SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA

Faculty of Civil Engineering

Ing. Eva Stopková

Autoreferat of the dissertation thesis

DEVELOPMENT AND APPLICATION OF 3D ANALYTICAL FUNCTIONS IN SPATIAL ANALYSES

to obtain academical degree of Philosophiae Doctor (PhD.)

in doctoral study programme: 5.1.3 Geodesy and Cartography

Bratislava 2014

Obsah

Introduction

Although our natural perception of the world where we live in is three-dimensional, a powerful tool for modelling our reality – geospatial information systems (GIS) – has been anchored in two dimensions, in a projection of Earth's surface into horizontal plane. The very first three-dimensional thoughts appeared in late 80ties.

There is also the difference in approach to vertical coordinate. So called 2,5 GIS deals with geoobjects with position given by coordinates in planar projection. If these geoobjects are described also using height data, these data are considered as attribute values only. This approach is suitable for representing phenomena located on the Earth surface. To deal with vertically stratified phenomena as archaeological or geological trench, it is necessary to describe data position in three-dimensional space.

This approach requires special data types, theoretical background and topology – as described more detailed in following chapters, there is still no complex solution of 3D GIS, just many partial implementations into existing GIS software. One of the software products that provides support not only for vector but also volumetric analyses, is *GRASS GIS (GRASS Development Team, 2014)*. This thesis deals with development of *GRASS GIS* modules that operate with the data in three-dimensional space.

In the thesis, there is summarized basic knowledge about 3D GIS, especially theoretical background of kriging in three-dimensional space, which differs from two-dimensional solution in variogram modelling, and mathematical background of statistical method *Nearest Neighbour Analysis*. Analytical tools, modules of open-source software *GRASS GIS (GRASS Development Team, 2014)* that are developed as the main goal of the thesis are based on these topics. Description of *v.kriging* and *v.nnstat* modules is included too. Practical usage of the modules is demonstrated in several case studies with statistical tests of the results and possibilities of optimization of the modules using spatial indexes.

1 Review

3D GIS was first described in 1989 and since early 90–ties, it has been developed in several partial solutions useful in geology, landscape architecture, ecology, hydrology etc. There are several papers dealing with topic of 3D GIS architecture and most of them propose web-based solutions of 3D GIS. Another branch of authors have been dealing with topological modelling in 3D GIS.

The development of 3D GIS has been gradual because three-dimensional objects need a special approach (in geometrical, topological and graphical sense) that differs from infrastructure usually implemented in GIS. Some of current softwares support 3D data, but only those may be considered as 3D GIS which enable to perform spatial analyses on 3D data. Open-source software *GRASS GIS (GRASS Development Team, 2014)* is one of the few GIS softwares that support volume data (3D vectors too).

1.1 Spatial statistics in 3D GIS

Statistical modelling of spatial data considers not just values of given phenomenon, but also positions of points to which the values are located. The values are supposed to be function of position – they are spatially correlated. This property is very useful in estimation, prediction and design that are the most important tasks of spatial statistics. Positions are defined in d-dimensional Euclidean space, mostly in two-dimensional. Idea of extension into three-dimensional space is usually simple, but it may be complicated because it should consider naturally different behaviour of phenomena in horizontal and vertical direction.

1.1.1 Variogram

Empirical variogram shows how the data behave according to distance which is usually supposed to be two-dimensional. To deal with variogram in three-dimensional space, several approaches have been proposed. *(Myers and Journel, 1990)* describes **nested variogram** which is given as a sum of data dependence of distances in each of n dimensions (n is usually equal to 2 or 3). This solution considers isotropic variograms only. In *(Fabbri and Trevisani, 2005)*, there is similar approach to 3D variogram described, but the authors tested (highly) anisotropic 3D variogram.

Univariate variogram with anisotropic ratio was described e.g. in *(Zimmerman, 1993)*. Author summarizes these types of anisotropy in two-dimensional space (analogically, they may be useful in three-dimensional space):

- **geometric anisotropy** depends on range that varies with direction. Rose diagram should be approximately elliptic and the axes of anisotropy are represented by major and minor axis of the ellipse.

- **zonal anisotropy** depends on sill (and range), but *(Isaaks and Srivastava, 1989)* define zonal anisotropy only with sill dependent on direction.

Most of papers that describe three-dimensional variogram are connected with geological modelling. It is also possible to compute 2D and 3D variogram in *(Pebesma, 2004)* under statistical open-source software *R (R Core Team, 2013)*, but none of them is implemented in (3D) GIS environment.

1.1.2 Nearest Neighbour Analysis

Nearest Neighbour Analysis is the method of spatial statistics that helps to assess whether points located in tested area are distributed randomly, in clusters or in regular pattern in two- or three-dimensional space. Proceedings *(Hertz, 1909)* and *(Chandrasekhar, 1943, p. 86–87)* outline the principle of method in three-dimensional space and *Nearest Neighbour Analysis* for two-dimensional space was described in *(Clark and Evans, 1954)*. The method (2D) have been implemented in *ArcGIS (ESRI, 2013a)*.

1.2 Interpolations in 3D GIS

In trivariate interpolation, unknown values are estimated using function of coordinates x, y and also z that is vertical coordinate, not just attribute. Also output is three-dimensional, the voxler is derived instead of raster. Open-source *GRASS GIS (GRASS Development Team, 2014)* provides the module *v.vol.rst (Mitáš et al., 2014)* that is based on *regularized spline with tension (Mitáš and Mitášová, 1988)*.

2 Objectives

The dissertation thesis deals with 3D principles of selected spatial analyses and with development of modules for open-source software *GRASS GIS (GRASS Development Team, 2014)*. *GRASS GIS (GRASS Development Team, 2014)* has been used because it has the widest support of 3D GIS functionality among the open-source solutions, open source code and also large community of developers and contributors.

- Development of module *v.kriging*

    - Theoretical background of 3D kriging, especially 3D variogram deriving
    - Implementation into module *v.kriging*
    - Improvement of the module (using anisotropy and interactive variogram modelling)
    - Testing on various 3D data

- Development of module *v.nnstat*

    - Extension of existing theoretical background
    - Implementation into module *v.nnstat*
    - Numerical tests

- Improvements (efficiency, processing time)

Note: Part of objectives (development of module to perform *3D buffer* and implementation of fuzzy logic) have been deleted because of issues that should be emerged in *v.kriging* and *v.nnstat*.

Expected contents

- module *v.kriging* for *GRASS GIS* (GRASS Development Team, 2014) for interpolating volumes from point data using *kriging in three-dimensional space* with extensions (anisotropic functions, interactive variogram modelling etc.)

- module *v.nnstat* for *GRASS GIS* (GRASS Development Team, 2014) for performing *Nearest Neighbour Analysis* in three-dimensional space

3 Theoretical background

In this chapter, there will be explained theoretical backgrounds of two main topics of this dissertation thesis – interpolation method *kriging* and method of spatial statistics *Nearest Neighbour Analysis* in three-dimensional space.

3.1 Kriging in three-dimensional space

Kriging is interpolation method that uses geostatistical approach to estimate unknown values. Interpolation is preceded by analysis of spatial structure of the data – variogram modelling that gives information about spatial variability of the sample *(Wackernagel, 2003)*.

The theory of 3D kriging design in this thesis is based on *Ordinary kriging* in two-dimensional space described e.g. in *(Wackernagel, 2003)*.

3.1.1 Basic principle of *ordinary kriging*

Unknown value $\hat{W}_i$ located on output point i is estimated as:

$$\hat{W}_i = \sum_{j=1}^{n} w_j \cdot W_j \tag{3.1}$$

In this equation, n is number of input points and j means index of each one. Value located on each of input points is represented by W_j and it is weighted using w_j. The weights are obtained solving the system of linear equations:

$$\begin{pmatrix} \gamma(x_1 - x_1) & \dots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix} \tag{3.2}$$

The large matrix $\mathbf{G}$ on the left side contains the dissimilarities between values located on the input points. They are based on variogram (see *Chapter 1.1.1*). The ones in the last matrix row and column express the condition that sum of weights $\mathbf{w}$ (vector in the left side) should be equal to one:

$$\sum_{i=1}^{n} w_i = 1 \tag{3.3}$$

This condition makes the differences of the input and interpolated value on the input point to be equal to zero. Constant μ^{OK} at the bottom of the $\mathbf{w}$ vector means Lagrange multiplier to attach the condition (*Eq. 3.3*) to the system of linear equations.

To compute the dissimilarities between input value and unknown value on the right side, it is necessary to solve the system of linear equations at first:

$$\mathbf{w} = \mathbf{G}^{-1} \cdot \mathbf{g_0} \tag{3.4}$$

Interpolation using *kriging* method usually consists of these steps:

- **Computing of experimental and theoretical variogram** Matrix of dissimilarities at input points $\mathbf{G}$ is derived. The dissimilarities between input and output points $\mathbf{g_0}$ are estimated to solve system of the equations (*Eq. 3.2*). Vector $\mathbf{g_0}$ is unknown in this phase of computing and we estimate its values using variogram.

  Variogram $\gamma(\mathbf{h})$ expresses dependence of change in values of random variable and of the distance between the pairs of the points according to equation *(Wackernagel, 2003)*:

$$\gamma(\mathbf{h}) = 0.5 \cdot (z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha))^2 \tag{3.5}$$

  $\mathbf{x}_\alpha$ means the position of the point in space, $\mathbf{h}$ is the vector connecting the pair of the points and $\gamma(\mathbf{h})$ expresses dissimilarity between values located on the points connected by vector $\mathbf{h}$. If the distances between points that are included in the vector $\mathbf{h}$, are split into k classes that are characterized by vector length of $\mathbf{h_k}$, average values of $\gamma_k(\mathbf{h})$, which are computed from $\gamma(\mathbf{h})$ characterizing the given interval $\mathbf{h_k}$, create *experimental variogram*.

