymbol{\lambda},\boldsymbol{\mu}) &= E\left(p(\mathbf{Z},\mathbf{s}_0)-Z(\mathbf{s}_0)\right)^2+2\cdot\boldsymbol{\mu}^{\mathrm{T}}\cdot\left(\mathbf{F}(\mathbf{s})^{\mathrm{T}}\cdot\boldsymbol{\lambda}-\mathbf{f}(\mathbf{s}_0)\right)= \\ &= \boldsymbol{\lambda}^{\mathrm{T}}\cdot\mathbf{C}(h)\cdot\boldsymbol{\lambda}+\sigma_0-2\cdot\boldsymbol{\lambda}^{\mathrm{T}}\cdot\boldsymbol{\sigma}+2\cdot\boldsymbol{\mu}^{\mathrm{T}}\cdot\left(\mathbf{F}(\mathbf{s})^{\mathrm{T}}\cdot\boldsymbol{\lambda}-\mathbf{f}(\mathbf{s}_0)\right)=\min\end{aligned} \tag{19}$$

Computing necessary partial derivatives against weights λ and Lagrange multipliers $\boldsymbol{\mu}$ and setting to zero we get the system of equations like shown below:

$$\begin{cases}\dfrac{\partial\Psi(\boldsymbol{\lambda},\boldsymbol{\mu})}{\partial\boldsymbol{\lambda}}=2\cdot\mathbf{C}(h)\cdot\boldsymbol{\lambda}-2\cdot\boldsymbol{\sigma}+2\cdot\mathbf{F}(\mathbf{s})\cdot\boldsymbol{\mu}=\mathbf{0} \\ \dfrac{\partial\Psi(\boldsymbol{\lambda},\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}=2\cdot\left(\mathbf{F}(\mathbf{s})^{\mathrm{T}}\cdot\boldsymbol{\lambda}-\mathbf{f}(\mathbf{s}_0)\right)=\mathbf{0}\end{cases} \tag{20}$$

and further:

$$\begin{cases}\mathbf{C}(h)\cdot\boldsymbol{\lambda}+\mathbf{F}(\mathbf{s})\cdot\boldsymbol{\mu}=\boldsymbol{\sigma} \\ \mathbf{F}(\mathbf{s})^{\mathrm{T}}\cdot\boldsymbol{\lambda}=\mathbf{f}(\mathbf{s}_0)\end{cases} \tag{21}$$

or in matrix notation:

$$\begin{bmatrix} c_{11}(h) & \cdots & c_{1n}(h) & 1 & f_1(s_1) & f_2(s_1) & \cdots & f_k(s_1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{n1}(h) & \cdots & c_{nn}(h) & 1 & f_1(s_n) & f_2(s_n) & \cdots & f_k(s_n) \\ 1 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \\ f_1(s_1) & \cdots & f_1(s_n) & 0 & 0 & 0 & \cdots & 0 \\ f_2(s_1) & \cdots & f_2(s_n) & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f_k(s_1) & \cdots & f_k(s_n) & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdots \\ \lambda_n \\ \mu_1 \\ \mu_2 \\ \cdots \\ \cdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} \sigma_1 \\ \cdots \\ \sigma_n \\ 1 \\ f_1(s_o) \\ f_2(s_o) \\ \cdots \\ f_k(s_o) \end{bmatrix} \tag{22}$$

Solving system of equations (22) for each and every unknown value we get a set of weights necessary for predicting this particular value. The measure of uncertainty is a minimized mean squared prediction error, called also universal kriging variance and can be written in the form:

$$\sigma^2(\mathbf{s}_0)=\boldsymbol{\lambda}^{\mathrm{T}}\cdot\mathbf{C}(h)\cdot\boldsymbol{\lambda}+\sigma_0-2\cdot\boldsymbol{\lambda}^{\mathrm{T}}\cdot\boldsymbol{\sigma} \tag{23}$$

Regression – kriging, called also kriging with external drift is a hybrid model (McBratney 2006). Regression – Kriging model is a modification of universal kriging. The trend model in RK is estimated on a basis of auxiliary variables rather than regression on coordinates like in universal kriging.

