Research Article **Open Access**

Robert Zůvala*, Eva Fišerová, and Lukáš Marek

Mathematical aspects of the kriging applied on landslide in Halenkovice (Czech Republic)

Abstract: Kriging is one of the geostatistical techniques for spatial data analysis that is usually used for a modelling of natural phenomena or a creation of digital elevation models. In this paper, we introduce kriging methods in the context of a landslide modelling in time. The proposed procedure, as well as most of the statistical methods, is designed for complete data sets, *i.e.* the observations at the beginning and the end of the study are available. In order to use all the information from the data and to avoid the loss of information after omitting observations with missing values, the algorithm for imputation of missing data values based on kriging techniques is proposed. The methodology was verified by the landslide modelling in Halenkovice, Czech Republic, during the period from June 2008 to March 2010. The obtained results showed a potential of kriging methods for the landslide modelling.

Keywords: universal kriging; variogram; landslide; missing values; data imputation

1 Introduction

One issue of the spatial data analysis is focused on the estimation of values of a certain phenomenon in a study area when values of this phenomenon are given only in certain locations. Such an analysis is based on interpolation. After the analysis, the values of the phenomenon are known at all points of the study area; the values are obtained either directly by measurement, or by estimation. The landslides monitoring is one of the particular areas, in which interpolation is commonly utilised.

Approaches for the spatial modelling of landslides can be divided into two groups. The first group includes deterministic and dynamic modelling methods of physical mechanisms that control slope failure [1]. These approaches are used scarcely because of the detailed data requirements. The second group consists of heuristic or statistical methods that use a relation between locations of previous landslides and geo-environmental variables to predict areas of landslide initiation with similar combinations of factors [2]. Generally, one can say that natural processes are spatially analysed and modelled by rather deterministic methods while social processes by stochastic methods. This is because the level of complexity of the system may appear simpler for the natural process than for the social process [1].

Although the complexity of natural phenomena might seem simpler, the number of initial conditions plays important roles in these processes. Conditions making a region susceptible to landmass movements encompass factors such as soil characteristics, topography, precipitation, vegetation, and others [3]. Various statistical, analytical and numerical approaches have been introduced for modelling of landslides. Eligibility of the artificial neural networks for mapping landslide susceptibility has been discussed in [4], the effect of water pressure and soil porosity on the occurrence of landslides in [5]. Rainfall intensity duration curves in regional [6] and global scales [7] have been used in developing of landslide models. Several attempts have been made to apply different methods of Landslide Hazard Zonation. The multivariate techniques are proved to be effective in spatial prediction of landslides with the high degree of accuracy [8].

The basic principle of spatial predictions and of geostatistical analyses in general, is that events that are closer to each other in a given area tend to be more similar than events that are more distant. With increasing distance from the prediction site, the impact of these sites on the

*Corresponding Author: Robert Zůvala:** Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacky University in Olomouc, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic; Department of Applied Informatics, Faculty of Science, Tomas Bata University in Zlín, Nad Stráněmi 4511, 760 05 Zlín, Czech Republic; Email: r.zuvala@seznam.cz
Eva Fišerová: Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacky University in Olomouc, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic
Lukáš Marek: Department of Geography, University of Canterbury, Private Bag 4800, 8140 Christchurch, New Zealand; Department of Geoinformatics, Faculty of Science, Palacky University in Olomouc, 17. listopadu 50, 771 46 Olomouc, Czech Republic

prediction drops and from a certain distance, the influence of these remote places on the prediction is equal to zero. If values of individual measurements are not fundamentally different and structure of the area remains constant, it is possible to interpolate area primarily from points lying close to each other. On this idea the Inverse Distance Weighting (IDW) method, which is one of the deterministic methods, is based. Other methods are *e.g.* Radial Basis Functions (RBF) method [9], Global Polynomial interpolation [10], Local Polynomial interpolation [10], trend [11], Thin Plate Spline [12] or even neural networks [13].

The most significant advantage of IDW is its (mathematical) simplicity, which causes that the method is fast and its results are easy to understand. The usage of IDW is the most appropriate for interpolation tasks with the high number of source points but IDW can be used also for showing barriers, discontinuous lines and fractures on the surface [14]. The disadvantage of the IDW is the formation of concentric contour lines (bull eyes) around entry points. The reason is a strong influence of these points in their surroundings, especially when a higher degree polynomial function is chosen. The smoothing parameter was introduced for attenuation of this phenomenon. Another disadvantage of the IDW method is that it cannot calculate higher or lower values than the values of the input data. So if observations with extreme values are not present in the data, there are distortions in missing locations.

The RBF method, unlike the IDW, can deal with this situation. RBF has various applications in practice, due to their simplicity, generality and fast learning stage. RBF is powerful and flexible method and is useful for gridding almost any type of data set [15]. It provides suitable results for phenomena that are changing gradually in space, *e.g.* elevation except the steep mountain areas. However, RBF is not suitable for interpolation of phenomena with step changes in values over short distances. The disadvantage of the trend method is a deformation of resulting structures in the periphery of the area, with the best estimate in the central part. In case that the observed values are low in the periphery, the method may provide negative estimates, which may not comply with the studied phenomenon.

The kriging method is one of the most commonly used geostatistical methods [16, 17]. Kriging, like the IDW interpolation, create weights from surrounding measured values to predict values at unmeasured locations. However unlike IDW, kriging directly incorporates spatial autocorrelation. Kriging weights come from a semivariogram, which was developed for modelling of spatial variability of a data structure. To create a map of the phenomenon, optimal linear predictions are made for locations in the investigated area. They are based on the semivariogram and spatial arrangement of measured values that are nearby. Kriging methods are usually used for modelling of mineral deposits. The aim of this paper is to introduce the kriging method in the context of landslide modelling and as an imputation method for missing values; and to discuss main issues of the kriging such as stationarity and isotropy. The application of this methodology is demonstrated on landslide modelling in Halenkovice, Czech Republic, for the observation period from June 2008 to March 2010.

The landslide in Halenkovice has been previously studied from the different point of view in [18, 19]. In these papers, regression models for estimation of shifts velocity in the *x*- and *y*-direction together with a deformation circle as a visualisation tool are proposed. Statistical tests for the landslide detection are proposed in [19].

The paper is organised as follows. Section 2 is devoted to the basic theoretical knowledge of semivariograms and kriging methods, especially to ordinary and universal kriging methods. Further, this section contains data description and processing, and the procedure for imputation of missing data values. The construction of a landslide model is described in Section 3. Kriging maps of movements slope are created along with the kriging standard errors at any location on the map. For better clarity, 3D model of the landslide is also constructed, where the third dimension shows the total shift for the whole period of observation. The last sections consider a conclusion and a discussion.

2 Methods

2.1 Universal kriging and variogram

The kriging method is one of the geostatistical interpolation methods that are used for modelling of spatial data [20]. The term kriging includes a variety of interpolation techniques whose aim is to find an optimal linear prediction. From the statistical point of view, kriging provides the best linear unbiased estimators [21].