  Function that approximates discrete values of experimental variogram is called *theoretical variogram*. Most commonly used and described types[1] are linear, exponential, spherical and Gaussian *(Wackernagel, 2003)*. Their shape depends on *sill* (the limit value of γ at higher values of h), *range* (distance that determines the zone of spatial correlation in the regionalized variable *(Omre, 1984)*) and *nugget effect* (the degree of discontinuity in the regionalized variable *(Omre, 1984)*).

  If experimental variogram rises after reaching sill and the fitting function is parabolic, this indicates that the values of modelled phenomenon are spatially correlated and there is some trend in the data. It is necessary to eliminate the trend and perform variogram modelling just using the residuals.

- **Estimation of optimal weights of input points** according to system of equations (*Eq. 3.2*) using dissimilarities approximated by theoretical variogram.

- **Estimation of unknown value** located on output point according to *Eq. 3.1*. Then estimation of next interpolated values continues.

The essential difference between interpolation using *ordinary kriging* in two- and three-dimensional space will be in variogram computing. In 2D space, it is function of horizontal distances between pair of the points only, but in 3D also vertical distance must be considered. Computation of $\mathbf{G}$ matrix and of $\mathbf{g_0}$ vector depends on variogram, but the principle is the same for both cases.

---

[1]If the curve of theoretical variogram is parabolic, this indicates that the values of modelled phenomena are strongly spatially correlated – parabolic variogram shows some trend in the data that should be also eliminated in further processing.

3.1.2 Variogram modelling in three-dimensional space

These main approaches to compute 3D variogram have been described:

- **nested variogram** according to *(Myers and Journel, 1990)*:

$$\gamma(x, y) = \gamma_1(x) + \gamma_2(y) = \gamma(h, \Delta z) = \gamma_1(h) + \gamma_2(\Delta z) \tag{3.6}$$

where $\gamma_1(x), \gamma_2(y)$ are isotropic variograms in 2D space, but we can assume that x represents horizontal distance and y can be vertical distance.

- **anisotropic variogram** according to *(Facas et al., 2010)*:

$$\gamma(x, y) = \gamma\left(\sqrt{x^2 + (R \cdot y)^2}\right) \tag{3.7}$$

where $R = a/b$ – ratio of maximum range to minimum range. This approach is useful in case of *geometrical anisotropy*, when just ranges differ in horizontal and vertical direction. If also sills are different, it indicates so-called *zonal anisotropy* and according to e.g. *(Gringarten and Deutsch, 2001)*, it is useful to use nested variogram in such a situation (*Eq. 3.6*).

3.2 Nearest Neighbour Analysis in three-dimensional space

Nearest Neighbour Analysis indicates randomness of points distribution in a sample. It can be useful in natural sciences or in solving social problems (e.g. crime analysis). If vertically divided phenomenon is analysed (e.g. artifacts in archaeological trench), three-dimensional distance should be considered in *Nearest Neighbour Analysis*.

There are N points located in the area of interest. Their average isotropic distance between the nearest neighbours $\overline{r}_A$ is expressed using arithmetic mean. If the points are randomly distributed, average isotropic distance of the nearest neighbours in three-dimensional space is expected to be (according to *(Chandrasekhar, 1943, p. 86–87)*):

$$\overline{r}_E = \int_0^\infty 4\pi n \cdot r^3 \cdot e^{-\frac{4\pi n}{3} \cdot r^3} dr = \frac{1}{\sqrt[3]{\frac{4\pi n}{3}}} \int_0^\infty e^{-x} \cdot \sqrt[3]{x}\, dx = \frac{1}{\sqrt[3]{\frac{4\pi n}{3}}} \cdot \Gamma(\frac{4}{3}) \tag{3.8}$$

where n is density of points in volume unit and $\Gamma(\frac{4}{3})$ is the gamma function[2]. This formula is based on the probability that the nearest neighbour of a point is located on the boundary of its spherical surrounding with radius r. The measure of degree to which observed points are randomly distributed R is expressed as the ratio of observed and expected distance *(Clark and Evans, 1954)* and it should be statistically tested.

According to *(Clark and Evans, 1954)*, statistical significance of the deviation from a random distribution of points (clustering or separation) can be tested using Student's t-test of significance of the mean *(Hughes and Hase, 2010)*. There is standard variate of the normal curve as test statistics and it depends on standard deviation of the average distance expected

---

[2] *Gamma function* $\Gamma(n)$ means extension of factorial for complex and real numbers *(Weisstein, 2013)*.

in case of randomly distributed points $\sigma_{\overline{r}_E}$. This value was derived in past for two-dimensional space *(Clark and Evans, 1954)*, but not for three-dimensional space. 3D solution has been derived analogically *(Stopková, 2013)*.

Null hypothesis H_0: $\overline{r}_A = \overline{r}_E$ means that points are randomly distributed in the space (according to *(ESRI, 2012)*). Alternative hypothesis H_a: $\overline{r}_A < \overline{r}_E \vee \overline{r}_A > \overline{r}_E$ and statistical tables of the cumulative standard normal distribution are base for determination of critical values $W = (-\infty, u_{\alpha/2}) \cup (u_{(1-\alpha/2)}, \infty)$. If null hypothesis is rejected with positive values of the test statistic, the points are assumed to be separated (patterned), otherwise (rejection with negative values) the points are assumed to be clustered.

4   Implementation

In this chapter, the modules for open-source software *GRASS GIS (GRASS Development Team, 2014)* will be described. There will be shown also results of case studies.

4.1   Kriging in 3D space

The structure of the module is shown here and also possibilities of variogram modelling are outlined in this chapter. Issues that emerged in module testing are described together with proposed solutions.

4.1.1   Module *v.kriging* in GRASS GIS

Interpolation was divided into three phases because of anisotropy detection that needs also interaction of the user:

- **initial phase** provides variogram in horizontal and vertical direction. In this phase, basic properties of experimental variogram are computed that are written to temporary text files for usage in following phases.

- **middle phase** is based on properties of experimental variograms in horizontal and vertical direction. The user should empirically set up ranges of both variograms (sills have been computed in initial phase and nugget effect is not important here).

   According to sills, the type of variogram function is chosen if not set up by the user. If sills difference is greater than 5% of greater of the sills, bivariate function should be used because of zonal anisotropy in the dataset. Otherwise, anisotropic function is appropriate to model just geometric anisotropy. Anisotropy ratio is derived as a ratio of ranges in horizontal and vertical direction. There may be chosen type of function (univariate or bivariate) according to sill ratio or user's opinion. Output shows experimental variogram anisotropic or bivariate.

- **final phase**. The user should set up parameters of anisotropic variogram (nugget effect and range – sill has been computed in initial phase, but it may be changed too). Parameters of bivariate function are computed using *Least Mean Square* method. Then, theoretical variogram is created and the values are interpolated.

Here may be seen the basic structure of the code:

- **graphical user interface** with options for:

  - **performing 2D, 2,5D or 3D interpolation**. This depends on user's decision and on the input layer.

  - **variogram settings**. The type of univariate or bivariate function belongs to the most important options here. This influences choice of variogram type and other parameters (coefficients of the function, number and size of lags and maximum distances). All these parameters are just optional, because the default values are derived later.

  - **outputs** with additional information about the interpolated data:
    * **variogram plot** in Gnuplot window or in various file formats,
    * **report file** with statistical tests of output data,
    * **cross validation results** that include differences of original and interpolated values located on the input points.

- **parameters of the region** are obtained. This is necessary for defining output (2D or 3D) raster extent and resolution and also for filtering points outside the region.

- **coordinates of input points** within the region are obtained.

- **values to be interpolated** are obtained from attribute table of vector layer. The function to get values from database is based on part of the module *v.buffer2 (GRASS Development Team, 2014)*.

- **variogram restrictions** are computed (maximum distance of variogram computation, size and number of lags). The user may set up also his/her own variogram settings, but at least on first performance of the interpolation, it is convenient to use the default settings.

- **experimental variogram** is estimated. The computation depends on required variogram function – bivariate function is based on *Eq. 3.6* and univariate function of any type uses in general *Eq. 3.5*.

- **anisotropy** computation is based on empirical estimation of variogram ranges in horizontal and vertical direction. In case that sills in both directions are different, also zonal anisotropy must be considered.

- **theoretical variogram** based on the experimental variogram is estimated. There are several types of functions available to model theoretical variogram.

- unknown values are estimated using the method of **ordinary kriging** and **cross validation** is performed optionally. The values located on the input points are interpolated using modelled variogram and compared with original values.

In special case that all the values located on input points are the same (coefficients of linear functions are equal to zero), constant 2D or 3D raster is created.

4.1.2   Case studies

To verify the results of *v.kriging* several datasets have been used:

- **synthetic data**. Testing of the data based on exact mathematical formulae (normal gravity) should not be influenced by error of measurement, physical properties of the environment and by other random factors.

- **the data acquired in real world** that describe the phenomenon including its random nuances. In this thesis, soil moisture data *(White, 2002)* and space-time cube that represents terrain evolution *(Mitášová, 2014)* in time were used.