The Regression – Kriging predictor is a sum of separately estimated drift (trend) and residual component. For a single location with unknown response variable (Price) and with known explanatory variables (property characteristics) the RK predictor written in matrix notation is as follows (compare Universal Kriging predictor e.g. Schabenberger & Gotway 2005 p. 241):

$$P(s_0)=\mathbf{x}(s_0)\cdot\hat{\boldsymbol{\beta}}+\mathbf{w}^{T}\cdot\hat{\boldsymbol{\varepsilon}}(s) \tag{24}$$

where:

$P(s_0)$	–	predicted response variable at location (s_0), here the value of the property
$\mathbf{x}(s_0)$	–	vector of explanatory variables at (s_0), here characteristics of the property
$\hat{\boldsymbol{\beta}}$	–	vector of parameters estimates of the model $\mathbf{P}(s) = \mathbf{X}(s) \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}(s)$, estimated through OLS or EGLS, depending on the stage of estimation process, according to the formulas:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS} = \left(\mathbf{X}(s)^T \cdot \mathbf{X}(s)\right)^{-1} \cdot \mathbf{X}(s)^T \cdot \mathbf{P}(s)$$
$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{EGLS} = \left(\mathbf{X}(s)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{X}(s)\right)^{-1} \cdot \mathbf{X}(s)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{P}(s)$$

$\mathbf{P}(s)$	–	vector of response variables for model parameters estimation, here the prices of the properties
$\mathbf{X}(s)$	–	matrix of explanatory variables for model parameter estimation, here characteristics of the properties
$\mathbf{C}$	–	covariance matrix for the residuals $\hat{\boldsymbol{\varepsilon}}$, constructed on the basis of corresponding semivariogram (under second – order stationarity condition), $C(h) = C = C(0) - \gamma(h) = (C_0 + C_1) - \gamma(h)$ C_0 - nugget effect C_1 - scale $(C_0 + C_1)$ - sill $\gamma(h)$ - semivariogram value at distance (h)
$\mathbf{w}$	–	vector of kriging weights $\mathbf{C}^{-1} \cdot \mathbf{c}(s_0)$, $\mathbf{c}(s_0)$ - covariances between (s) and (s_0)
$\hat{\boldsymbol{\varepsilon}}(s)$	–	vector of OLS or EGLS residuals of the model, depending on the stage

The RK predictor can be written also, as follows:

$$P(s_0) = \mathbf{x}(s_0) \cdot \hat{\boldsymbol{\beta}} + \mathbf{w}^T \cdot \left(\mathbf{P}(s) - \mathbf{X}(s) \cdot \hat{\boldsymbol{\beta}}\right)$$

and a prediction variance at (s_0) is equal to (compare to universal kriging variance):

$$\sigma^2(s_0) = (C_0 + C_1) - \mathbf{c}(s_0)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(s_0) +$$
$$+ \left(\mathbf{x}(s_0) - \mathbf{X}(s)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(s_0)\right)^T \cdot \left(\mathbf{X}(s)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{X}(s)\right)^{-1} \cdot \left(\mathbf{x}(s_0) - \mathbf{X}(s)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(s_0)\right)$$

thus, standard error of prediction is:

$$\sigma(s_0) = \sqrt{\sigma(s_0)^2}$$

The entire iterative process of obtaining RK predictor can be written as follows:

1. *Obtain a starting estimate of* $\boldsymbol{\beta}$ *through OLS,* $\hat{\boldsymbol{\beta}}$
2. *Compute residuals* $\hat{\boldsymbol{\varepsilon}}(s) = \mathbf{P}(s) - \mathbf{X}(s) \cdot \hat{\boldsymbol{\beta}}$

3. *Produce empirical semivariogram and fit theoretical variogram of the residuals, get the parameters of the semivariogram and construct corresponding covariance matrix*
4. *Obtain a new estimate of* $\boldsymbol{\beta}$ *using*
   $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{EGLS} = \left(\mathbf{X}(s)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{X}(s)\right)^{-1} \cdot \mathbf{X}(s)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{P}(s)$
5. *Repeat steps 2-4 until the relative or absolute change in estimates of* $\hat{\boldsymbol{\beta}}$ *is small*

On the basis of (SCHABENBERGER 2005 p. 256)

6. Final step, predict the value at s_0 on the basis of drift (trend) component $\mathbf{x}(s_0) \cdot \hat{\boldsymbol{\beta}}$ and weighted residuals $\mathbf{w}^T \cdot \left(\mathbf{P}(s) - \mathbf{X}(s) \cdot \hat{\boldsymbol{\beta}}\right)$ and assess the accuracy of prediction.

