

# Contribution to the application and terminology of geostatistical mapping methods in Croatia - Universal Kriging

The Mining-Geology-Petroleum Engineering Bulletin  
UDC: 550.8:552.1  
DOI: 10.17794/rgn.2017.4-3

Review scientific paper



Ivana Mesić Kiš

ES kralja Tomislava, Ulica Matice hrvatske 1, 31 500 Našice

## Abstract

The advancement of science and computer development enabled the processing and analysis of large amounts of data for which the geostatistical methods have become indispensable when researching and presenting the results in a variety of geosciences. One of the most noted and most accurate geostatistical estimation method is the Kriging method, which estimates the value of the analysed variables at unsampled locations. There are several Kriging techniques, while the most widely used is Ordinary Kriging, at the global level, as well as in Croatian geology. Here, Universal Kriging is described, the purpose of its use and the terminology associated with it in order to facilitate the understanding and use of this technique. It is a technique that is most appropriate when the input data is marked by a trend, and as such can often be applied in geology.

## Keywords:

geostatistical methods, residuals, terminology, Universal Kriging

## 1. Introduction

Geostatistical methods are commonly used for interpolation of, e.g., geological data. Field or other measurements are converted into a continuous surface allowing the distribution of analyzed variables. The purpose of spatial prediction is to estimate variable values at unsampled locations. Since the 1970s, one of most used estimators for interpolation of spatial data are Kriging techniques, of which Ordinary Kriging (abbr. OK) is most frequently used. It was confirmed by numerous Croatian authors as the best method for displaying the spatial distribution of reservoir variables (e.g., **Malvić and Đureković, 2003; Balić et al., 2008; Malvić, 2008; Husanović and Malvić, 2014**). Even so, the choice of which Kriging technique to use depends on the characteristics of the data. It is characterized as an advanced interpolation method since it is preceded by variogram analysis which determines spatial dependencies. Simpler interpolation methods such as inverse distance, nearest neighborhood and moving average are also used in Croatian geology (e.g., **Balić et al., 2008**).

Kriging techniques used for geological mapping in Croatia are diverse. For example, Factorial Kriging Analysis (abbr. FKA) is used for the mapping of chemical variables to determine soil contamination (e.g., **Cas-trignano and Romić, 2007; Sollitto et al., 2010**). Some authors use linear Kriging for geochemical mapping and

the assessment of pollution in soils (e.g. **Miko et al., 2000**), or Regression Kriging for predictive mapping of top soil organic matter (e.g., **Hengl et al., 2007**). According to this, Kriging found its place as an irreplaceable mapping method in geochemistry when assessing soil pollution and distribution of organic matter.

As mentioned before, the most commonly used Kriging technique is OK. In Croatia, it is used for estimating porosity and thickness (e.g., **Husanović et al., 2015**), and depth (e.g., **Špelić et al., 2016**). Of other Kriging techniques, it is important to mention Indicator Kriging which was used for mapping of the Late Miocene sandstone facies (**Novak Zelenika et al., 2010**).

The aim of this paper is to describe the Universal Kriging (abbr. UK) technique as another statistical tool for mapping geological variables and to contribute to new Croatian geological terminology and overall usage regarding the afore-mentioned technique. The purpose is to give appropriate terminology guidelines for using the UK technique as an advanced mapping method. It was previously described as a better interpolation method compared with OK in the case of the oil and gas Šandrovac Field where a depth variable of e-log marker Z' (the border between Pannonian and Pontian) with an expressed linear spatial trend was mapped (**Mesić Kiš, 2016a**).

## 2. About geomathematical terminology and its development in Croatia

When working with advanced geostatistical and geomathematical methods, knowing the appropriate termi-

nology is crucial for obtaining accurate results and understanding them. So far, there are several geostatistical and geomathematical dictionaries in Croatian literature (e.g., **Malvić et al., 2008; Malvić and Vrbanac, 2013; Malvić and Novak Zelenika, 2013**). The terminology is constantly expanding (e.g., **Mesić Kiš and Malvić, 2015; Mesić Kiš, 2016b, Malvić et al., 2016**) as a result of more research and popularization of the mentioned methods due to their accuracy in geological assessments of different variables.