Let the symbol $Z(s)$ denotes the observed value of the studied spatial process at the point s. For example, in the study of mineral deposits, $Z(s)$ is the observed amount of mineral resources at a given point s; in landslide modelling of the slope, $Z(s)$ is the observed overall shift at a point s in studied time period. The principle of prediction is based on a weighted average of neighbouring values $Z(s_i)$, where the weights λ_i depend on the distance and spatial relationship between observed points. Mathematically, the prediction at the point s can be described by the equation $\hat{Z}(s) = \Sigma \lambda_i Z(s_i)$.

There are many different methods for determination of the weights and each of them is associated with some kriging method. Basic methods, which are discussed in the paper, are ordinary and universal kriging [22].

Ordinary kriging is based on the assumption that the correlation between two random variables depends only on their spatial distance that separates them and is independent of their position. Moreover, the variance of the difference of two random variables $Z(s)$ and $Z(s + h)$ also depends only on their spatial distance h. More precisely, $var[Z(s + h) – Z(s)] = 2\gamma(h)$, where $2\gamma(h)$ is called variogram, $\gamma(h)$ is semivariogram [21]. The variogram is described later in this section. Furthermore, the ordinary kriging assumes stationarity of the first moment of all random variables, *i.e.* it assumes constant mean $E[Z(s_i)] = \mu$, which is unknown.

If the first moment of the observed field is not stationary and a polynomial trend occurs in the data, it is appropriate to use the universal kriging [23] instead of the ordinary kriging. The ordinary kriging assumes the stationarity of the first moment. Under this assumption, the average value of the scatter points is relatively constant. Whenever there is a significant trend in the data values, such as sloping surface or a locally flat region, this assumption is violated. Universal kriging is much like the ordinary kriging, except that instead of fitting just a local mean in the neighbourhood of the estimation point a linear or higher-order trend is fitted in the (x, y) coordinates of the data points [2]. The trend of the first-order is given by $E[Z(s_i)] = m(x_i, y_i) = \mu + a_1 x_i + a_2 y_i$, where μ, a_1 and a_2 are local trend coefficients of the first-order trend, and x_i, y_i are coordinates of input point s_i. The resulting kriging equations at the location s_0 are

$$\sum_{j=1}^{N} \lambda_j \gamma\left(\left|s_i - s_j\right|\right) + \omega + a_1 x_i + a_2 y_i = \gamma\left(\left|s_i - s_0\right|\right),$$

$$i = 1, 2, \ldots, N,$$

$$\sum_{i=1}^{N} \lambda_i = 1, \quad \sum_{i=1}^{N} \lambda_i x_i = x_0, \quad \sum_{i=1}^{N} \lambda_i y_i = y_0.$$

Here the symbol $|s_i - s_j|$ means the distance between the locations s_i and s_j; ω, a_1, a_2 are Lagrange multipliers. The variance of the universal kriging estimator with the linear trend at the location s_0 is

$$\sigma^2\left(\hat{Z}(s_0)\right) = \sum_{i=1}^{N} \lambda_i \gamma\left(\left|s_0 - s_i\right|\right) + \omega + a_1 x_0 + a_2 y_0.$$

The square root of the kriging variance is called the kriging standard error (KSE).

The regionalized variable theory assumes that the variation of the variable under study is the same in all directions. Under this assumption, the spatial variation is isotropic, and it is possible to use the interaction of all possible pairs of points of the input data to estimate the variogram. However, in practice geostatistical data very often show anisotropy [24] in some direction, and thus autocorrelation functions vary in different directions. In the case of anisotropy, it is necessary to determine direction estimates for variogram that does not show such loading [21]. Elliptical rose diagrams [25] typically indicate the presence of anisotropy. A rose diagram is a circular histogram that displays directional data and the frequency of each class. If the spatial correlation pattern is anisotropic, the range varies in different directions. In the rose diagram, one can identify the major and minor axes of the elliptical range distribution. These new coordinate systems are suitable for construction of variograms.

The empirical form of a variogram is defined as

$$2\gamma(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \left[Z(s_i) - Z(s_i + h)\right]^2,$$

where $N(h)$ is a number of h-distant point pairs $(s_i, s_i + h)$. The result of the point estimates is a set of variogram values for various distances h between points. Nevertheless, knowledge of these variogram values is not enough for compiling the kriging equations. One also need to know the variogram values for any distance h. Therefore, when quantifying spatial dependence, obtained estimates are approximated with a theoretical curve. The most common semivariogram models are linear, spherical, exponential, or Gaussian model [16, 26].

There are specific characteristics describing the semivariogram models [27]. Range defines the distance, where the semivariogram value is constant. Sample locations separated by a distance shorter than the range are spatially autocorrelated; locations farther apart than the range are uncorrelated. Sill is the semivariogram value at the range. The nugget effect is defined as the limit of the semivariogram for h decreasing to zero. The nugget effect can be caused by measurement errors or by the fact that the process includes spatial variability at distances smaller than the sampling interval [28].

A better fitted semivariogram model (with respect to the experimental semivariogram) produces a better kriging estimate. A semivariogram model that takes due cognisance of any anisotropic behaviour in the data will also produce a better kriging estimate. According to general recommendations, more sophisticated semivariogram model should be used just in cases if simple semivariogram models cannot capture the major spatial structure

of the experimental semivariogram to a satisfactory level. Overfitting of an experimental semi-variogram will not produce better kriging estimates than a simpler model that adequately captures the overall spatial structure of that regionalized variable. The most important aspect of the fitted semivariogram model for the kriging estimate is the nugget-effect and the slope near the origin [16, 29]. In general, fit the modelled semivariogram sill to the experimental semivariogram sill rather than to the sample variance.

Frequently used measures that indicate how well a statistical model fits data are *e.g.* [30] the root mean square error (RMSE) and the coefficient of determination R^2. The RMSE characterizes an achieved precision of the fit and it can be interpreted as an average deviation of the fitted and observed values. The lower the values of RMSE are, the better is the fit. The coefficient of determination R^2 indicates the proportion of variation explained by the model. It takes values from zero to one; the higher the values, the better the model. The value can be interpreted as how accurately a model explains a phenomenon. Formulas for the RMSE and R^2 are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left[Z(s_i) - \hat{Z}(s_i) \right]^2}{N}},$$

$$R^2 = \frac{\sum_{i=1}^{N} \left(\hat{Z}(s_i) - \bar{Z}(s) \right)^2}{\sum_{i=1}^{N} \left(Z(s_i) - \bar{Z}(s) \right)^2},$$

where $Z(s_i)$ is the observed value, $\hat{Z}(s_i)$ is the predicted value at the location s_i, and $\bar{Z}(s)$ is the average value of $Z(s_i)$.