Each case study is based on very similar methodology. Horizontal and vertical experimental variogram were computed to estimate the range, which shows geometrical anisotropy, and the sill that may indicate zonal anisotropy. In this experiment, both types of variogram were computed because of comparison. Parameters of anisotropic theoretical variogram were estimated empirically, coefficients of bivariate linear variogram were derived using method Least Mean Squares (LMS).

Interpolation results were verified comparing cross validation results and also differences between reference and interpolated values. Cross validation was performed on input points – original values of normal gravity were compared with values interpolated on the same points. Condition *Eq. 3.3* should make the differences to be close to the zero (because of data imperfection and other random errors).

Reference values to be compared with interpolation result were derived using physical model or as results of *Regularized Spline with Tension* interpolation (by *v.vol.rst (Mitáš et al., 2014)*). The differences between 3D rasters were obtained using the module *r3.mapcalc (GRASS Development Team, 2014)* and statistical characteristics were compared.

Normal gravity   The values were derived using physical model according to *(Hoffmann-Wellenhof and Moritz, 2006)*. Parabolic curve of experimental variogram in vertical direction indicated trend in the data that was eliminated before interpolation using kriging method. Following polynomial expression seems to be the most appropriate:

$$w = a \cdot x + b \cdot z + c \cdot x \cdot y \cdot z + d \cdot x \cdot y + e \cdot x \cdot z + f \cdot y \cdot z + g \qquad (4.1)$$

where x, y, z are coordinates of the points and w means the values located on these points. In *Fig. 4.1* and *Fig. 4.2*, there might be seen experimental and theoretical variograms computed for residuals in horizontal and vertical direction. The variogram in vertical direction (*Fig. 4.2*) may still indicate some trend, as it rises even after reaching the sill.
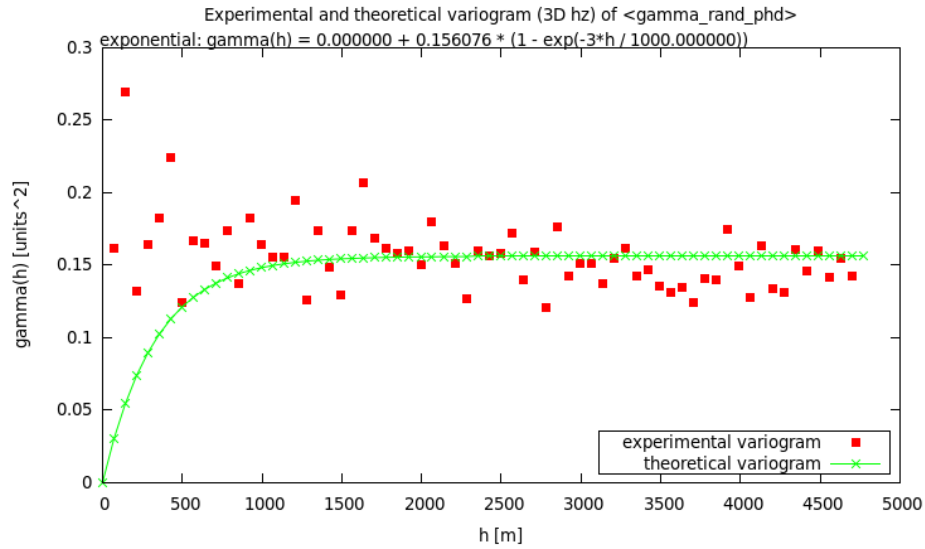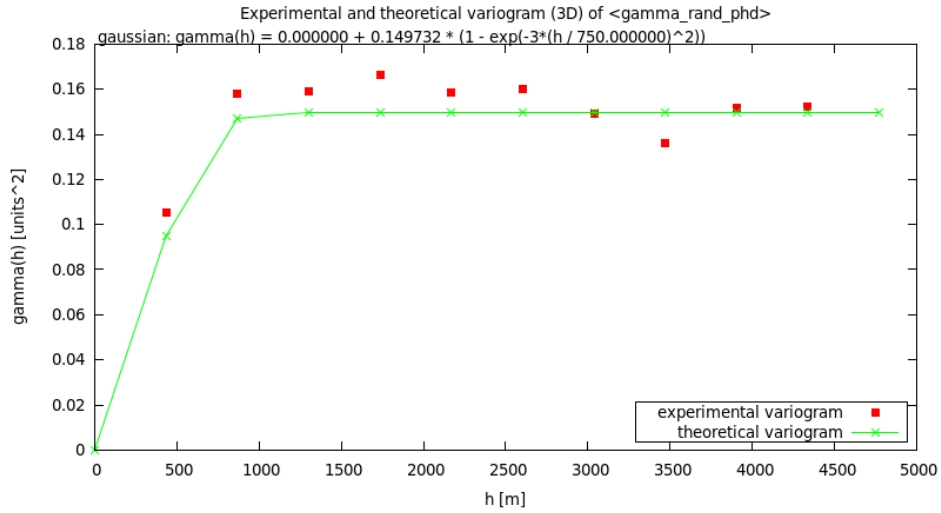

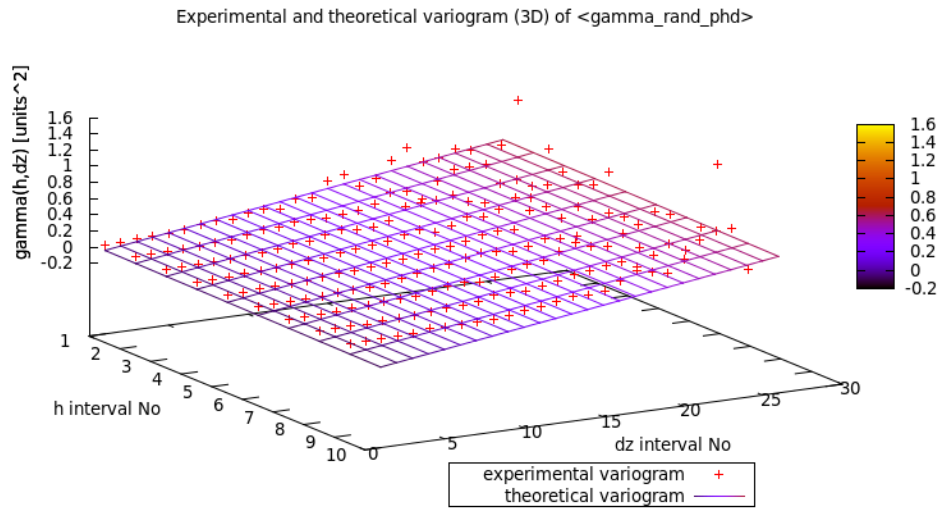Obr. 4.1: Normal gravity (after eliminating trend) – horizontal variogram
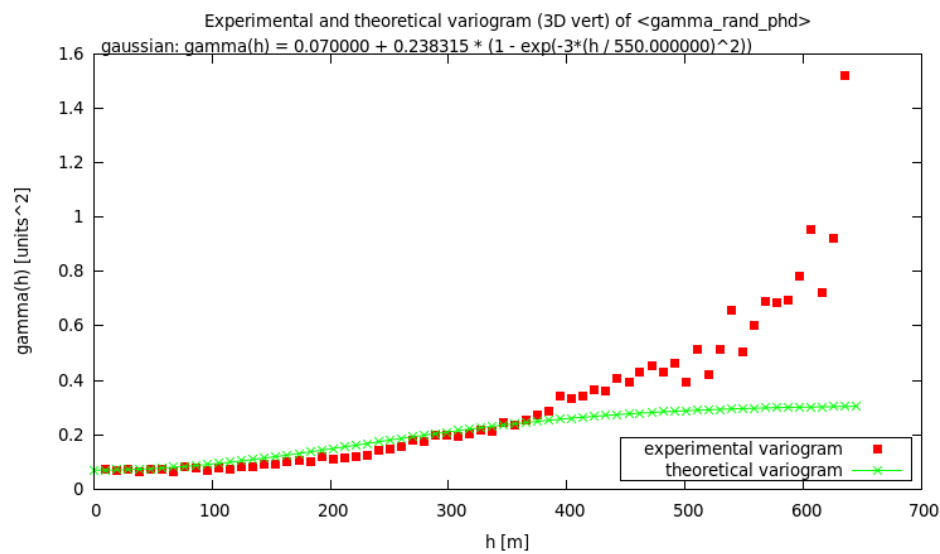
Obr. 4.2: Normal gravity (after eliminating trend) – vertical variogram

The difference between sills in horizontal and vertical direction indicates also zonal anisotropy.

Obr. 4.3: Normal gravity (after eliminating trend) – anisotropic variogram

Obr. 4.4: Normal gravity (after eliminating trend) – bivariate variogram

According to statistical characteristics of cross validation results, Gaussian variogram (*Fig. 4.3*) seems to be the most appropriate to model this dataset. Although bivariate variogram (*Fig. 4.4*) was expected to fit the data better, linear function is not the most appropriate to model this dataset. It generates negative values of squared dissimilarities located on the pairs of the points and also the plane does not respect curvature of the experimental bivariate variogram very well.

Comparison of reference and interpolated values indicates that spherical variogram could be the most appropriate for modelling this dataset. As the range of theoretical variogram is quite small, there are no significant differences in shape of the Gaussian and spherical curve while it reaches the sill sharply.

In *Fig. 4.5*, there can be seen 3D raster interpolated using kriging with Gaussian variogram as function of distance in three-dimensional space, where horizontal and vertical direction are anisotropic. The values are visualized using almost parallel isosurfaces – they

11

represent equipotential planes that are in fact sections of ellipsoids.