Some authors are confused when trying to work with the Universal Kriging technique because there are two other terms for it and it is more mathematically advanced in comparison to the OK. The use of this technique is still in its initial stages in Croatian geology, therefore the terminology and the theory of the technique itself will be explained hereinafter.

The Universal Kriging is known as a Kriging technique with a trend and was originally described by **Matheron (1969)**. However, there are other terms that describe this technique, such as Regression Kriging (abbr. RK) and Kriging with external drift (abbr. KED). Most of the authors (e.g., **Deutsch and Journel, 1992; Wackernagel, 1998**) think that the UK technique refers to a case where the trend is modelled as a function of the coordinates only. When the drift is defined 'externally' through some auxiliary variables (such as parameters derived from digital elevation modelling, remote sensing/imagery or thematic maps), rather than coordinates, KED term is most commonly used (e.g., **Chiles and Delfiner, 1999; Wackernagel, 1998**). **Odeh et al. (1994, 1995)** suggested that the drift and residuals can be fitted separately and then summed afterwards, naming it RK. Compared with UK, the difference is that the regression in the UK is done with the spatial coordinates as the explanatory variables. Mathematically speaking, they are the same techniques, it is the co-variables that are different. According to **McBratney et al. (2000)**, UK, KED and RK belong to the group of the so-called "hybrid", i.e. non-stationary geostatistical methods (**Wackernagel, 1998**). A stationary process is one whose statistical properties such as mean, variance and autocorrelation, are constant.

### 3. The Application of Universal Kriging

The process of using UK can be described through the following steps:

1) It is necessary to understand why the trend exists based on the nature of our data. For example, when creating palaeostructural maps, one can presume a trend in data since mapped surfaces often have a structural trend towards one cardinal direction, or a sequence of geological structures.

2) It is required to use a simple form of the trend if possible and avoid extrapolation beyond available data. In geology, a linear trend can often be identified in input data, which is simple to use.

3) After the trend selection, it is subtracted from the observed data to obtain the residuals. Residuals ( $e$ ) represent the difference between the observed value of the dependent variable ( $y$ ) and the predicted variable ( $\hat{y}$ ), which can be expressed by the simple equation  $e = y - \hat{y}$ . A method for predicting the value of a dependent variable  $Y$ , based on the value of an independent variable  $X$ , is least squares linear regression.

If there are two independent variables, multiple regression analysis is used for trend removal. An estimated model would have following **Equation 1**:

$$\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \quad (1)$$

The estimation of  $\beta$  coefficients is given by resolving normal equations of the least square method. The sum of the residuals is always zero, whether the data set is linear or nonlinear. A residual plot is used to show the appropriate model for the data. It represents a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. A linear regression model is appropriate for the data if the points are randomly dispersed around the horizontal axis, otherwise, a non-linear model is more appropriate (<http://stat-trek.com/regression/residual-analysis.aspx?Tutorial=AP>).

4) Those residuals are used for variogram computation and spatial estimation. This step represents a big difference compared to the OK technique where variogram computation is used on observed data.

5) Finally, kriged residuals are added back to the trend (e.g., **Isaaks and Srivastava, 1989**). The result is a map interpolated with UK. Interpolation can be done in different software like Surfer Golden software (e.g., **Mesić Kiš, 2016a**) and ArcGIS (both are commercial), or freeware such are Gstat or SAGA GIS.