5. CONCLUSIONS

The geostatistical methods are the best choice to be applied when there is a spatial autocorrelation detected (i.e. over the dataset concerning real estate there is nearly always spatial autocorrelation). The geostatistical methods require more effort and computer's processing power than classical methods. Despite afore mentioned the gain both in accuracy and effectiveness in mapping the results is great. The greatness of the result renders the additional effort worth undertaking.

The idea of universal kriging might be presented in few words in the following way: 1^o estimate the trend in data (regression on coordinates), 2^o subtract estimated trend from observed data, 3^o basing on residuals obtained estimate the structure function (semivariogram), 4^o compute kriging weights solving suitable system of equations, 5^o basing on the trend and interpolated residuals predict the value at $Z(s_o)$, 6^o map the results if needed.

For real estate analysis purposes universal kriging might be used for analyzing spatial distribution of real estate prices and for creating maps of real estate values. Another possible application emerges in authors' heads, that is, calculation of indexes of price property changes over time (PPI – Property Price Index). Periodical application of kriging method (as depicted at Fig. 3) enables, for very large number of real estates, recalculation of real estate values over various time points what in turn constitutes the base for finding relative change in real estate price. In order to achieve a proper estimation of afore mentioned indicators the proper grouping of real estates together with averaging in groups must be carried out.

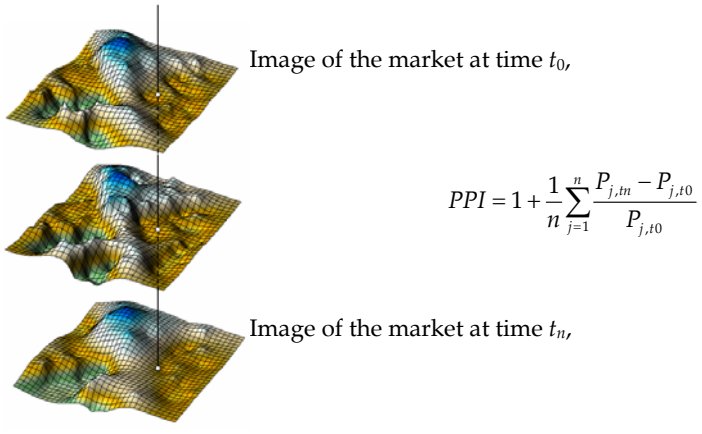

Figure 3. An example of application of kriging method to data derived from real estate market – property price index.

Regression – Kriging method shows an advantage over traditional approach to prediction based on regression models. In an unlikely case of data showing no spatial autocorrelation Regression – Kriging model reduces to classical model, thus in the worst case we will get the same results as using classical methods. Spatial interpolation methods seem to be very valuable and useful both during analyzing the real estate market as well as during appraisal. They perfectly fit for mapping results and as everybody knows "one picture is worth a thousand words" (old Chinese saying).

BIBLIOGRAPHY

Cressie N. A. C.: Statistics for spatial data. John Wiley & Sons 1993

Hengl T., A practical guide to geostatistical mapping of environmental variables, EUR 22904 EN Scientific and Technical Research series, Office for Official Publications of the European Communities, 2007

Kitanidis P. K.: Introduction to geostatistics – applications in hydrogeology. Cambridge, Cambridge University Press 1997

McBratney A., Odeh I., An overview of pedometric techniques for use in soil survey, available on http://natres.psu.ac.th/Link/SoilCongress/bdd/symp17/622-t.pdf, 2006

Olea R. A.: Geostatistics for engineers and earth scientists. Kluwer Academic Publishers 1999

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

Wackernagel H.: Multivariate geostatistics – an introduction with applications. Springer 2003

CONTACTS

Marcin Ligas
University of Science and Technology
Faculty of Mining Surveying and Environmental Engineering
Department of Geomatics
Al. A. Mickiewicza 30
30 – 059 Krakow
POLAND
Tel. + 48 12 617 44 80
Fax + 48 12 617 22 77
E-mail: marcin.ligas@agh.edu.pl

Marek Kulczycki
University of Science and Technology
Faculty of Mining Surveying and Environmental Engineering
Department of Geomatics
Al. A. Mickiewicza 30
30 – 059 Krakow
POLAND
Tel. + 48 12 617 44 80
Fax + 48 12 617 22 77
E-mail: marek.kulczycki@agh.edu.pl