The differences between values obtained using physical model and values interpolated using Gaussian variogram in *Fig. 4.6* show some systematic error in the result.
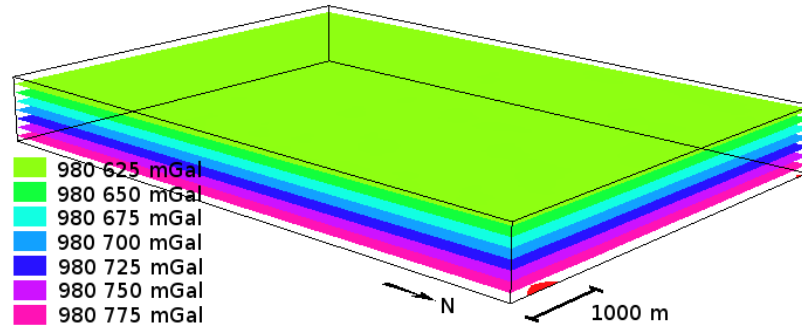
Obr. 4.5: Normal gravity (kriging based on Gaussian variogram)
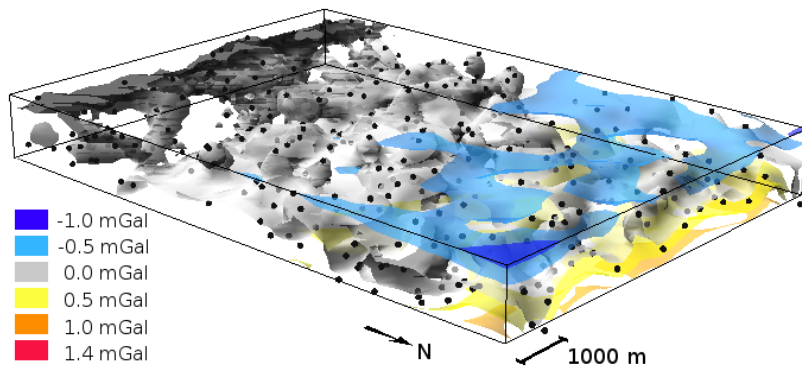
Obr. 4.6: Normal gravity (differences between physical model and kriging)

The difference between real and interpolated values (based on Gaussian variogram) is less than 1.5 mGal (in absolute value) – original values vary around 980 702 mGal. In *Fig. 4.6* might be seen that the errors are not distributed randomly. In northern part of the study area, interpolated values have wider range than original dataset. This trend might be caused by coordinate transformation (normal gravity was computed for ellipsoidal coordinates in ETRS89 and then the points were transformed to S-JTSK) or there might be also some residual trend.

Space-time cube   *(Mitášová, 2014)* represents terrain evolution in time. As vertical co-ordinate, we consider time t in this case[3], and values of z coordinate were interpolated as function of horizontal coordinates and time $z = f(x, y, t)$.

As sill of horizontal and vertical variogram indicates (*Fig. 4.7, 4.8*), bivariate variogram should be more appropriate to model dissimilarities in this dataset. There may be also some trend in the data, because horizontal variogram rises quite regularly after reaching the sill. Anyway, curve of this experimental variogram seems to be not too regular as in case of normal gravity, so for this experiment we did not eliminate trend.

[3]One unit of vertical axis represents 100 days
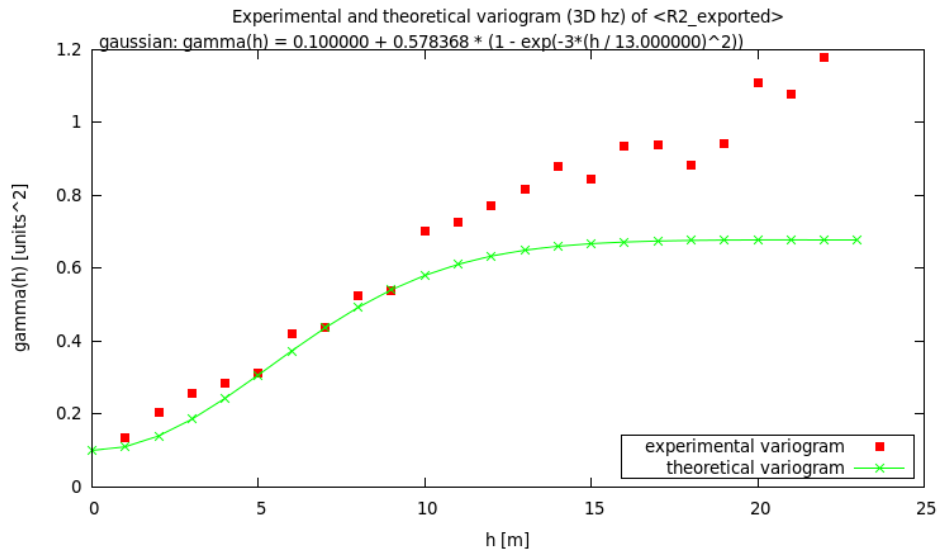
Obr. 4.7: Space–time cube – horizontal variogram
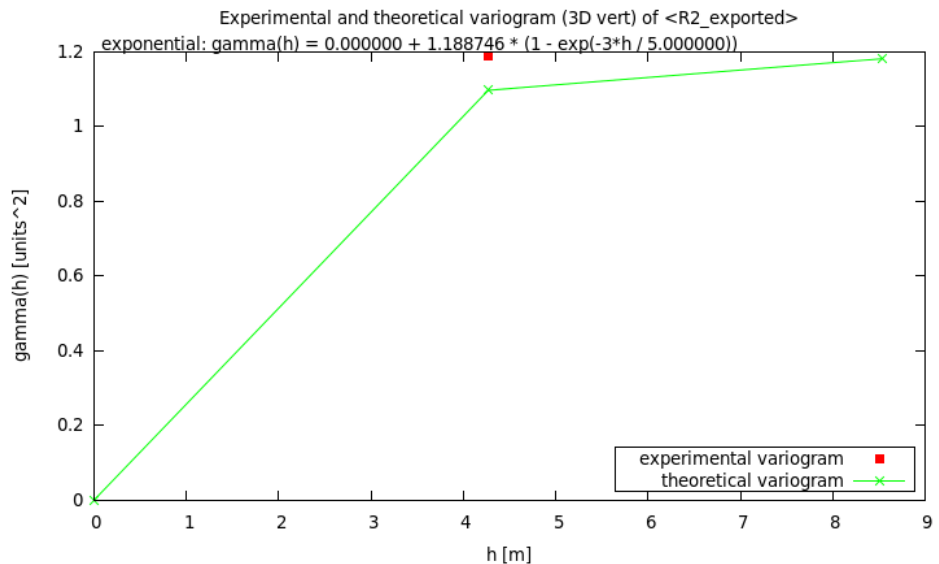

Obr. 4.8: Space–time cube – vertical variogram

Another tricky phenomenon is represented in *Fig. 4.8*. As the values of elevation have been acquired in within constant time intervals and the sample is quite thin (just three levels), there is only one vertical lag with real value. In this case, we tested just setting up range to make the curve of theoretical variogram fit the only value of experimental variogram.

Comparing cross validation results and differences between *RST* and *kriging*, exponential variogram (*Fig. 4.9*) was assumed to be optimal for this dataset. Bivariate variogram seems to be not appropriate to model this dataset. This may be caused by too sparse configuration of points along vertical axis and also constant step might spoil the result, despite of the function that fits experimental variogram quite well (*Fig. 4.10*).
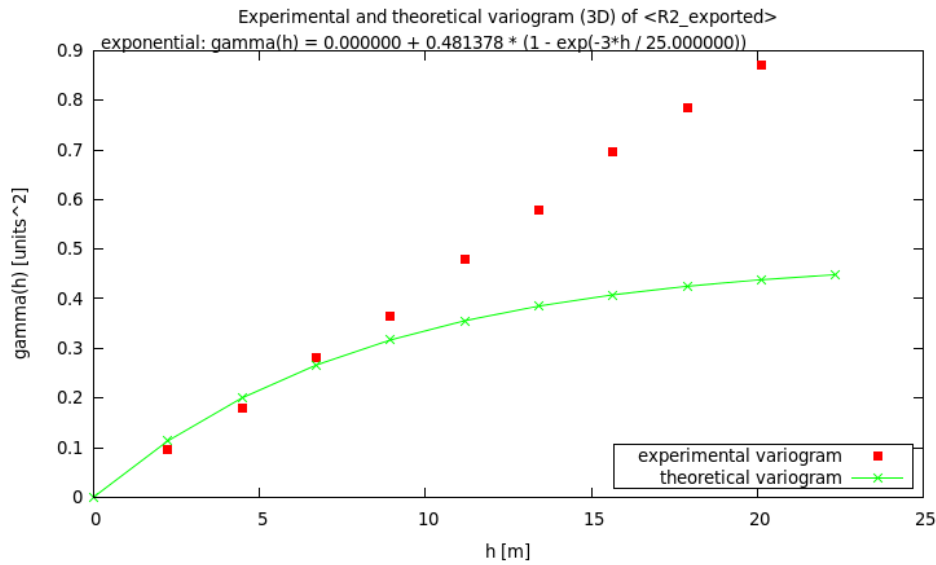
Obr. 4.9: Space–time cube – anisotropic exponential variogram