The predictive performance of a fitted model can be evaluated using the leave-one-out cross-validation (LOOCV) method [31, 32]. The idea of the cross-validation is to perform several splitting of the data into the training sample used for fitting the model, and into the validation sample (remaining data) used for evaluating the predictive accuracy. In the LOOCV procedure, each data point is left out from the sample and used for validation. For large datasets, further splitting procedures are available. The predictive accuracy of a model can be measured by the RMSE and the mean absolute error (MAE) on the test set of data not used in estimation. In the MAE, all errors are weighted equally in the average; in the RMSE, the errors are squared, therefore the RMSE gives a relatively high weight to large errors. This means that the RMSE is more useful when large errors are undesirable. The RMSE is always larger or equal to the MAE; the greater difference between them, the greater the variance in the individual errors. If the RMSE and MAE are equal, all errors are of the same magnitude. The explicit formula for the MAE is given as

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| \hat{Z}(s_i) - Z(s_i) \right|$$

2.2 Data description and processing

The case study, in which the usage of presented methods is demonstrated, is dealing with the shallow landslide situated near Halenkovice municipality in Zlín Region, Czech Republic. The study area is located in the outer part of the Western Carpathians, which is created by the Mesozoic and Tertiary deposits. This part of Carpathians is called Flysch Carpathians. The geological structure of the flysch is characterized by alternating sandstones and conglomerates with clay slates that create layers with varying thickness and strength. This fact is one of a cause of frequent slope instabilities in the area [33]. There are different types of slope instabilities and slope movements occurring in the flysch, but the most common are shallow landslides. They evolve on the inclined slopes in the mantle rock and may be caused by increased precipitation. The study area is represented by the slope with inclination 10–15°, which has NW orientation. The area has been devastated by recent deformations. It has not been properly stabilized since its initialization in March 2006, so the further evolution and reactivation stages are expected. A landslide lies in the pasture, and it is not near any residential or other buildings. Hence, there are no people in immediate danger, implying that stabilization is not necessary. Area of the slope deformation is approximately 5 000 m²; its dimensions reach a length of 120 m and width of 80 m [34]. All three zones of landslide progression (detachment, transport and storage) are easily recognizable within the landslide. Scarp of the landslide varies in size and height, ranging from 30 to 250 cm. Debris thickness at the toe reaches as much as 50 cm. There are 1 m high lateral accumulations on flanks of the landslide.

The place of the landslide, marked with a red marker, is displayed in Figure 1. In the year 2008, 28 fixed ground points were located as it is demonstrated in Figure 2. Twenty four fixed ground points were installed in June 2008; in October 2008 were added four more points (7, 10, 12, 28). The point 0 was the basic position. Three points (25, 26, 27) served as orientation points. The measurement was finished in March 2010. All measurements were performed by the total station Trimble 5503 DR Standard. Coordinates of all points are based on the S-JTSK/Krovak East North (EPSG code 5514, Datum of Uniform Trigonometric Cadastral Network) coordinate system.

    (a)   **(b)**
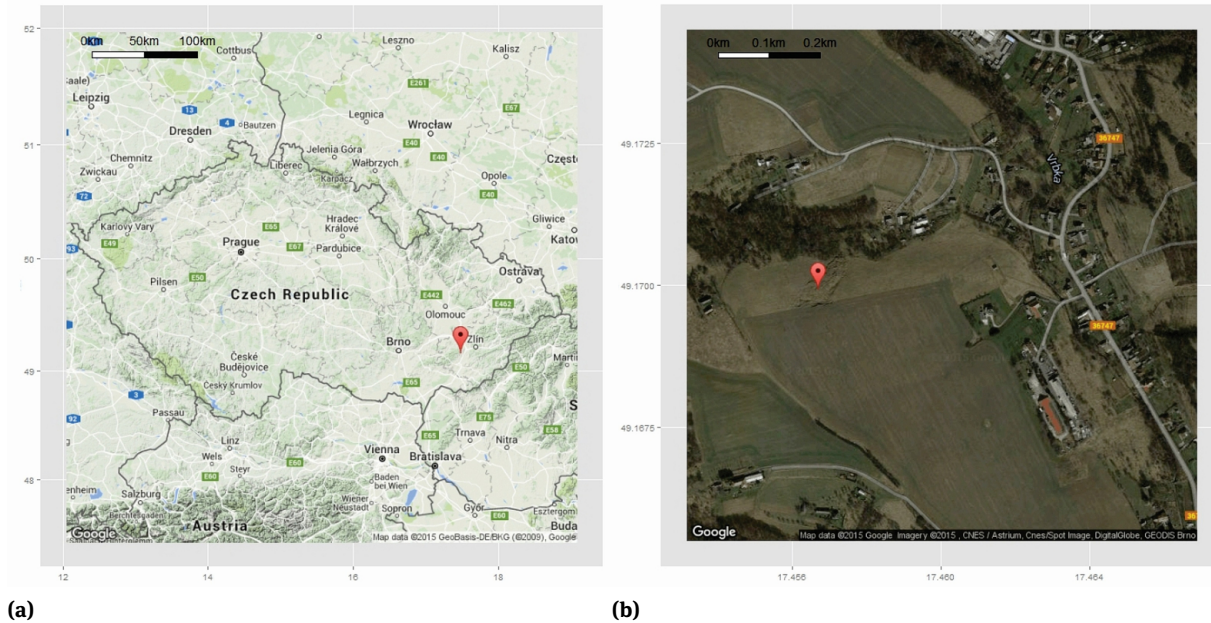
Figure 1: Localization of the landslide in the Czech Republic (a); the study area in detail (b).

Due to significant movement occurred between June and October 2008, we set June as the starting month for landslide modelling. The coordinates from June are not available for four points installed in October, and thus they cannot be directly used for modelling. In order to use all the information from the data and to avoid the loss of information after omitting these points, it is necessary to impute the missing coordinates from June for them. The proper method is described below.

The coordinates at the points 8, 9 and 15 are known only from the initial measurement. These points have never been found later. Since their movement in the study area is unknown, they are unusable for modelling, and thus they were excluded from the modelling. Therefore, they are already no longer displayed in the graphical output.

2.2.1 The imputation method for missing data values

For landslide modelling by the kriging, it is necessary to have observations at both, the beginning and the end of studied period, say at time T_0 and T_n. In practice, however, it often happens that data are observed at different initial times, say T_0 and T_1. Therefore, it is necessary to select the start and the end of the modelling period properly and, if possible, to impute the missing data values at the locations that have been observed later. For these situations, when enough observation at both times T_0 and T_1 is available, we recommend to use kriging method for the period from T_0 to T_1 and to predict missing values at time T_0.

The imputation of missing values plays an important role in the evaluation of resulting kriging model. Since the accuracy of kriging predictions depends on data configuration in a studied area, knowledge of movements of more points can substantially affect the resulting model.

