FUZZY SURFACE MODELS BASED ON KRIGING OUTPUTS

Jan, CAHA[1], Pavel, TUČEK[2], Alena, VONDRÁKOVÁ[3], Lenka, PACLÍKOVÁ[4]

Department of Geoinformatics, Faculty of Science, Palacký University in Olomouc,

tr. Svobody 26, 771 46, Olomouc, Czech Republic

jan.caha@upol.cz[1], pavel.tucek@upol.cz[2], alena.vondrakova@upol.cz[3], lenka.paclikova01@upol.cz[4]

Abstract

Surface and its analyses are important topic in geosciences. Surface models are obtained from data with uncertainty in both position and attribute. The procedure which is used for modelling the surface from the original data results in biased outputs. Fuzzy set theory and fuzzy logic dispose of methods and tools for modelling surfaces based on biased input data. Especially fuzzy numbers are suited for modelling surfaces with uncertainty.

Kriging has been proved to be one of the best interpolation methods and it allows calculation of the standard deviation of the output surface. Based on those inputs, fuzzy surface can be constructed. The fuzzy-surface represents each point of the grid as a fuzzy number. Triangular and trapezoidal fuzzy numbers are the most common ones that are used in applications because of their easy implementation. The kernel of the fuzzy number is the result given by the kriging method and the support of the fuzzy number is calculated from the standard deviation of the kriging method. The output surface represents the interpolation of the data with the uncertainty that was present in the original data as well as the uncertainty that arises from the interpolation procedure. The aim of this paper is the creation of fuzzy surface based on the results of kriging calculation.

Fuzzy surface can be further used in geosciences for analyses of situations where the uncertainty of the result is important for decision making. Knowledge of uncertainty in calculations also allows much better risk management and provides more information for better crisis management.

Keywords: fuzzy, surface, kriging, fuzzy number, uncertainty

INTRODUCTION

All types of surfaces and their analyses have an important role in geosciences [1]. Often they are used as error free models even if significant errors and uncertainty can be and usually is present in those surfaces. Most of the time surfaces are created from data that do not cover the whole area of interest. Interpolation methods are used to create surface from discrete data. Creation of surface through different interpolation methods can lead to significantly different results. For necessary evaluation of the precision and quality of the result surface there are developed several methods that are widely used. These techniques include calculations of different kinds of errors and indexes that show surface quality. The most common known and used method is calculation of root mean square error (RMSE), next are absolute error and Hammock index [1]. Currently the uncertainty in the surface estimation is a topic that has the same importance as the surface quality. There are several ways how the uncertainty is introduced to the surface; however two of them are the most important: uncertainty in dataset and uncertainty in the interpolation process [2]. There exists also several methods for handling and propagation of uncertainty such as interval arithmetic, Bayesian statistic, Dempsey-Schaffer methods and fuzzy sets and fuzzy logic [3,4].

Fuzzy sets and logic become widely used theory for handling uncertainty in various disciplines. Most likely it is because of relatively simple concept and the fact that it is very close to the style of human thinking. People unlike machines or mathematical theories do not think in exact values but rather in vague terms. Fuzzy theory allowed using of vague terms in reasoning as well as in calculations. Because of this fuzzy theory became one of the most popular theories for handling uncertainty.

This contribution is related to the approach for creating fuzzy surfaces from the results of interpolation of input data by kriging method. Existing methods used in geosciences for the interpolation have been extended for use on fuzzy data, so there exists fuzzy IDW, fuzzy spline and fuzzy kriging methods [2,5]. However these methods are not yet implemented in GIS or in mathematical software such as R, Octave or Scilab. Also named methods are intended for use on fuzzy input data. In practical applications such data are not common. But there can be utilized methods of existing techniques of estimation provided by kriging algorithm and combined with expert opinion to create a fuzzy surface.

FUZZY SET THEORY

Nowadays it is well known that uncertainty is present in almost every information [3]. According to [3,6] and [7] a lot of phenomena are not statistical by nature and thus probability theory is not well suited for handling their uncertainty. Fuzzy set theory and fuzzy logic were developed as tools for analytical solving of problems that are not suited for probability theory and classic logic [8]. Main purpose of fuzzy theory can be described as precise description of imprecision (or vague) objects or phenomena [9]. Today fuzzy theory is used in wide area of disciplines for handling different data and processes that contain uncertainty.