UK splits the random function into a linear combination of deterministic functions, the smoothly varying and nonstationary trend, that is also called a drift  $\mu(x) \in R$ , and a random component  $Y(x) := Z(x) - \mu(x)$  representing the residual random function (**Wackernagel, 2003**). Therefore, this technique assumes a nonstationary regionalized variable (the mean value varies and is not constant) that can be considered having two components (**Davis, 1973**): drift (average or expected value of the regionalized variable) and a residual (difference between the actual measurements and the drift). The random function  $Y(x)$  has a stationary covariance function  $C_y(h)$  that describes the covariance between any pair of random variables  $\{Y(x), Y(x+h)\}$  independently of the position of the point  $x$  in the domain (**Wackernagel, 1994**):

$$C_y(h) = E[Y(x) \times Y(x+h)] \quad (2)$$

The method of UK assumes that the mean function  $m(x)$ , also called the drift, has a functional dependence on spatial location and can be approximated by model with form (e.g. **Kumar, 2007**):

$$\mu(x) = \sum_{i=1}^k a_i f_i(x) \quad (3)$$

Where are:

- $a_l$  –  $l$ th coefficient to be estimated from the data,
- $f_l$  –  $l$ th basic function of spatial coordinates that describes the drift,
- $k$  – the number of functions used in modelling the drift.

Spatial trend or a drift represents any detectable tendency for the values to change as a function of the coordinate variables. It is assumed that the drift is a smooth function of the coordinate vector  $x$  and that it can be represented by a polynomial of order  $k$  with coefficients  $a_l$ .

The mean can be a function of the coordinates in a linear, quadratic or higher form. The mathematical model with two variables considering a linear trend is (e.g. **Kastelec and Košmelj, 2002**):

$$Z(s) = \beta_0 + \beta_1 x_1(s) + \beta_2 x_2(s) + \delta(s) \quad (4)$$

Where are:

- $Z(s)$  – random variable at location  $s$ ,
- $x_1(s)$  – longitude at location  $s$ ,
- $x_2(s)$  – latitude at location  $s$ ,
- $\beta_0, \beta_1, \beta_2$  – coefficients of the linear trend,
- $\delta(s)$  – intrinsic stationary random process with existing variogram  $2\gamma(h)$ , where  $h = s_i - s_j$  is a distance vector between locations  $s_i$  and  $s_j$ .

The kriging estimate is of the form (e.g. **Kumar, 2007**):

$$Z'(s_0) = \sum_{i=1}^N \lambda_i Z(s_i) \quad (5)$$

Where are:

- $Z'(s_0)$  – estimated value at  $s_0$ ,
- $Z(s_i)$  – observed values at points  $s_i$ ,
- $N$  – sample size,
- $\lambda_i$  – weights chosen for  $s_i$  so as to satisfy following statistical conditions:

(a) *Unbiasedness*

$$E\{Z'(s_0) - Z(s_0)\} = 0 \quad (6)$$

(b) *minimum variance*

$$\text{var}\{Z'(s_0) - Z(s_0)\} = \text{minimum} \quad (7)$$

According to **Journel and Huijbregts (1978)**, the minimisation of **Equation (7)** that is subject to the constraint of **Equation (6)**, using the Lagrange multiplier ( $\mu$ ), results in the following UK system:

$$\begin{cases} \sum_{j=1}^N \lambda_j \gamma(s_i, s_j) + \sum_{i=1}^k \mu_i f_i(s_i) = \gamma(s_i, s_0) & i = 1, 2, 3, \dots, N \\ \sum_{i=1}^N \lambda_i f_i(s_i) = f_i(s_0) & i = 1, 2, 3, \dots, k \end{cases} \quad (8)$$

Where are:

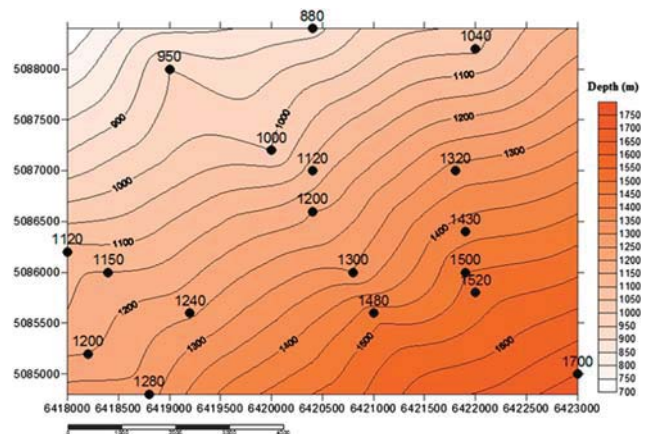
- $\gamma(s_i, s_j)$  – semivariogram between two points  $s_i$  and  $s_j$ ,
- $\mu_l$  – Lagrange multiplier associated with the  $l$ th unbiased condition.