The results of *RST* seem to be smoother than results of kriging and also than input values – this is visible also on horizontal slices. This may depend on settings of tension and smoothing of spline function in *RST* method. Parameters of *RST* function have been chosen according to *(Mitášová, 2014)*.
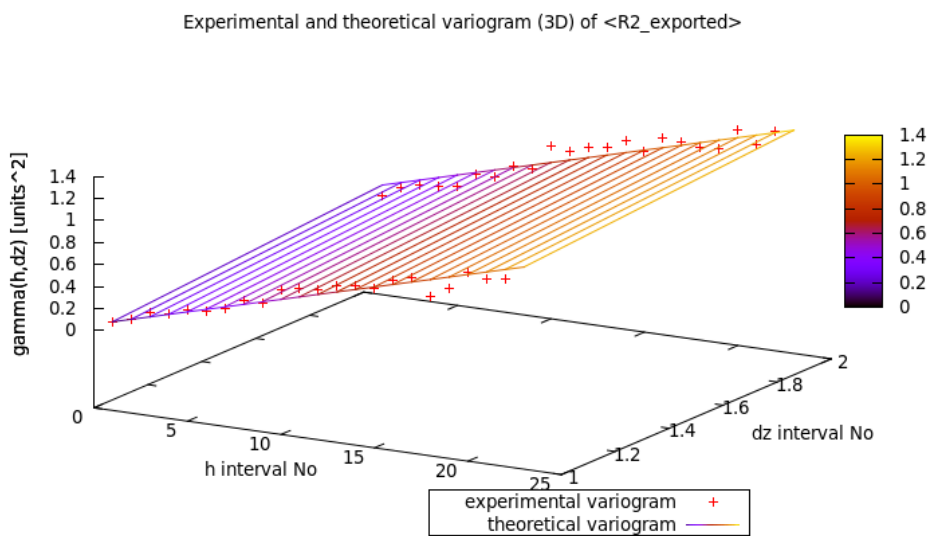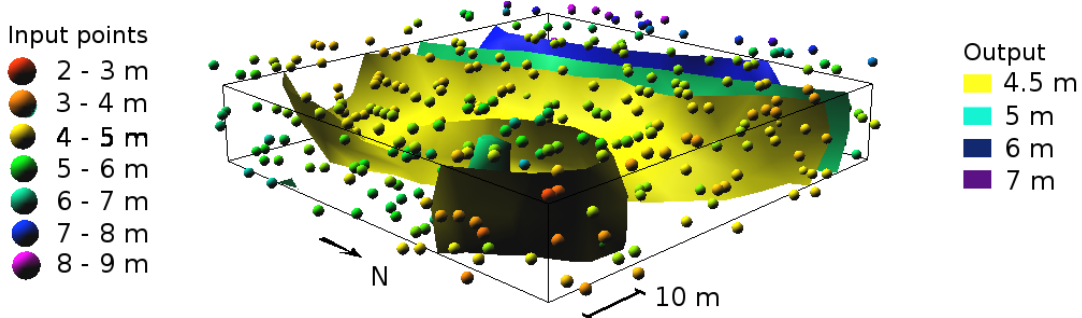

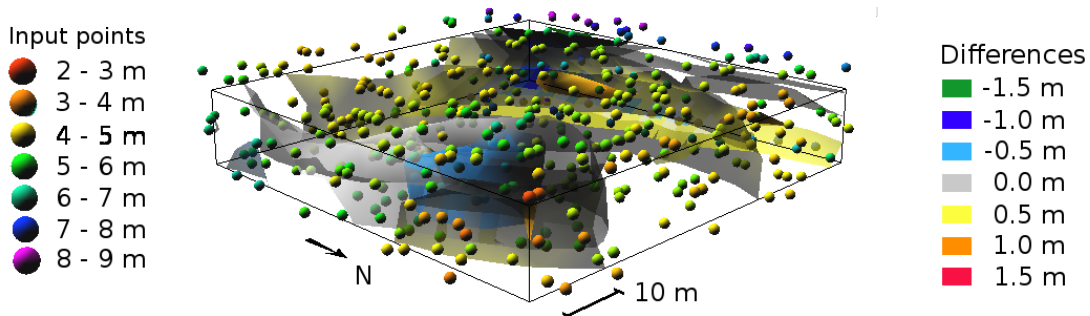Obr. 4.10: Space–time cube – bivariate variogram
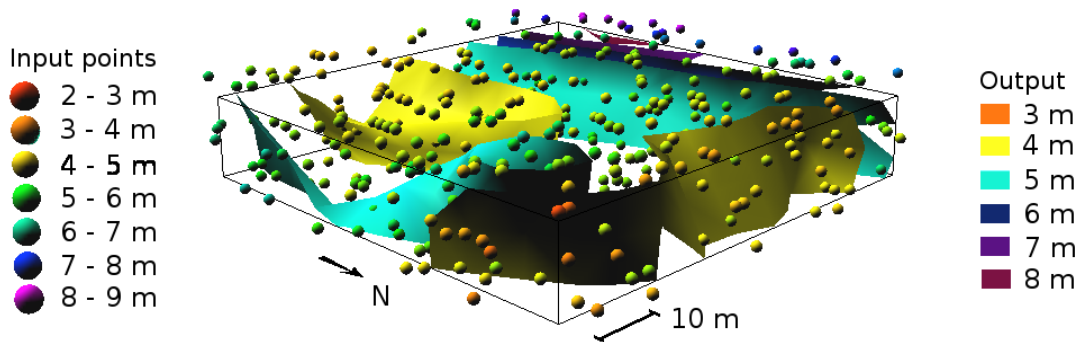
Obr. 4.11: Space-time cube (kriging based on exponential variogram)

Obr. 4.12: Space-time cube (Regularized Spline with Tension)

Obr. 4.13: Space-time cube (differences between RST and kriging)

The differences are quite high – does not matter about variogram function, but comparing with statistical characteristics of input data it might be assumed that the result of RST interpolation is smoother than original values. This is visible also in *Fig. 4.11* and *4.12*.