Fuzzy set theory and fuzzy logic were first introduced by L. A. Zadeh in 1965 [8]. A fuzzy set is a collection of ordered pairs of objects and their membership grades. According to [2] the fuzzy set is defined as:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x))\} : x \in U \tag{1}$$

Where $\tilde{A}$ denotes a fuzzy set, U is a universe on which $\tilde{A}$ is defined, x is object from U and $\mu_{\tilde{A}}(x)$ is a degree membership of x to the fuzzy set $\tilde{A}$. Such fuzzy set is characterized by the membership function ($f_{\tilde{A}}(x)$) which associates each x to its degree of membership – number from interval [0,1]. The closer is the membership value to the 1 the more is the object part of the fuzzy set. Value 1 indicates complete membership to the set and value 0 means that the object does not belong to the set at all. Fuzzy set theory is generalization of classic (crisp) set theory, where $x \in A$ or $x \notin A$ as those are the extreme cases of membership degrees associated with values of 1 and 0 [fig.1]. Fuzzy logic utilizes similar concept using degrees of truth as a measure of correctness of the predicament. Some important terms connected with fuzzy set theory are:

- kernel – set of all x where $\mu_{\tilde{A}}(x) = 1$
- support – set of all x where $\mu_{\tilde{A}}(x) > 1$
- α-cut – set of all x where $\mu_{\tilde{A}}(x) \geq \alpha$ for $\alpha \in [0,1]$

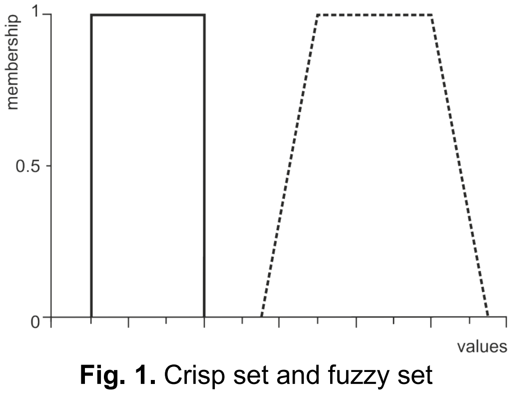

Fig. 1. Crisp set and fuzzy set

It is important to note, that all α-cuts, which special cases are both kernel and support, are crisp sets. This is important for constructing and calculating with fuzzy sets. For details about fuzzy set theory and fuzzy logic see [8].

Fuzzy number

Fuzzy numbers are considered as special cases of fuzzy sets. Fuzzy number is normalized convex fuzzy set used to represent vague value or number. Different types of fuzzy numbers exist i.e. Gaussian, Triangular, Trapezoidal, Piecewise Linear, etc. [fig.2] [2, 5, 10, 14, 16]. Another terminology uses notion fuzzy number for triangular fuzzy number and fuzzy interval for trapezoidal fuzzy number [18, 19, 20]. Because of simple implementation of both definition and calculations the triangular and trapezoidal fuzzy numbers are the most common in geosciences applications [2].

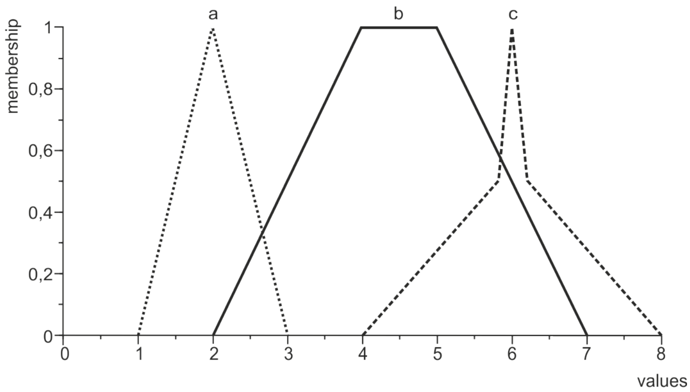

Fig. 2. Fuzzy numbers a) triangular b) trapezoidal c) piecewise linear

Basic triangular fuzzy number is often defined as a triplet $[a^-,a^0,a^+]$ in the same way the trapezoidal number can be defined as a quaternion $[a^-,a^{0-},a^{0+},a^+]$. Piecewise linear numbers can be defined by ordered pair values and their membership value or for easier computational treatment as set of α-cuts. Each α-cut is represented by interval which allows easier computational operations with such fuzzy numbers [10].