3 Results of the landslide modelling in Halenkovice

When using kriging methods for landslide modelling, the value of the process $Z(s)$ is the observed overall shift at the point s for the period from June 2008 to March 2010; $Z(s+h)$ is the observed overall shift at the point $s + h$, the point situated at the distance h from the point s, for the same period. All calculations were performed using the R project library geoR [35].

For the construction of a model, the dataset completed by the universal kriging for the period from June 2008 to October 2008 was considered. The imputation was performed using all available points, *i.e.* 22 points. The imputed coordinates of the points 7, 10, 12, and 28 are listed in Table 3. Here, the discussion of the procedure how these values were obtained is skipped. The procedure is analo-
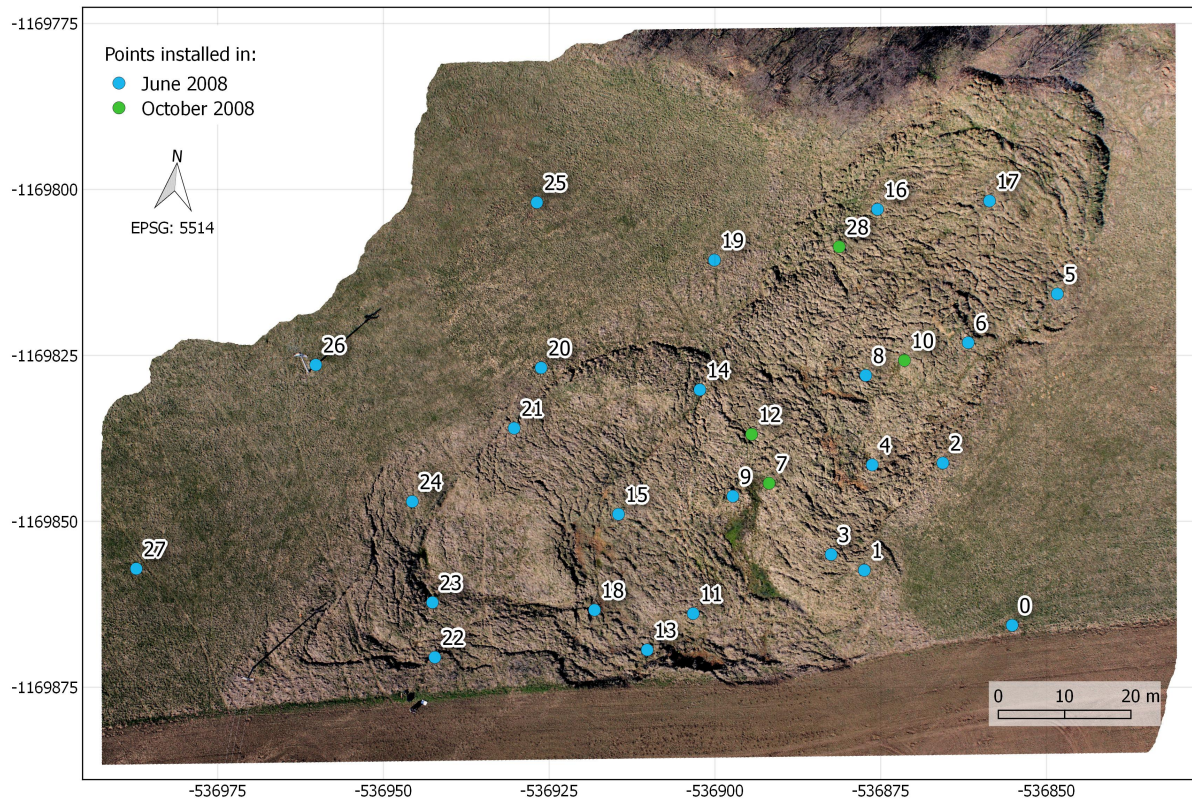
Figure 2: Layout of fixed ground points in landslide area Halenkovice. Lower part (accumulation zone) of the landslide is in the right upper part of the figure.

gous to the construction of a landslide model in the period from June 2008 to March 2010 which is thoroughly described in the following.

The assumption of normality of the data was verified by Shapiro-Wilk test with *p*-values equal to 0.31 for *x*-direction, 0.28 for *y*-direction and 0.37 for *z*-direction. The presence of anisotropy in *x*-, *y*- and *z*-direction was indicated in rose diagrams (Figure 3). One can see the major and minor axes of the elliptical range distribution in each direction that should be used for the construction of individual variograms. Namely, for *x*-direction, the axes are rotated by the angle of 22.5°, for *y*-direction by the angle of 30°, and for *z*-direction by the angle of 18°.

A presence of linear trend $m(s)$ was detected in the field during the structural analysis, and thus the universal kriging was used. The correctness of the trend is assessed by verifying the condition of stationarity of the first moment applied on residuals $R(s) = Z(s) - m(s)$. The obtained residuals indicate stationarity in all directions (Figure 4).

The semivariogram models with linear trend (Figure 5) vary in different directions as expected with respect to rose diagrams. Therefore, when creating kriging maps,

they were constructed separately with shifts in *x*-, *y*- and *z*-directions. For the construction of kriging maps, Gaussian semivariogram models were used without nuggets and with ranges about 65 metres for the *x*-, *y*-directions, and about 70 metres for the *z*-direction. The sill for the *x*-direction is equal to 0.3 m², for the *y*-direction it equals 1.4 m², and for the *z*-direction it is 0.08 m². Curves for exponential, linear and spherical models in semivariograms in the *x*- and *y*-directions were overlapping; exponential and spherical models also for *z*-direction were overlapping.

The different scales of shifts were found in various directions (Figures 6–8). In the *x*-direction, points with the largest shifts are located nearby the upper right corner of Figure 6 representing the lower part of the slope. According to the density of contours, the landslide occurred gradually from the south-west direction from the point 23 to the point 17. Conversely, a steep landslide occurred near the points 5 and 28, where contours are the densest. In the *x*-direction (Figure 6), the points 6, 16, and 17 had the greatest shifts by approximately 1.0–1.2 m. At the points 4, 10, 12, 14 and 28 shifts about 50 cm were observed and in

Rose diagram in x− direction   **Rose diagram in y− direction**   **Rose diagram in z− direction**

(a)   (b)   (c)
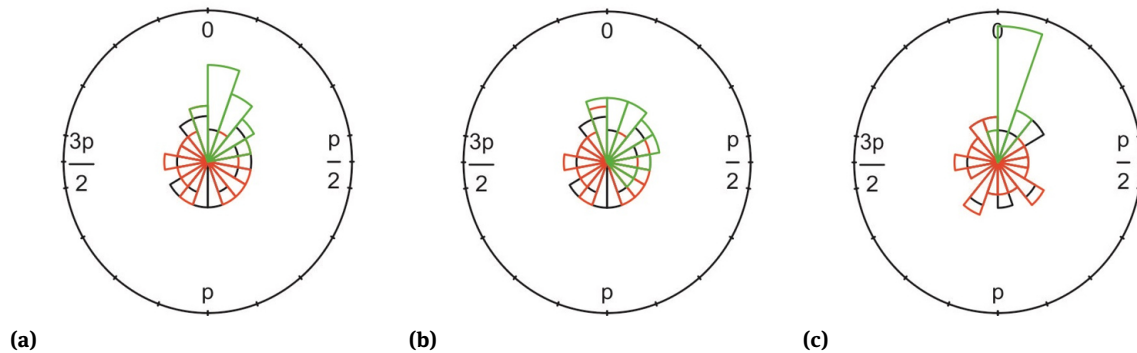
Figure 3: Rose diagrams of shifts in *x*-direction (a), *y*-direction (b), and *z*-direction (c) for the completed dataset.