The optimum weights  $\lambda_i$  can be obtained by solving these equations simultaneously. For the calculation of  $\lambda$ -s, a variogram which is a measure of spatial continuity of the data, is required (**Kastelec and Košmelj, 2002**). The variance of this estimation is given in **Equation 9** (e.g. **Kumar, 2007**):

$$\sigma_k^2(s_0) = \sum_{i=1}^N \lambda_i \gamma(s_i, s_0) + \sum_{i=1}^k \mu_i f_i(s_0) \quad (9)$$

As mentioned in the Introduction, the UK technique was used for the interpolation of a depth variable of e-log marker  $Z'$  in the case of the oil and gas Šandrovac Field (**Mesić Kiš, 2016a**). It represents a classical uplifted structure in the Bjelovar Subdepression (the southwestern part of Drava Depression and the Croatian part of the Pannonian Basin). Input data were depths of e-log marker  $Z'$ , which has chronostratigraphical importance in the Drava Depression. It is situated approximately in the middle of the 2<sup>nd</sup> transtension and represents the border between the Pannonian and Pontian (and between the formations Ivanić Grad and Kloštar Ivanić). The 2<sup>nd</sup> transtension lasted from the Late Pannonian (9.3-7.1 Ma) to the Early Pontian (7.1-6.3 Ma) and was marked by the deposition of a monotonous series of sandstones and marlstones.

**Figure 1** shows the successful application of the UK technique in the above mentioned area. It shows a homocline, where the entire Šandrovac Field sinks towards the SE. Because of that, it was relatively easy to determine a simple linear trend using a scatter plot that showed a positive linear trend of depth with longitude and a negative linear trend with latitude. Since there were two independent variables, multiple regression analysis was used for trend removal. Residuals were given by subtracting the real value of dependent variable from the estimated, which was then used for variogram analysis.



**Figure 1:** The Universal Kriging structural map (m) of Šandrovac Field (**Mesić Kiš, 2016a**)



### 4. Discussion

All interpolation methods estimate the value at a given location as a weighted sum of data values at adjacent locations. According to **Isaaks and Srivastava (1989)**, Kriging assigns weights according to a (moderately) data-driven weighting function, rather than an arbitrary function. Even so, it is still just an interpolation algorithm that can give similar results compared to other methods. However, an advantage of Kriging when compared to inverse distance or moving average methods, is that it provides a measure of the probable error associated with the estimates. It is most appropriate in a case of spatially correlated distance or directional bias in the data. Universal Kriging is recommended in a case of non-stationary data, when input data are described with a trend. It represents an adaptation of the OK technique that allows for the accommodation of a trend, e.g. when estimating temperature or a depth that are modelled as a function of the coordinates (north=cooler/south=warm; or when interpolating palaeostructural maps: N-S deeper, E-W lower depth, e.g., **Mesić Kiš, 2016a**). The UK technique produces good local estimates in the presence of a trend, and in a case when an experimental semivariogram indicates the non-stationarity of the data. The trend is easier to follow on a small surface (like Šandrovac Field) along some direction, in comparison to a bigger surface where it can be interrupted (for example by faults in geological surfaces). Uncertainties using Kriging techniques are determined through variogram analysis. Variograms are made using computer programs, but they can be incorrectly interpreted depending on the experience of an interpreter. In UK, a variogram model is used on residuals from the trend. Calculating this model properly is not simple.

Ordinary Kriging is the most used method for data interpolation in Croatian geology, but as aforementioned, UK has an advantage when having a trend in input data (especially on small surfaces). **Figure 2** shows a simple flow chart describing when to use OK or UK.