Fuzzy numbers can be used for calculations since all arithmetic operations can be defined for them through the extension principle [7]. However using extension principle was proven to be computationally complicated and methods for calculations with fuzzy numbers using α-cuts were developed [10]. Fuzzy arithmetic is important since it allows propagation of fuzzy numbers through any mathematical operation.

Fuzzy surface

As a fuzzy surface can be defined surface that instead of using crisp values uses fuzzy numbers. Such surface has included uncertainty since for each location is possible range of values. This is consistent with so called possibilistic fuzzy set theory [3,7].

CONVERTING RESULTS OF KRIGING TO FUZZY SURFACE

Kriging

Kriging is a set of geostatistical methods used to predict value of variable at location where it was not measured from the set of nearby points where the value of variable is known. Input for kriging is so called random field, that presents points where the variable was measured. For interpolation over the whole area of interest the regularly spaced grid of points is created and the prediction is made for each of those points [1]. Algorithm used by kriging is a linear least square estimator because it minimizes variance of the prediction error.

There are several types of kriging and the most common are simple, ordinary and universal kriging. Each of these methods have different assumptions: simple kriging presumes known constant trend of the variable, ordinary kriging presumes unknown constant trend and universal kriging assumes polynomial trend [11].

Before choosing which method is the most suitable to use statistical analysis that searches for trends in data have to be performed. However universal kriging seems to be the best choice since most of the geographical data contain some sort of trend [5].

Generally the universal kriging is defined [11]:

$$Z(x) = m(x) + \varepsilon(x) \qquad (2)$$

Z(x) is estimation of variable *Z* at the point *x*, *m(x)* is a structural component associated with the trend and *ε(m)* is regionalized variable which is stochastic and spatially correlated. Universal kriging estimator is defined:

$$Z^*(x) = \sum_{i=1}^{N} \lambda_i Z(x_i) \qquad (3)$$

where *N* is number of neighbors and *λ* is a vector of kriging coefficients. Results of universal kriging are obtained after minimalization of function:

$$\sum_{j=1}^{N} \lambda_j \gamma_\varepsilon(x_1 - x_j) + \sum_{l=1}^{P} \mu_l f_l(x_i) \qquad (4)$$

in the equation *γ* is a vector of variogram values, *μ* is vector of Lagrange multipliers and *f(x)* is a vector of values at sampled locations.

Kriging Standard Error

Same as many other statistical methods kriging allows besides estimation of value that have to be predicted also estimation of standard error of this prediction.

In [12] and [5] was proven that kriging standard error has no direct connection to the value that kriging is predicting, while the only connection is to the distance of the neighbour measure points. That is consistent with the kriging definition. The uncertainty (standard error) is higher if the points used for prediction are further away from the prediction location. If the neighbour points are close to the point of estimation the uncertainty is smaller. If the measurement points are located in regular grid the standard error prediction repeats for each part of the grid without any connection to the local variance of data in the grid [12] [fig.3].

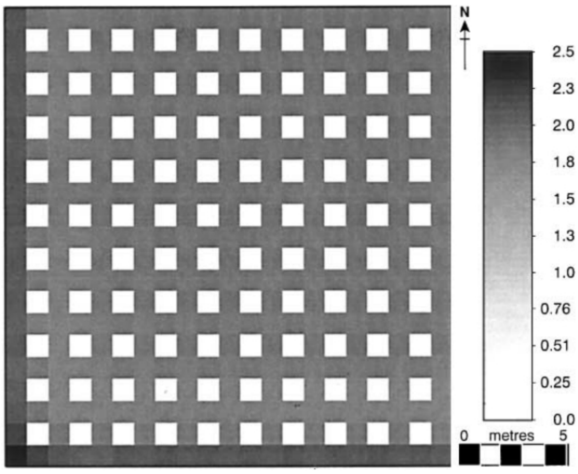

Fig. 3. Universal kriging standard error for regular grid of input data (source [12])

The value of standard error depends on the type of semivariogram model used for kriging. For different semivariograms the prediction of standard errors should look different, values vary significantly however the general trend that standard error is dependent on distance to measurement points is still significant [12].