Table 1: Results of the leave-one-out cross-validation for the completed dataset.

	x-direction	*y*-direction	*z*-direction
RMSE	36.6 cm	53.4 cm	12.3 cm
MAE	25.7 cm	33.4 cm	7.9 cm
Median of AE	13.3 cm	9.8 cm	2.1 cm
Standard error of AE	26.7 cm	42.8 cm	9.6 cm
Correlation coefficient	0.53	0.70	0.66

other locations it was only about 10 cm. The KSE for the *x*-direction equals 10 cm in the observed points and grows with increasing distance from these points.

In the *y*-direction (Figure 7), the shifts were more distinct. The points 6, 10, 16 and 17 had shifts from 1.5 m to 2.2 m.; the points 4, 12, 14 and 28 had similar behaviour as in *x*-direction with the shifts about 50 cm. The shifts at other locations were less than 10 cm. The KSE for the *y*-direction equals 25 cm in the observed points.

The smallest shifts were observed in the *z*-direction (Figure 8). The points 6, 16, 17 and 28 had the greatest shifts from 20 to 30 cm; the points 2, 4, 5 and 10 had shifts about 15 cm; other locations moved only about 5 cm. The KSE for the *z*-direction equals 5 cm in the observed points.
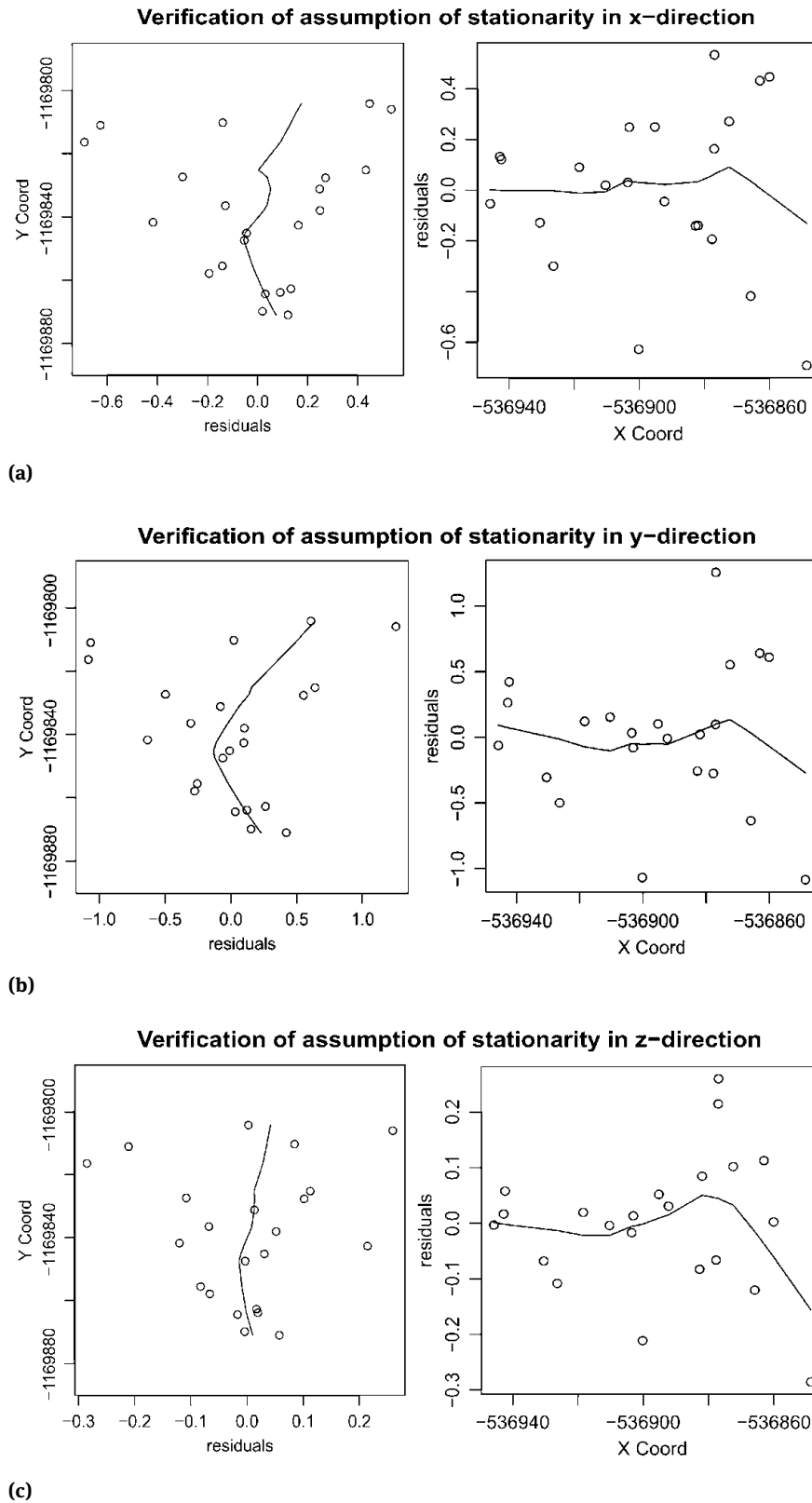
Index of determination and RMSE for modelling the shift is equal to $R^2 = 0.997$, $RMSE = 11.1$ cm in the *x*-direction; $R^2 = 0.998$, $RMSE = 15.4$ cm in the *y*-direction; and $R^2 = 0.998$, $RMSE = 3.5$ cm in the *z*-direction. All characteristics indicate that models very well fit data, and average deviations of the fitted and observed values are sufficiently small and comparable with the observed shifts.

For better clarity, an approximate 3D model (the third dimension represents the total shift in meters) for the landslide is displayed in Figure 9. The ridges of the landslide

(the area with the largest shift) stretch in a curve from the south-west to the north-east. It is obvious that the biggest shift, about 2.5 metres, was around the points 16 and 17 at the lower part of the slope.