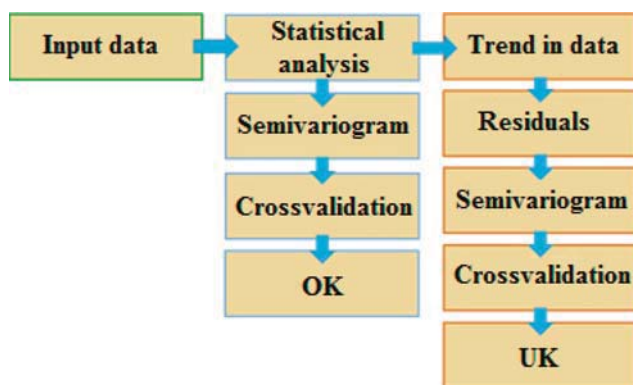


Figure 2: Flow chart

### 5. Conclusions

The purpose of this paper was to enable the reader to understand one of the Kriging techniques: Universal Kriging. The application of this technique is still in its initial stage in the mapping of geological variables in Croatia, and is suitable for the mapping of the area in which there is a structural trend. It is used on non-stationary data. A trend can be determined by using a scatter plot, after which residuals are obtained by the removal of the observed trend in the data. If the trend is not properly determined, the results will not be accurate. If an experimental semivariogram shows second order stationarity of the data, Ordinary Kriging would be a more acceptable method. Cross validation helps in the selection of an appropriate method of interpolation, where one with a smaller cross validation value is advisable to use.

### 6. References

Balić, D., Velić, J., Malvić, T. (2008): Selection of the most appropriate interpolation method for sandstone reservoirs in the Kloštar oil and gas field. *Geologica Croatica*, 61, 1, 23–75.

Castrignano, A., Romić, M.: Application of multivariate geostatistics to describe soil pollution by heavy metals in Northwestern Croatia (2007): Book of Abstracts TIES2007 Computational Environmetrics: Protection of Renewable Environment and Human and Ecosystem Health / Horova I. and Hřebiček J. (eds.).-Brno: Masaryk University, 21, (ISBN: 978-80-210-4333-6).

Chiles, J., Delfiner, P. (1999): *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 483 p.

Davis, J.C. (1973): *Statistics and Data Analysis in Geology*. Wiley, New York, 550 p.

Deutsch, C., Journel, A. (1992): *Geostatistical Software Library and User’s Guide*. Oxford Univ. Press, New York, 340 p.

Hengl, T., Heuvelink G.B.M., Rossiter, D.G. (2007): About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33, 1301-1315.

Husanović, E., Malvić, T. (2014): Review of deterministic geostatistical mapping methods in Croatian hydrocarbon reservoirs and advantages of such approach. *Nafta*, 65, 1, 57-63.

Husanović, E., Novak, K., Malvić, T., Novak Zelenika, K., Velić, J. (2015): Prospects for CO<sub>2</sub> carbonation and storage in Upper Miocene sandstone of Sava Depression, Croatia. *Geological Quarterly*, 59, 1, 91-104

Isaaks, E., Srivastava, R. (1989): *An Introduction to Applied Geostatistics*, Oxford University Press Inc., New York, 561 p.

Journel, A. G., Huijbregts, C. J. (1978): *Mining Geostatistics*, Academic Press, London, 600 p.

Kastelec, D., Košmelj, K. (2002): Spatial interpolation of mean yearly precipitation using universal kriging. In: Mrvar, A. and Ferligoj, A. (eds) *Developments in Statistics*, FDV, Ljubljana, Slovenia, 149-162.