Soil moisture data *(White, 2002)* consist of *time domain reflectometry* (TDR) readings of volumetric soil moisture (in %) located on coordinates x, y and depth. Horizontal and vertical variogram (see *Fig. 4.14, 4.15*) indicate that there is no zonal anisotropy in this dataset.
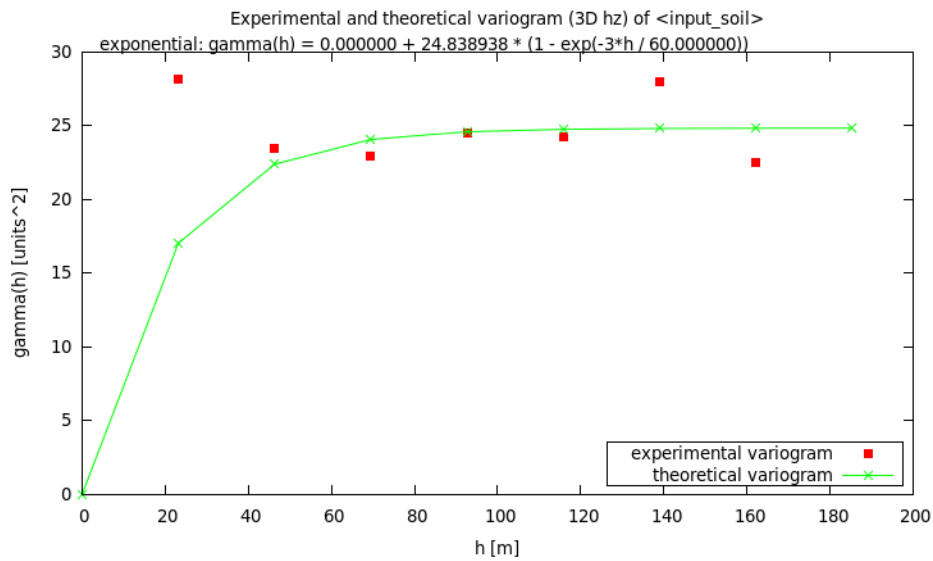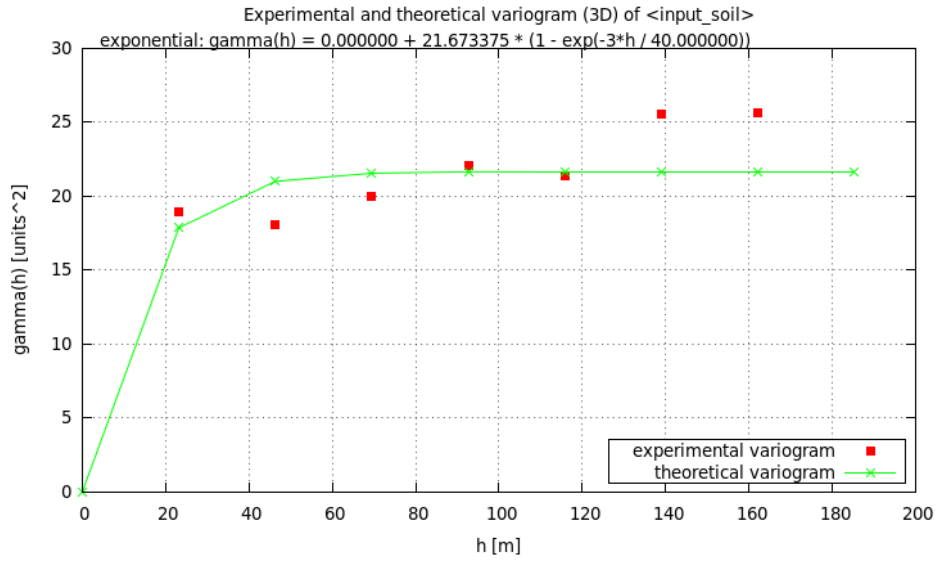
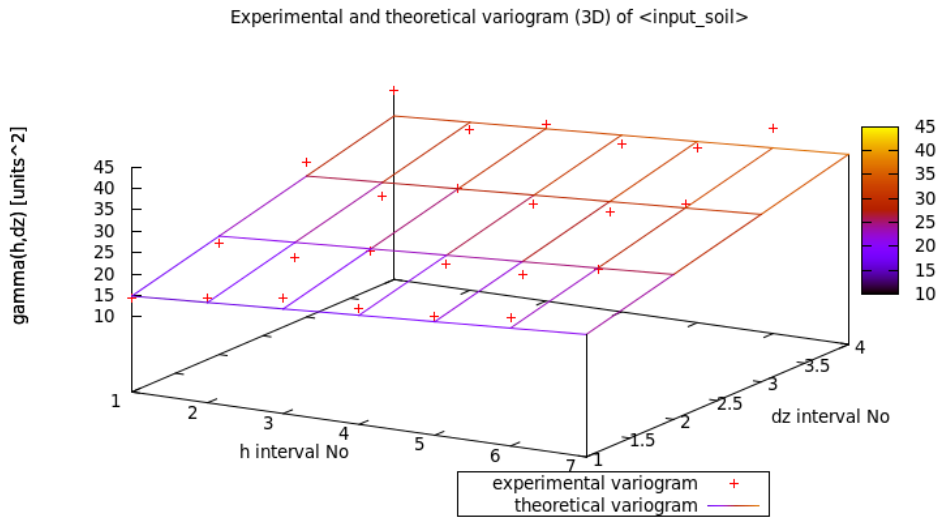Obr. 4.14: Soil moisture – horizontal variogram
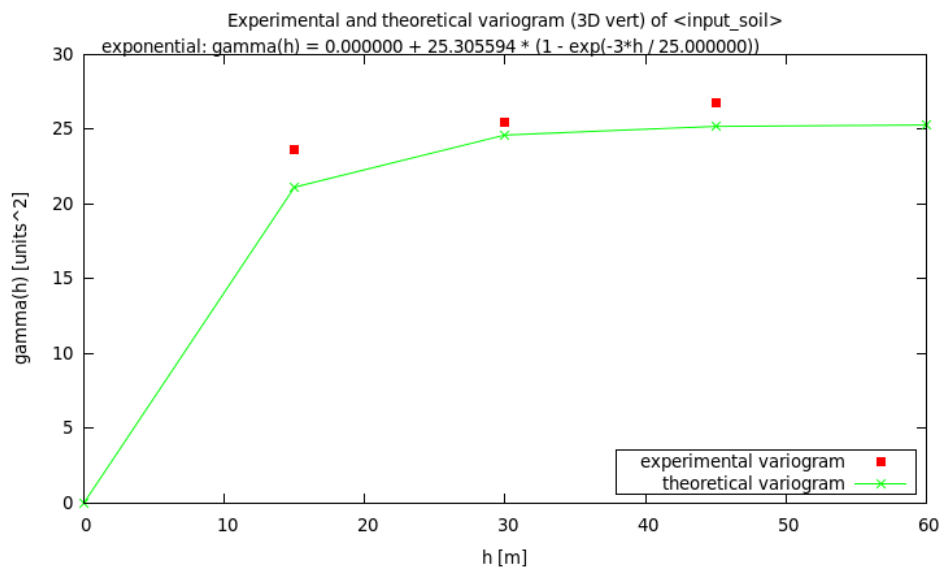
Obr. 4.15: Soil moisture – vertical variogram

It is correct then to model the dissimilarities using just anisotropic variogram (*Fig. 4.16*), but experimentally also bivariate variogram (*Fig. 4.17*) was computed.

Obr. 4.16: Soil moisture – anisotropic variogram

Obr. 4.17: Soil moisture – bivariate variogram

According to statistical characteristics of cross validation, Gaussian variogram (see *Fig. 4.16*) seems to be the most appropriate to model this dataset. On the other side, bivariate variogram (*Fig. 4.17*) causes extremely high values of differences between original and interpolated values. Linear bivariate function fits experimental variogram quite well (even better comparing with function in *Fig. 4.4*). The reason why interpolated values are so incorrect may be spatial variability of input values. Soil moisture is much more variable phenomenon than normal gravity and it probably requires different approach (maybe segmentation of study area for variogram modelling purposes).

Again, comparing differences of interpolation methods, Gaussian variogram (*Fig. 4.18*) seem to generate very close values to the results of *RST* (*Fig. 4.19*). Exponential and spherical function are probably too steep to model this dataset. In *Fig. 4.20* might be seen

that the highest values are grouped in surrounding of points that are characterised by value significantly different comparing with other values in neighbourhood.
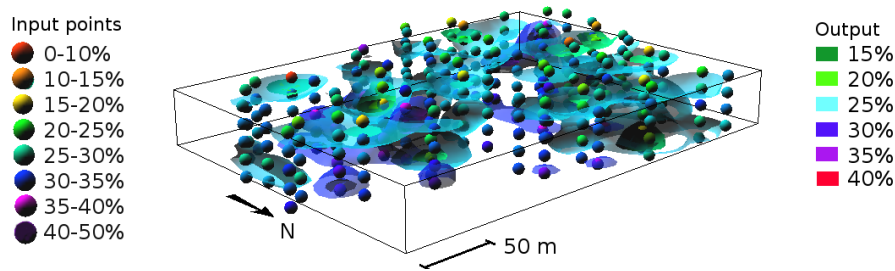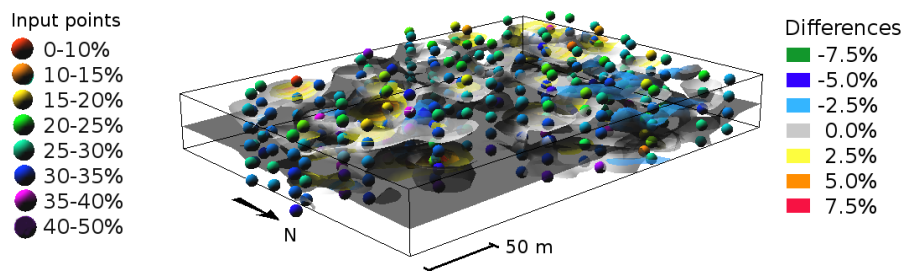

Obr. 4.18: Soil moisture (kriging based on Gaussian variogram)
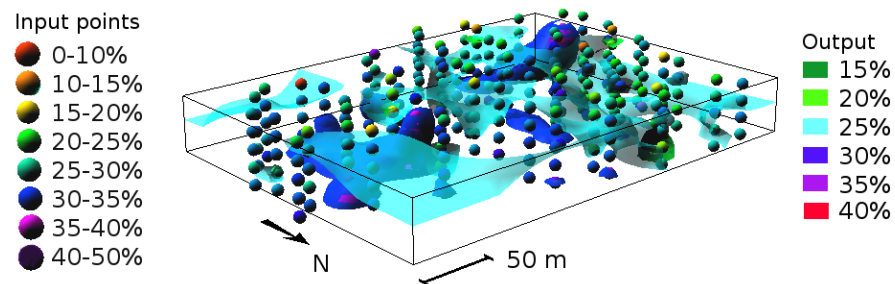
Obr. 4.19: Soil moisture (Regularized Spline with Tension)

Obr. 4.20: Soil moisture (differences between RST and kriging)

Comparison with *RST* shows that Gaussian function generates too smooth values – soil moisture is quite variable and it is distributed in space quite randomly. Spline functions are able to better fit values that are changing quite steeply.

4.2   Nearest Neighbour Analysis

The purpose of this experiment is to outline the way how to implement *Nearest Neighbour Analysis* in three-dimensional space into the GIS environment.

4.2.1   Module *v.nnstat* in GRASS GIS

The module contains functionality of NNA in two-dimensional space (implemented also in the analytical tool *(ESRI, 2013b)* of *ArcGIS (ESRI, 2013a)*) and solution for objects in three-dimensional space. The most significant complications during development were connected with determining of Minimum Bounding Box (MBB). MBB determines the density

of points in three-dimensional space (or the volume may be set up by the user). Methods how area (volume) of MBR (MBB) can be determined are principally quite similar:

– <u>Coordinates of convex hull</u> must be obtain. In 3D case it is necessary to know also reference of each vertex to the faces. Partially modified functions of the module *v.hull (GRASS Development Team, 2014)* that enables to build convex hull in new vector layer were used.

– The coordinates of vertices must be <u>transformed</u> to coordinate systems, which x axes are parallel to lines between neighbouring vertices in 2D case. In 3D case, their xy planes are parallel to planes of faces of the convex hull.

– <u>Extent of transformed coordinates</u> should be determined,

– <u>Area / volume of the extent</u> must be counted and the values should be compared to obtain the smallest one. This value becomes input in determining the density of points.

4.2.2 Case studies

For points located in two-dimensional space, the module was tested comparing numerical results with outputs of the analytical tool in *ArcGIS 10.1 (ESRI, 2013a), Average Nearest Neighbor (ESRI, 2013b)*. 3D variant of NNA is not implemented in any of accessible softwares, numerical results were verified comparing them with values computed by scripts in the software *Mathematica (Wolfram Research, Inc., 2008)* and *Matlab (MathWorks, Inc., 2010)*.