The predictive performance of fitted models for *x*-, *y*- and *z*-directions was evaluated using the leave-one-out cross-validation method (Table 1). The obtained results are satisfactory. The MAE values correspond with the ranges of shifts in different directions. In the *x*-direction, maximum shifts are about 1.2 m; in the *y*-direction the maximum shifts are about 2.5 m; and in the *z*-direction it is only about 0.5 m. The RMSE values are greater (< 1.6-fold) than MAE values, which indicate the variance in individual errors. The biggest errors occur in the prediction of some specific points such as points with extreme shifts or points with hardly predictable movements (*e.g.*, point 19), because the resulting surface of kriging maps is smooth. The values of correlation coefficients between observed and predicted values indicate sufficient predictive ability of the fitted models.

In order to show the effect of imputed points 7, 10, 12 and 28, kriging maps were created without these points (Figure 10). In locations, where the omitted points are situated, interpolation resulted according to the nearest points to these locations. The KSE for the *x*-direction in the ob-

Verification of assumption of stationarity in x−direction

(a)

Verification of assumption of stationarity in y−direction

(b)

Verification of assumption of stationarity in z−direction

(c)

Figure 4: Behaviour of residuals in *x*-direction (a), *y*-direction (b), and *z*-direction (c) for the completed dataset. Residuals against to *y*-coordinates are shown on left panels, against to *x*-coordinates on right panels.

(a)

(b)

(c)
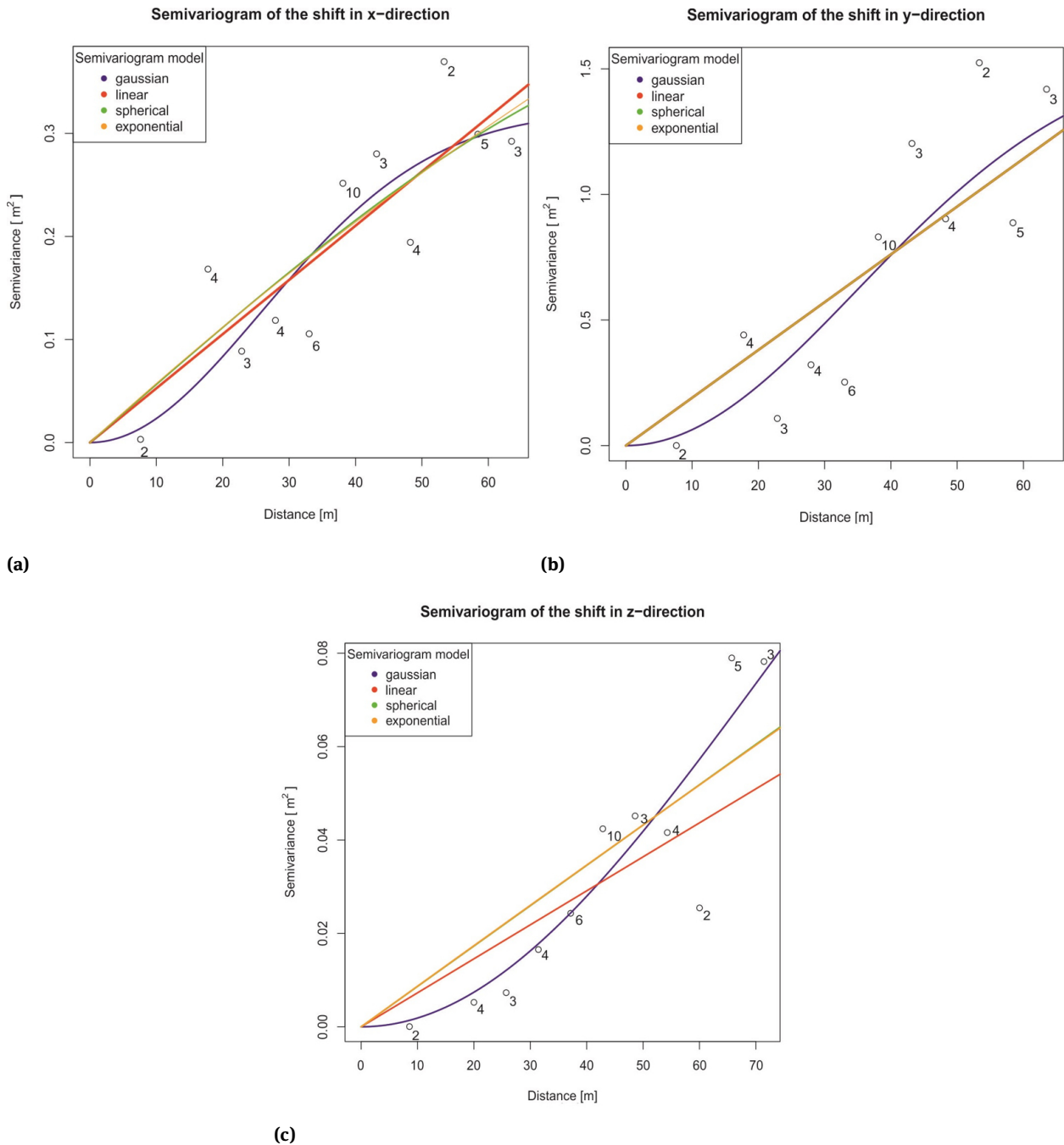
Figure 5: Semivariogram models of the shifts in the *x*-direction (a), *y*-direction (b), and *z*-direction (c) for the completed dataset. Numbers indicate the number of pairs used for sample semivariogram points.

served points equals 15 cm, for *y*-direction equals 30 cm and for *z*-direction equals 5 cm. Index of determination and the RMSE are equal to $R^2 = 0.997$, $RMSE = 9.5$ cm in the *x*-direction; $R^2 = 0.997$, $RMSE = 14.7$ cm in the *y*-direction; and $R^2 = 0.997$, $RMSE = 3.5$ cm in the *z*-direction. The predictive quality of fitted models was again evaluated by means of the LOOCV (Table 2). Finally, pre-

dicted coordinates of the points 7, 10, 12 and 28 were compared them with imputed ones (Tables 3). The differences are not greater than 5 cm what indicate sufficient accuracy of imputed coordinates.

Summarizing the characteristics of models with and without imputed values, the RMSE and the index of determination are almost the same. The KSE increased only

Table 2: Results of the leave-one-out cross-validation for the original dataset.