Fuzzification of kriging standard error

It was defined that the standard error of kriging has no direct connection to the data since it is only function of distance to sampling points and covariance function [13]. This provides important information about the amount of uncertainty in the predicted surface, however this information is still not enough to allow a direct fuzzification of the predicted value to create a fuzzy surface.

All interpolating methods are based on presumption that value of interpolated phenomena is exactly known at set of exact locations. The resulting surface is then often treated as containing no uncertainty in its predicted values [2]. Several studies suggested how information sampled at discrete locations should be fuzzified for further analytical procession in the form of surface that contains uncertainty. In [14] where contour lines are used as a data source it is suggested that in 90% of cases the error in the input data should be smaller than the half of contour lines interval. This error is directly propagated from input data to the surface that was created. Based on this fact triangular number is constructed for each grid cell of the surface according to the equation:

$$\tilde{A} = \left[x_{ij} - \frac{ci}{4}, x_{ij}, x_{ij} + \frac{ci}{4}\right] \quad (5)$$

x_{ij} is predicted value of the surface at coordinates *ij* and *ci* is a contour interval. The output fuzzy number has support range equal to the half of contour lines interval.

Study [15] suggest using value of standard error - σ 0.2, 0.5, 1 and 2 meters as optimistic, two realistic and pessimist presumptions of surface vertical error created from contours lines with density of 1 meter. So in this attitude 95% of data are contained in interval denoted by:

$$I = [x_{ij} - 2\sigma, x_{ij}, x_{ij} + 2\sigma] \quad (6)$$

Another source [16] suggests creating of a triangular fuzzy number from the crisp input data by this formula:

$$\tilde{A} = [x_{ij} - \varepsilon, x_{ij}, x_{ij} + \varepsilon] \quad (7)$$

Where ε is the error of the measurement. This variant is based on uncertainty in data before the interpolation process, and the uncertainty is equal to the maximal possible error of measurement. The interpolation process is repeated three times to create surfaces: a^-, a^0, a^+. In such case the surface a^- represents the lowest possible surface of variable to predict, a^0 the most possible one and a^+ the highest possible surface.

The first and second study show similarities as the formulation about 90% of data with error lower then half of the interval size is close match with σ 0.2 for 1 meter density of contour lines, in this case 95% of data has error lower then 40% of the interval size. In second study this serves as an optimistic presumption.

Compromise solution from those studies is to assume that the maximal error in the surface estimation is equal to the measurement error in the input data. This only stands for estimations made by such algorithms that never exceeds the data or at least does not exceed them by much. So this presumption can not be used for spline without tension because spline algorithm tends to exceed the data range in many situations but spline with tension does not exceed the range of data so this presumption can be made for this algorithm. Based on these three studies can be identified the assumption that results of kriging prediction have the maximal possible error equal to the measure precision.

According to the [15] and [16] when assumption about error in data can be made it can be used for the creation of fuzzy surface. However the outlined principle does not incorporate uncertainty in the surface prediction, the assumption of uniformly distributed uncertainty over the surface is made. However kriging provides information about the distribution of uncertainty over the predicted surface. We use feature normalization to rescale the kriging standard error to interval of values [0,1]. Values of 1 indicate the areas with high uncertainty in estimation of the surface while values of 0 indicate areas with low uncertainty.

Based on these presumptions there can be created fuzzy surface that has at each point fuzzy number constructed by this equation:

$$\tilde{A} = [x_{ij} - (E_d + \varepsilon_{ij} * E_e), x_{ij}, x_{ij} + (E_d + \varepsilon_{ij} * E_e)] \quad (8)$$

where E_d is a minimal estimation error that we assume, E_e is an error that arise from the uncertainty of prediction. ε_{ij} denotes the normalized standard error of kriging at location *ij*. Together E_d plus E_e should be equal to the half of maximal possible error in input data, so that the range of the support of the fuzzy number is equal to the maximal possible error.

CASE STUDY – RUSAVSKÁ HORNATINA MOUNTAINS

Area Rusavská hornatina mountains is located on the east side of Czech republic [fig.4]. It is a part of Hostýnské vrchy mountains. The area of interest is a square of extent 4x4 kilometres. Input data set were contour lines with density of 5 meters. This quality of altitude data is usual in the Czech Republic.