The module was tested on various sets of synthetic data. Configuration of observed points in the area of interest 20 km x 20 km was designed to represent all possible cases: randomly distributed points, separated points and clustered points. The samples of randomly distributed points were generated in the software *GRASS GIS (GRASS Development Team, 2014)* using the module *v.random (GRASS Development Team, 2014)*. Except numerical accuracy, process speed was tested too. Here is an example (*Tab. 4.2.1*) from testing:

		n = 5000		
	v.nnstat	*Average Nearest Neighbor (ESRI, 2013b)*	difference	
$\bar{r}_A[m]$	140.981861	140.981861	0.000	mm
$\bar{r}_E[m]$	141.307761	141.307761	0.000	mm
R	0.997694	0.997694	0.000000	
c	-0.311984	-0.311986	-0.000002	
$A[m^2]$	399357668.904297	399357668.904439	142	mm^2
$t[s]$	0.324	5		

Tabuľka 4.2.1: Test of *v.nnstat* in 2D space – 5000 randomly distributed points

According to the test statistic $c \in (-1.96; 1.96)$, it can be assumed that in both cases the points were randomly distributed. Differences in the values of MBR area may be caused by

the different way of storing data in computer memory, as the results of other experiments show.

Nowadays, there is no accessible tool for process NNA in three-dimensional space, so it is not possible to compare the results of the module with outputs of any verified software. That is the reason to control the results using parts of code scripted in mathematical software *Mathematica (Wolfram Research, Inc., 2008)* or *Matlab (MathWorks, Inc., 2010)*. Just results of volume of Minimum Bounding Box (MBB) were verified.

4.3 Optimization methods

This chapter is based on the article *(Stopková, 2014)*. Simple tasks of searching points are often quite compute intensive and time-consuming, especially for large datasets. That is the reason why spatial index for three-dimensional data (*kd-tree*) implemented in *PCL* library *(Rusu and Cousins, 2011)* was tested.

4.3.1 Spatial indexing

The nearest neighbour of observed point can be simply found computing all the distances between all the pairs of the points, but this way is too inefficient. If the sample contains n points, computation of all the distances means $0.5 \cdot n \cdot (n-1)$ operations. Moreover, if it is necessary to interpolate unknown values in 3D raster of $(n_x \cdot n_y \cdot n_z)$ cells, where n_x means number of columns, n_y means number of rows and n_z means number of depths (or vertical levels), number of operations during search of the relevant surrounding will increase to $n \cdot (n_x \cdot n_y \cdot n_z)$ because of need to compute distance of reference point of each cell to all the input points. Spatial indexing *(Rusu and Cousins, 2011)* was used to improve these tasks:

- nearest neighbour search for computing **average minimum distance between points** in *Nearest Neighbour Analysis* (PCL function *nearestKsearch*)

- **search of points within relevant distance to observed point** for variogram modelling (PCL function *radiusSearch* following additional conditions because of anisotropy)

- **estimate optimal lag distance** (PCL function *nearestKSearch* with k equal to 5% of total number of the points)

These functions return vector of indices of the nearest neighbour(s) and also vector of squared distances between point pairs. However, these distances may be not useful in next steps, because *PCL* library *(Rusu and Cousins, 2011)* uses float precision only *(PCL Users mailing list, 2012)*. Because of accuracy of results it is better to use just vector of indexes to select relevant coordinates of original points.

$\gamma(h, \Delta z) = a \cdot \Delta z^2 + b \cdot h^2 + c$	a	b	c
without kd-tree	0.043714	0.000020	-387.254400
with kd-tree	0.043301	0.000026	-499.359364

Tabuľka 4.3.1: Theoretical variogram coefficients

4.3.2 Case studies

Both of the modules were tested comparing the results of version with and without using *kd-tree*. The most interesting results have been obtained in comparing results and processing time in *Nearest Neighbour Analysis* of the sample of 211 631 points of gravity mapping *(Kubeš et al., 2001)*. Module *v.nnstat* was tested on computer with parameters as follows:

Processor:	Intel(R) Core(TM) i5 CPU, 2.80 GHz, 4 cores
Memory:	4.00 GB
Operation system:	Ubuntu 12.04

With acceleration of processing time from 16 to 9.5 minutes, spatial index *kd-tree* seems to be really useful. Anyway, this module needs to speed up processing of *Minimum Bounding Box* searching.

To test the module *v.kriging*, interpolation was performed on input points with normal gravity γ. This sample of points was selected from gravity mapping *(Kubeš et al., 2001)*.

In statistical characteristics of interpolated normal gravity values, there were noticeable numerical differences. We suppose that they were caused by experimental variogram $\gamma(h, \Delta z)$ that was determined using sample of points selected by *kd-tree* due to distance between pairs of points. This sample was probably slightly different of the sample from selection in original module because of precision (*float* vs. *double*). This changed also coefficients of theoretical variogram (see *Tab. 4.3.1*). Variable of h means horizontal distance and Δz means vertical distance between pairs of points. It is not correct to use parabolic variogram, but this was older version of the module tested. Actual version used spatial indexes.

Spatial index in *v.kriging* is helpful just in case of larger samples of points. If the sample is too small, then interpolation should be performed without spatial index – as it is necessary to create new spatial index for each output point, it slows down the process. In other parts of kriging, e.g. variogram modelling or estimation of optimal lag distance, spatial index is quite helpful.

Conclusion

The development of analytical tools for 3D GIS included **summation or extension of theoretical background** of selected spatial analyses in three-dimensional space, **finding optimal way** how to implement these principles, making **case studies** to verify the results.

The main contribution of this thesis are two modules:

- **v.kriging** that provides functionality to interpolate data using geostatistical approach. The module implements univariate anisotropic variogram and bivariate variogram (although this needs additional research). Interpolation using kriging may be alternative to the methods that have been implemented in GIS already (e.g. *Regularized Spline with Tension* in module *v.vol.rst (Mitáš et al., 2014)*). Variogram modelling is useful not just for interpolation but also as information about spatial dependencies between the data.

- **v.nnstat** that helps to assume if the points are distributed in three-dimensional space randomly, in clusters or in patterns. The module provides two-dimensional functionality that may be found e.g. in analytical tool *Average Nearest Neighbour* in *ArcGIS (ESRI, 2013a)*. More important is three-dimensional functionality that has been described, but not implemented in GIS before. For purposes of GIS usage, computation of standard deviation in three-dimensional space was derived.

The modules are integrated in open-source *GRASS GIS (GRASS Development Team, 2014)* to be free and accessible to anybody in the world to use, modify or extend if necessary. As modules of well-known open-source software that is used by (not only) scientific communities all over the world, they have a chance to be used more often as if they were part of some not very commonly used software. They may be also starting point for other research – this probably will not be possible if the modules were part of proprietary software.

Résumé

Dizertačná práca poskytuje prehľad súčasného stavu 3D GIS s dôrazom na teoretické princípy a praktické aplikácie vybraných 3D priestorových analýz. Hlavným cieľom bolo vytvoriť analytické nástroje na prácu s trojrozmernými dátami – moduly pre open-source softvér *GRASS GIS (GRASS Development Team, 2014)*.

Modul *v.kriging* umožňuje interpolovať dáta na základe modelovania zmien ich hodnôt pomocou 3D variogramu, ktorý môže byť funkciou anizotropnej trojrozmernej vzdialenosti alebo horizontálnej vzdialenosti a prevýšenia. Modul *v.nnstat* predstavuje nástroj priestorovej štatistiky na posúdenie, či sú body v danej vzorke usporiadané náhodne, v zhlukoch alebo v pravidelnej vzorke.

Oba moduly boli testované s použitím syntetických aj reálnych dát a boli navrhnuté možnosti optimalizácie ich výkonu.

Literatúra

Chandrasekhar, S. (1943). The Law of Distribution of the Nearest Neighbor in a Random Distribution of Particles. *In Reviews of Modern Physics. Stochastic Problems in Physics and Astronomy*, 15:1–89. Available at: `http://rmp.aps.org/abstract/RMP/v15/i1/p1_1` [date accessed 2013-03-21]. ISSN 1539-0756.

Clark, P. J. and Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, 35(4):445–453. Available at: `https://courses.washington.edu/bio480/Week1-PAPER-Clark_and_Evans1954.pdf` [date accessed 2013-03-21]. ISSN 0012-9658.

ESRI (2012). *What is a z-score? What is a p-value? In ArcGIS Help 10.1 [online]*. Available at: `http://resources.arcgis.com/en/help/main/10.1/index.html#/What_is_a_z_score_What_is_a_p_value/005p00000006000000/` [date accessed 2013-03-21].

ESRI (2013a). *ArcGIS Desktop: Release 10.1*. Redlands, CA: Environmental Systems Research Institute.

ESRI (2013b). *Average Nearest Neighbor [computer software]. In ArcGIS Desktop: Release 10.1, Spatial Statistics Toolbox*. Redlands, CA: Environmental Systems Research Institute.

Fabbri, P. and Trevisani, S. (2005). A Geostatistical Simulation Approach to a Pollution Case in Northeastern Italy. *Mathematical Geology*, 37(6):569–586. Available at `http://link.springer.com/article/10.1007%2Fs11004-005-7307-6#page-1` [date accessed 2014-05-10].

Facas, N., Mooney, M., and Furrer, R. (2010). Anisotropy in the Spatial Distribution of Roller-Measured Soil Stiffness. *International Journal of Geomechanics*, 10(4):129–135.