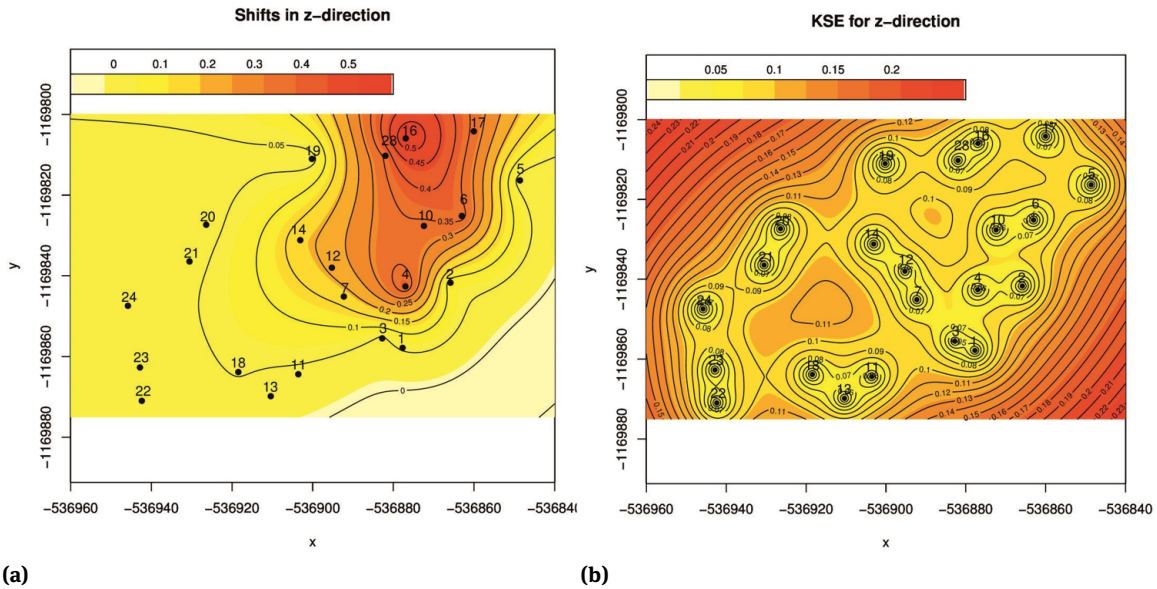|	*x*-direction	*y*-direction	*z*-direction
RMSE	42.1 cm	64.4 cm	17.2 cm
MAE	29.0 cm	39.4 cm	10.9 cm
Median of AE	9.3 cm	5.3 cm	2.1 cm
Standard error of AE	31.4 cm	52.5 cm	13.7 cm
Correlation coefficient	0.41	0.57	0.54

(a) (b)
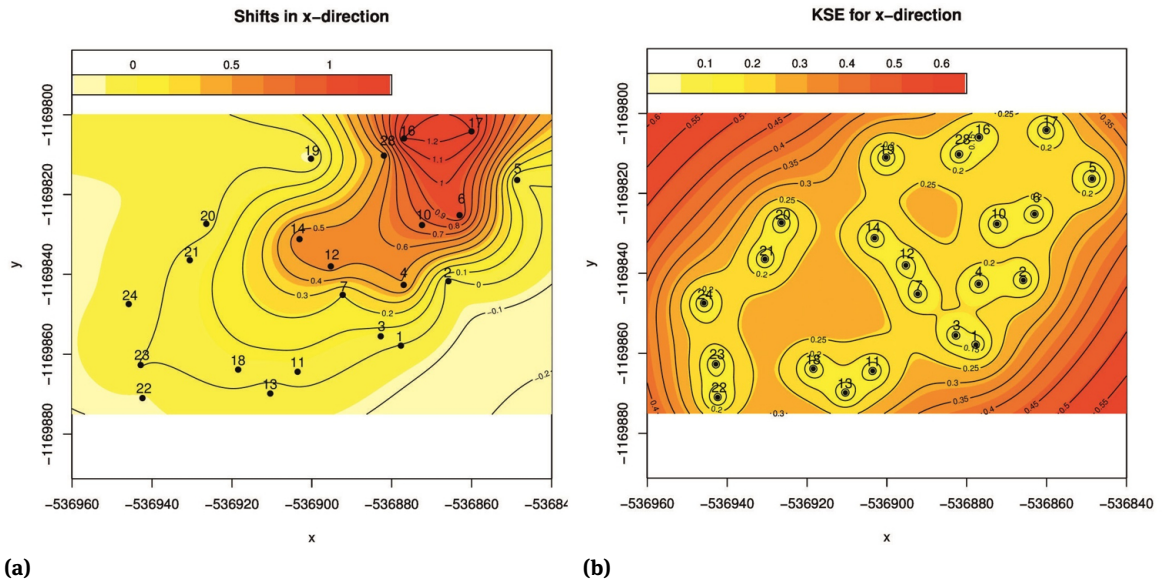
Figure 6: The total estimated shifts in the *x*-direction (a) together with KSE (b) for the completed dataset.


(a) (b)
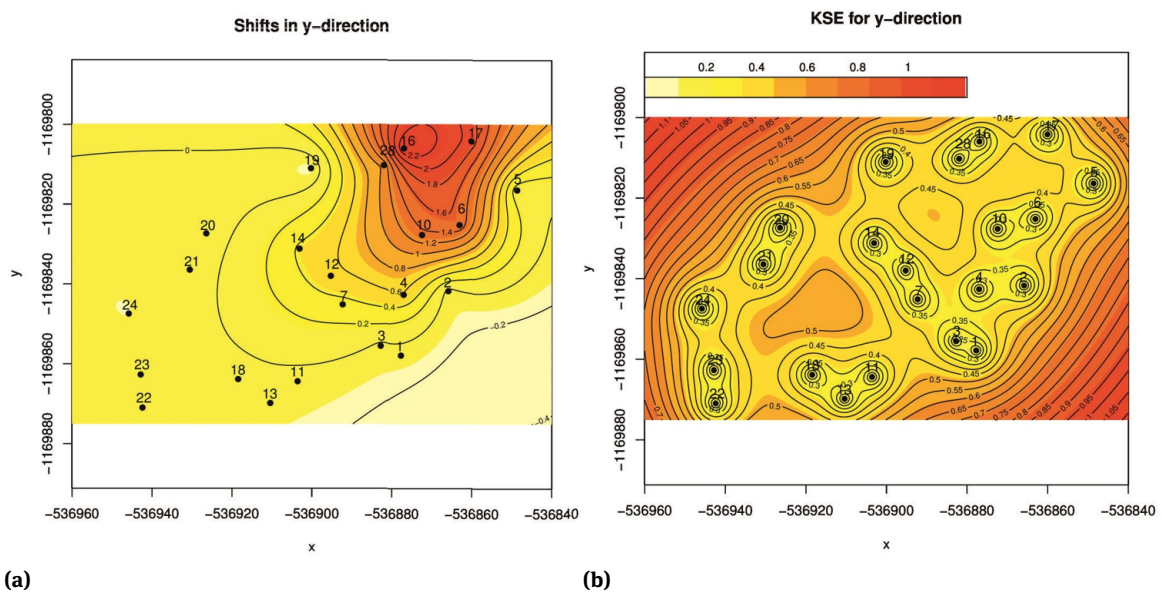
Figure 7: The total estimated shifts in the *y*-direction (a) together with KSE (b) for the completed dataset.

Table 3: The comparison of the imputed coordinates (denoted by a subscript imp) obtained by the universal kriging for the period from June 2008 to October 2008 and those predicted (denoted by a subscript pred) by the universal kriging for the original dataset for the period from June 2008 to March 2010.