Fig. 4. Location of the area of interest

Model area dataset was previously studied by [17]. In this study optimal interpolating algorithm for several areas of interest were determined. For Rusavská hornatina mountains the best algorithm with best evaluation was universal kriging. As the most correct setting kriging with second order trend removal with spherical theoretical semivariogram with 20 neighbours was identified. The reason for using spherical semivariogram is the need of preserving the local variability of the surface. Spherical or exponential semivariograms fulfil this need while gaussian preserves more of global trend then local variability. The surface was first created for area 4.1x4.1 km and later just area of interest of size 4x4 km was extracted to avoid errors that occur during interpolation process near the edges of data.

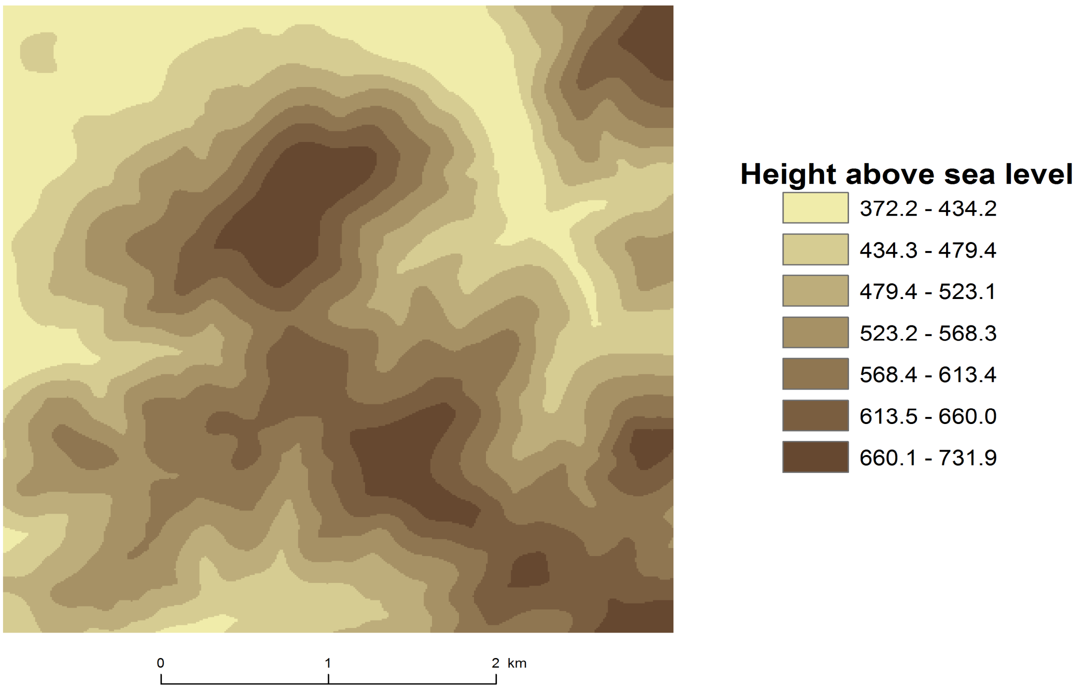

Fig. 5. Prediction of the terrain elevation

In previous chapter number of studies showing that the actual value of standard error does not have direct connection to the prediction values. More likely it shows the spatial distribution of the error prediction over the estimated surface. While the standard error of kriging has metrics it is not connected to the metrics of the prediction at least not in the sense of classic statistical description [12]. That means that it can not be used directly as a value of standard error to predict the range of possible values but it can be used as a metrics of uncertainty of the surface.