GRASS Development Team (2014). *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7*. Open Source Geospatial Foundation, USA. Available at: `http://grass.osgeo.org`.

Gringarten, E. and Deutsch, C. V. (2001). Teacher's Aide: Variogram Interpretation and Modeling. *Mathematical Geology*, 33(4):507–534. Available at `http://link.springer.com/article/10.1023%2FA%3A1011093014141#page-1` [date accessed 2014-05-14].

Hertz, P. (1909). Über den geigenseitigen durchschnittlichen Abstand von Punkten, die mit bekannter mittlerer Dichte im Raume angeordnet sind. *Mathematische Annalen*, pages 387–398. According to: Clark, P. J., Evans, F. C. (2003). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. In Ecology [online]. Vol. 35, Is. 4. October 1954, pp. 445-453. Available at: `https://courses.washington.edu/bio480/Week1-PAPER-Clark_and_Evans1954.pdf` [date accessed 2013-03-21]. ISSN 0012-9658.

Hoffmann-Wellenhof, B. and Moritz, H. (2006). *Physical Geodesy*. Vienna, Austria: Springer–Verlag. 403 p. ISBN-10-3-211-23544-7.

Hughes, I. G. and Hase, T. P. A. (2010). *Measurements and their Uncertainies: A practical Guide to Modern Analysis*. 1. edition. New York : Oxford University Press Inc., New York. 136 p. ISBN 978-0-19-956633-4.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York. ISBN: 0195050134. In Zimmerman, D. L. (1993). Another Look at Anisotropy in Geostatistics. In *Mathematical Geology*, 25(4):453–470.

Kubeš, P., Grand, T., Šefara, J., Pašteka, R., Bielik, M., and Daniel, S. (2001). Atlas geofyzikálnych máp a profilov : Záverečná správa geologickej úlohy, Časť gravimetria [Atlas of geophysical maps and profiles : Final report of geological survey. Gravity measurements]. Štátny geologický ústav Dionýza Štúra, Bratislava.

MathWorks, Inc. (2010). *MATLAB [computer software]. Version 7.11.0*. Natick, Massachusetts: The MathWorks, Inc.

Mitáš, L. and Mitášová, H. (1988). General variational approach to the interpolation problem. *Computers & Mathematics with Applications*, 16(12):983–992. ISSN 0898-1221. Available at `http://www4.ncsu.edu/~hmitaso/gmslab/papers/CMA1988.pdf` [date accessed 2014-05-13.

Mitáš, v., Mitášová, H., Kosinovsky, I., Gerdes, D., and Hofierka, J. (2014). *v.vol.rst*. GRASS Development Team. Available at `http://trac.osgeo.org/grass/wiki/DownloadSource#GRASS7` [date accessed 2014-03-18].

Mitášová, H. (2014). Space–time points JR9908_xytz_100d.txt. online. In *Geospatial Analysis and Modeling MEA792*. Available at `http://courses.ncsu.edu/mea792/common/Assign_GISamodel/Volume_GISamodel.html` [date accessed 2014-04-23].

Myers, D. E. and Journel, A. (1990). Variograms with Zonal Anisotropies and Noninvertible Kriging Systems. *Mathematical Geology*, 22(7):779–785. Available at todo [date accessed 2014-05-11].

Omre, H. (1984). *Introduction to geostatistical theory and examples of practical applications*. Norwegian Computing Center. Available at `http://www.math.ntnu.no/emner/TMA4250/info/geostat84.PDF` [accessed date 2014-05-14].

PCL Users mailing list (2012). c++ float type and precision issues. online. Available at `http://www.pcl-users.org/c-float-type-and-precision-issues-td3959221.html`.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7):683–691. Available at `http://www.gstat.org/` [date accessed 2014-05-11].

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at `http://www.R-project.org/` [date accessed 2014-05-11].

Rusu, R. B. and Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China. Available at `http://pointclouds.org/about/` [date accessed 2014-05-12].

Stopková, E. (2013). Extension of mathematical background for Nearest Neighbour Aanalysis in three-dimensional space. *Geoinformatics FCE CTU*, 11:25–36. ISSN 1802-2669. Available at `http://geoinformatics.fsv.cvut.cz/pdf/geoinformatics-fce-ctu-2013-11.pdf` [date accessed].

Stopková, E. (2014). Optimalizácia vybraných metód priestorovej štatistiky v trojrozmernom priestore s využitím kd-stromu [Optimization of selected methods of spatial statistics in three-dimensional space using kd-tree]. In todo, editor, *GIS Ostrava 2014 : Sborník symposia*, page 6, Ostrava, Czech Republic. Ostrava : VB - TechnickĂ˘ univerzita Ostrava. todo: isbn, access.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications.* Heidelberg: Springer-Verlag. 387 p. ISBN 3-540-44142-5.

Weisstein, E. W. (2013). Gamma Function. *MathWorld. A Wolfram Web Resource [online].* Available at: `http://mathworld.wolfram.com/GammaFunction.html` [date accessed 2013-03-21].

White, J. G. (2002). Soil Moisture Data. online. Available at: `http://courses.ncsu.edu/mea592/common/gisdata/tdr_obsday1.dat`.

Wolfram Research, Inc. (2008). *Mathematica [computer software]. Version 7.0.* Champaign : Wolfram Research, Inc.

Zimmerman, D. L. (1993). Another look at anisotropy in geostatistics. *Mathematical Geology*, 25(4):453–470.

List of publications

Publications

2014

Eva Stopková. *Optimalizácia vybraných metód priestorovej štatistiky v trojrozmernom priestore s využitím kd-stromu* [Optimization of selected methods of spatial statistics in three-dimensional space using kd-tree]. In GIS Ostrava 2014 [online]: Geoinformatics in Motion. Ostrava, Czech Republic, 27th - 29th January 2014. Ostrava: VSB – Technical University of Ostrava campus, 2014, 6 p. ISBN 978-80-248-3310-1.

2013

Eva Stopková. *Extension of mathematical background for Nearest Neighbour Analysis in three-dimensional space*. In Geoinformatics Vol.11. p. 25-36. ISSN 1802-2669.
`http://geoinformatics.fsv.cvut.cz/pdf/geoinformatics-fce-ctu-2013-11.pdf`

Eva Stopková. *Nearest Neighbour Analysis in three-dimensional space*. In Geoinformatics 2013 [online] : Praha, Czech Republic, 11th -12th[th] July 2013. Prague : Faculty of Civil Engineering, Czech University of Technology in Prague, 2013. `http://geoinformatics.fsv.cvut.cz/gwiki/Nearest_Neighbour_Analysis_in_three-dimensional_space`

Eva Stopková. *Interpolácia dát v 3D priestore rozšírenou metódou kriging* [3D data interpolation using extended method kriging]. In GIS Ostrava 2013 [online] : Geoinformatics for Society. Symposium. Ostrava, Czech Republic, 21st – 23rd January 2013. Ostrava: VSB – Technical University of Ostrava campus, 2013, p. 8. ISBN 978-80-248-2951-7. `http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2013/sbornik/papers/gis20135076d......889cc867.pdf`

Renata Ďuračiová, Tibor Lieskovský, Eva Stopková, Kristína Kročková: *The Benefit of Fuzzy Logic to Protection of Cultural and Historical Heritage*. In GIS Ostrava 2013 [online]: Geoinformatics for Society. Symposium. Ostrava, Czech Republic, 21st - 23rd January 2013. Ostrava: VSB – Technical University of Ostrava campus, 2013, p. 12. ISBN 978-80-248-2951-7. `http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2013/proceedings/papers/gis2013508e634f37269.pdf`

2012

Tibor Lieskovský, Eva Stopková, Peter Demján: *"Slovenské Stonehenge" – využitie geografických informačných systémov pri skúmaní archeologických lokalít* ["Slovak Stonehenge" – using geographical information systems for archaeological sites research]. In Proceedings of seminar "60th Anniversary of the Department of Theoretical Geodesy". Kočovce, Slovak Republic, 4th - 5th December 2012. Bratislava: FCE SUT, 2012, p. 46. ISBN 978-80-227-3843-9.

Eva Stopková. *Možnosti tvorby a využitia 3D obalových zón* [Possibilities of creation and usage of 3D buffer]. In Advances in architectural, civil and environmental engineering [CD-ROM] : 22nd Annual PhD Student Conference. Bratislava SR, 15th November 2012. Bratislava: FCE SUT, 2012, p. 203-205. ISBN 978-80-227-3853-8.

Eva Stopková. *Návrh riešenia interpolačnej metódy kriging v priestore* [Design of interpolation method 3D kriging solution]. In Juniorstav 2012. 14. International Conference of PhD Students [CD-ROM]. Brno: Faculty of Civil Engineering, Brno University of Technology. 26th January 2012. 11 p.

2011
Eva Stopková. *Publikovanie archeologických dát vo webovom prostredí* [Publishing of archaeological data in web environment]. In Juniorstav 2011. 13. International Conference of PhD Students [CD-ROM]. Brno: Faculty of Civil Engineering, Brno University of Technology. 4th of February 2011. 7 p.

2010
Eva Stopková, Tibor Lieskovský. *Možnosti publikovania archeologických dát vo webovom prostredí* [Possibilities of publishing of archaeological data in the web] (in Slovak). In GIS in Archaeology 2010 [CD-ROM] : Proceedings. Nitra: Institute of Archaeology of Slovak Academy of Sciences. ISBN 978-80-89315-25-3. 8 p.