- Kumar, V. (2007): Optimal contour mapping of groundwater levels using universal kriging – a case study, *Hydrological Sciences Journal*, 52, 5, 1038-1050. DOI: 10.1623/hysj.52.5.1038
- Malvić, T. (2008): Primjena geostatistike u analizi geoloških podataka (*Application of geostatistics in geological data analysis*). INA, SD Istraživanje i proizvodnja nafte i plina, Grupa za tiskarske i grafičke poslove, Zagreb, 105 p. (*In Croatian*)
- Malvić, T., Đureković, M. (2003): Application of the methods: Inverse distance weighting, ordinary kriging and collocated cokriging in porosity evaluation, and comparison of results on the Beničanci and Stari Gradac fields in Croatia. *Nafta*, 54, 9, 331–340.
- Malvić, T., Cvetković, M., Balić, D. (2008): Geomatematički rječnik (*Geomathematical dictionary*). HGD, Zagreb, 77 p. (*In Croatian*)
- Malvić, T., Novak Zelenika, K. (2013): Hrvatski rječnik odabranih geostatističkih pojmova (*Croatian geostatistical dictionary*). Rudarsko-geološko-naftni zbornik, 26, 1–9. (*In Croatian*)
- Malvić, T., Vrbanac, B. (2013): Geomatematički pojmovnik (*Geomathematical dictionary*). Hrvatski matematički elektronički časopis, 23, 1–49. (*In Croatian*)
- Malvić, T., Velić, J., Mesić Kiš, I. (2016): Treća nadopuna baze geomatematičkoga nazivlja (*Third upgrade of the geomathematical terminology*), Hrvatski terminološki portal Instituta za hrvatski jezik i jezikoslovlje, <http://nazivlje.hr> (*In Croatian*)
- Matheron, G. (1969): Le krigeage universel (*Universal Kriging*). Cahiers du Centre de Morphologie Mathématique, vol. 1. Ecole des Mines de Paris, Fontainebleau, 83 p. (*in French*)
- McBratney, A., Odeh, I., Bishop, T., Dunbar, M., Shatar, T. (2000): An overview of pedometric techniques for use in soil survey. *Geoderma*, 97, 3–4, 293–327.
- Mesić Kiš, I. (2016a): Comparison of Ordinary and Universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac Field. *Rudarsko-geološko-naftni zbornik*, 31, 2, 41–58.
- Mesić Kiš, I. (2016b): Hrvatski rječnik odabranih geomatematičkih i geostatističkih pojmova (*Croatian dictionary of selected geomathematical and geostatistical terms*). Zbornik radova “Matematičke metode i nazivlje u geologiji 2016”, Malvić, T.; Velić, J. (ur.). Zagreb: Rudarsko-geološko-naftni fakultet, 1-11. (*In Croatian*)
- Mesić Kiš, I., Malvić, T. (2015): Druga nadopuna baze geomatematičkoga nazivlja u terminološkoj bazi Struna Instituta za hrvatski jezik i jezikoslovlje (*Second upgrade of the geomathematical terminology in Struna database maintain by Institute for Croatian Language and Linguistics*), <http://struna.ihjj.hr> (*In Croatian*)
- Miko, S., Peh, Z., Bukovec, D., Prohić, E., Kastmüller, Ž. (2000): Geochemical baseline mapping and lead pollution assessment of soils on the karst in Western Croatia. *Nat. Croat.*, 9, 1, 41-59.
- Novak Zelenika, K., Malvić, T., Geiger, J. (2010): Mapping of the Late Miocene sandstone facies using indicator kriging. *Nafta*, 61, 5, 225-233.
- Odeh, I., McBratney, A., Chittleborough, D. (1994): Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, 63, 3-4, 197-214.
- Odeh, I., McBratney, A., Chittleborough, D. (1995): Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*, 67, 3-4, 215-226.
- Sollitto, D., Romić, M., Castrignano, A., Romić, D., Bakić, H. (2010): Assessing heavy metal contamination in soils of the Zagreb region (Northwest Croatia) using multivariate geostatistics. *Catena*, 80, 3, 182-194.
- Wackernagel, H. (1994): Cokriging versus kriging in regionalized multivariate data analysis. *Geoderma*, 62, 83-92.
- Wackernagel, H. (1998): *Multivariate Geostatistics: An Introduction With Applications*, 2nd ed. Springer, Berlin, 386 p.
- Wackernagel, H. (2003): *Multivariate Geostatistics: An Introduction with Applications*, (3rd ed.) Heidelberg: Springer, Berlin, 388 p.
- URL: <http://stattrek.com/regression/residual-analysis.aspx?Tutorial=AP> (accessed 5th March 2017)