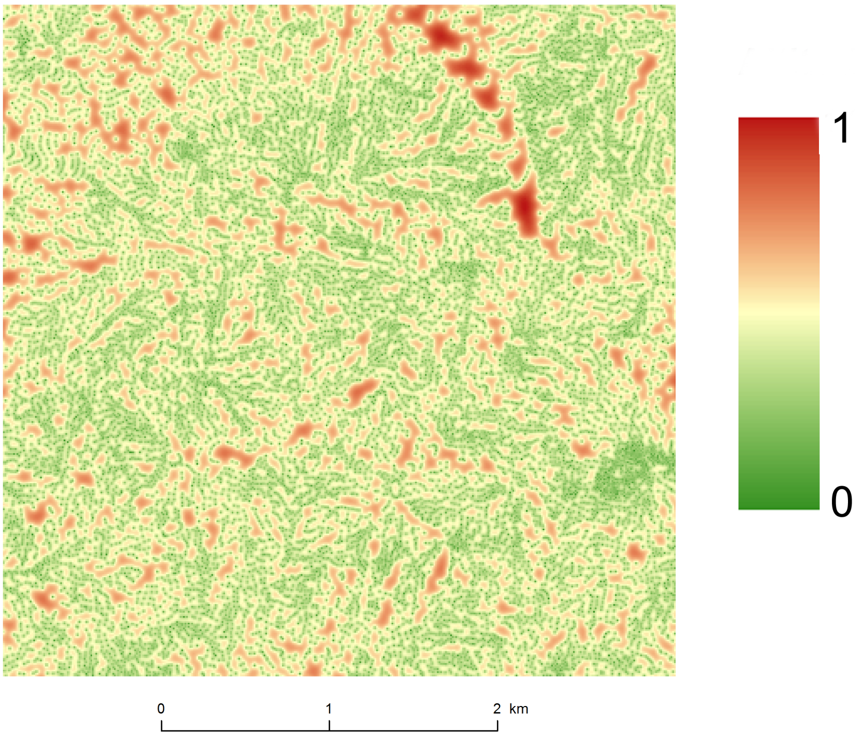

Fig. 6. Normalized standard error of kriging estimation

It is possible to use standard error in the estimated surface to point out areas that have higher uncertainty in estimation [fig.6]. Values of standard error have normal distribution so their normalization can be done through linear scaling to unit range by the equation:

$$x_n = \frac{x - min(X)}{max(X) - min(X)} \quad (\mathbf{9})$$

where x_n is normalized value of x, X is vector of all x. Then fuzzification of the result is realized by use of equation [8]. As E_d value of 0.5 meter was chosen and E_e is equal to 0.75 meters, together this gives range of the fuzzy number 2.5 meters as the worst case situation. Result can be visualized in 3D form, where the axis x,y corresponds to the location. Axis z shows estimated value at location which is also the value that has membership degree equal to 1 in the resulting fuzzy surface. Colour shows the maximal possible error in estimation of value at each location [fig.7].

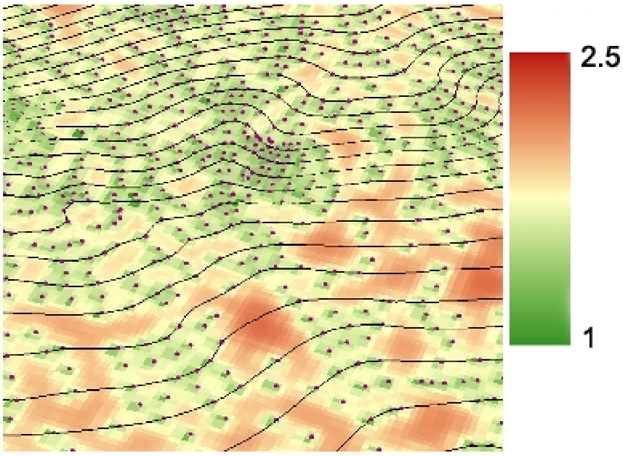

Fig. 7. Small area of 3D visualization showing input points and their relation to the resulting uncertainty (units are meters)

DISCUSSION

With the development of modern technologies, measurements are more precise and the effort of scientists is to make the presentation of these data in the most precise form. Therefore it is important in addition to the actual results also present the accuracy of this result. In the field of surface modelling it is mainly the uncertainty in the interpolation process. Current technologies do not allow efficient visualization of fuzzy surfaces, so it is necessary to find suitable alternatives. Possible way how to provide the most accurate information is presenting of information of surface confidence or possible erroneous of surface based on fuzzy model.

Presentation of possible uncertainty of the model is very important in every discipline such as monitoring of landslides. If the information is low quality because the derived model has a high uncertainty proposed measure can be completely inappropriate. This can occur in a situation where the points in the specific area did not occur at all and the surface model is only the result of interpolation of points more distant. Conversely, if there are many entry points on the measured site, the potential uncertainty of the model is low and the surface model can be used for detailed studies. Example in [fig.7] presents a situation where are entry points of surface model in a large number and perhaps the model uncertainty is very low in this area and very far can be location with very low number of entry points and the model uncertainty can by high and possible error can reach the maximum values.

In various models of natural processes there occurs uncertainty or vagueness arising from the very nature of the monitored phenomenon. Therefore it is very important to avoid bringing an error into other possible calculation. This error can be based on the fact that the model is considered as error free. Such an error could negatively affect the results of each study.

CONCLUSION

The proposed method allows creation of fuzzy surface containing uncertainty based on output of kriging interpolation method and expert knowledge. Creation of such surface is not computationally demanding and all the necessary methods are implemented in common GIS software. The output fuzzy surface can be further used in operations where the calculation of uncertainty is an important part of the decision making process. Derivative characteristics can be created from such surface and uncertainty of the surface can be propagated to these derivatives. This allows much better treatment of uncertainty during following analysis of both surface and its derivatives.

REFERENCES

[1] Burrough, P. A. and McDonnell, R. (1998) Principles of geographical information systems. Oxford University Press , Oxford.

[2] Lodwick, W. A. (2008) Fuzzy surfaces in GIS and geographical analysis : theory, analytical methods, algorithms, and applications. CRC Press, Boca Raton.

[3] Bandemer, H. (2006) Mathematics of uncertainty : ideas, methods, application problems. Springer, Berlin.

[4] Lodwick, W. A. and Santos, J. (2003).Constructing consistent fuzzy surfaces from fuzzy data. Fuzzy Sets and Systems, 135(2), 259-277.

[5] Bandemer, H. and Gebhardt, A. (2000) Bayesian fuzzy kriging. Fuzzy Sets and Systems, 112(3), 405-418.

[6] Zadeh, L. A. (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems, 90(2), 111-127.

[7] Celikyilmaz, A. and Turksen, I. B. (2009) Modeling uncertainty with fuzzy logic : with recent theory and applications. Springer, Berlin.

[8] Zadeh, L. A. (1965). Fuzzy Sets. Information and Control, 8(3), 338-353.

[9] Zadeh, L. A. (2008). Is there a need for fuzzy logic?. Information Sciences, 178(13), 2751-2779.

[10] Dutta, P., Boruah, H. and Ali, T. (2011) Fuzzy Arithmetic with and without using α-cut method: A Comparative Study. International Journal of Latest Trends in Computing, 2, 99-107.

[11] López, C. and Samper, J. Numerical aspects of the universal kriging method for hydrological applications. http://www.thedigitalmap.com/~carlos/papers/rep96_1/geoenv96.pdf , 31.10.2011.

[12] Lloyd, C. D. and Atkinson, P. M. (2001) Assessing uncertainty in estimates with ordinary and indicator kriging. Computers & Geosciences, 27(8), 929-937.

[13] Bohling, G. (2005) Kriging. http://people.ku.edu/~gbohling/cpe940/Kriging.pdf. 31.10.2011.

[14] Fonte, C. and Lodwick, W. (2005) Modelling the Fuzzy Spatial Extent of Geographical Entities. In: Petry, F., Robinson, V. and Cobb, M. (eds.) Fuzzy modeling with spatial information for geographic problems. Springer, Berlin.

[15] Oksanen, J. and Sarjakoski, T. (2005) Error propagation of DEM-based surface derivatives. Computers & Geosciences, 31(8), 1015-1027.

[16] Waelder, O. (2007) An application of the fuzzy theory in surface interpolation and surface deformation analysis. Fuzzy Sets and Systems, 158(14), 1535-1545.

[17] Svobodová, J. (2011) Hodnocení kvality digitálních výškových modelů pro environmentální aplikace, Ph.D. Thesis , Ostravská universita v Ostravě.

[18] DeBaets, B. and Marková-Stupňanová, A. (1997) Analytical expressions for the addition of fuzzy intervals. Fuzzy Sets and Systems, 91(2), 203-213.

[19] Chalco-Cano, Y., Jimenez-Gamero, M., et al. (2008) An approximation to the extension principle using decomposition of fuzzy intervals. Fuzzy Sets and Systems, 159(24), 3245-3258.

[20] Mesiar, R. (1997) Shape preserving additions of fuzzy intervals. Fuzzy Sets and Systems, 86(1), 73-78.