Point	X_{imp}	X_{pred}	$\|\Delta X\|$	Y_{imp}	Y_{pred}	$\|\Delta Y\|$	Z_{imp}	Z_{pred}	$\|\Delta Z\|$
7	−536892.269	−536892.278	0.009	−1169845.140	−1169845.840	0.044	266.900	266.890	0.010
10	−536872.418	−536872.398	0.020	−1169827.643	−1169827.593	0.050	263.050	266.038	0.012
12	−536895.245	−536895.239	0.006	−1169837.987	−1169837.971	0.016	266.220	266.213	0.007
28	−536881.959	−536881.952	0.007	−1169810.249	−1169810.241	0.008	259.600	259.602	0.002

(a)   (b)

Figure 8: The total estimated shifts in the *z*-direction (a) together with KSE (b) for the completed dataset.
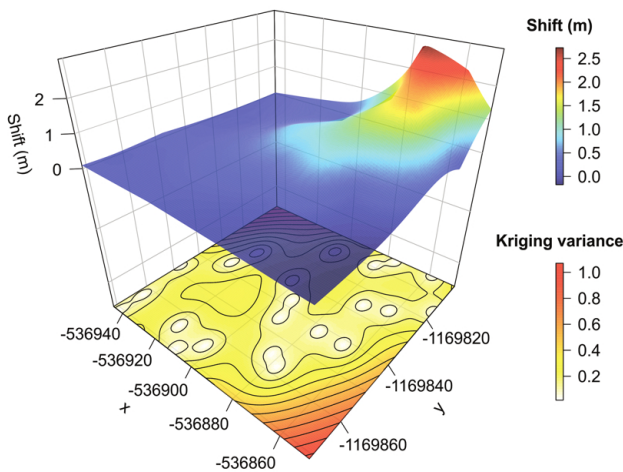

Figure 9: 3D model of the landslide in Halenkovice obtained from the completed dataset.

slightly at observed points for models without imputed values. It is obvious since there are fewer observations in

these models. The biggest difference between models with and without imputed values can be seen in the KSE map for *y*-direction where an important role of the imputed points 7, 10, 12, and 28 is noticeable. Their imputation to the model leads to a reduction of the KSE in the central part of the slope. The effect of imputed values is also apparent in predictive performance of fitted models (Tables 1 and 2). While the results are satisfactory for fitted models with imputed values, the results are much weaker for models without imputed values.

4 Discussion

Kriging is a stochastic method, which, unlike the deterministic methods, involves an element of randomness, including measurement and model errors. Thus, the resulting spatial prediction is considered one of the many that could be created. The resulting surface of kriging is smooth, so it may not necessarily respect the "true" sur-

(a)

(b)

(c)
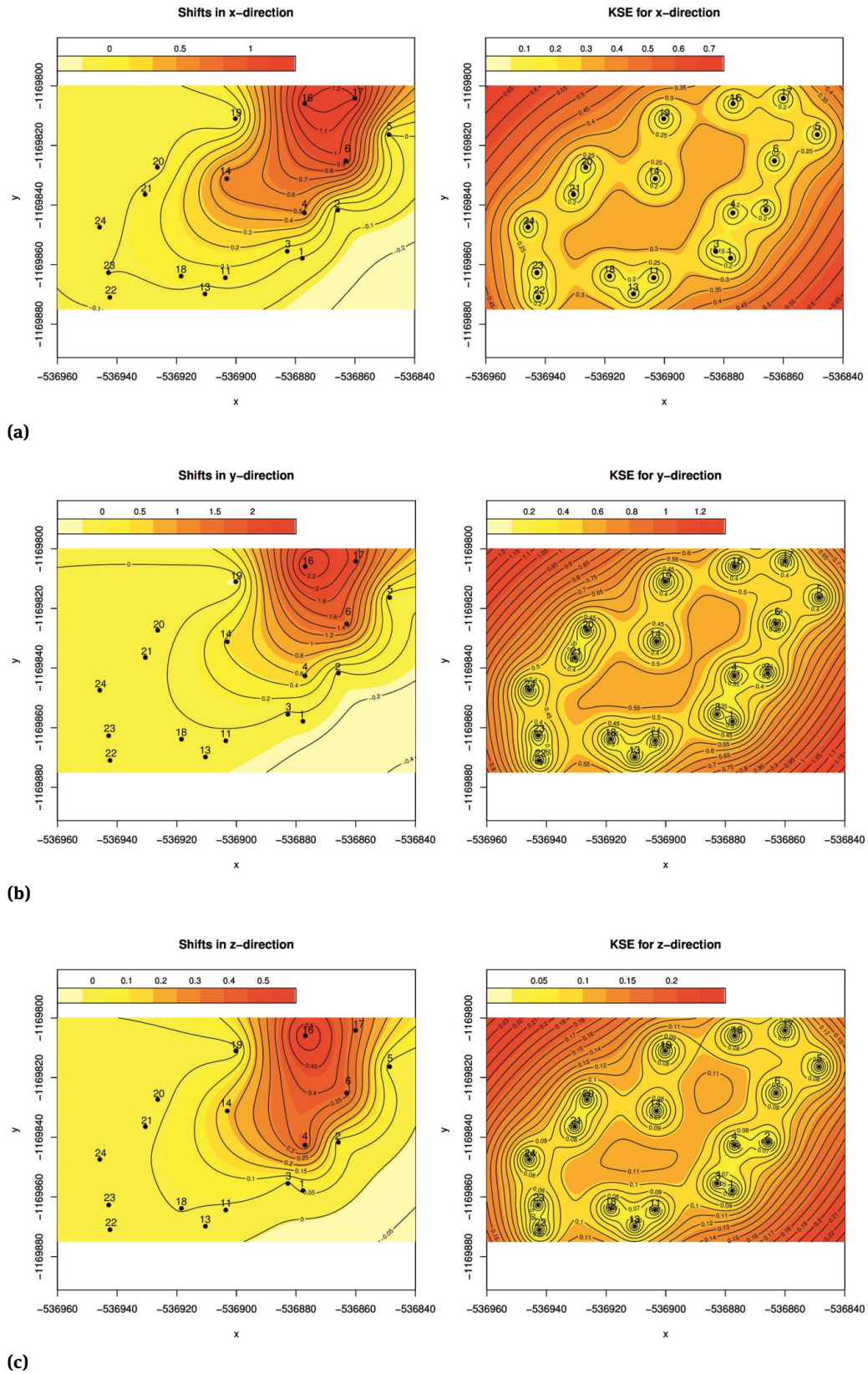
Figure 10: The total estimated shifts (left panels) in the *x*-direction (a), *y*-direction (b), and *z*-direction (c) together with KSE (right panels) for the original dataset.

face. The advantage of kriging consists of involving all the data information in the analysis. Weights do not depend only on the distance among the measured points and location of the predictor, but also on the spatial arrangement of measured points. Kriging estimates are optimal in the sense that they have a minimum degree of variability.

In the paper, we present the universal kriging in the context of the landslide modelling. When landslide varies in different directions, *i.e.* anisotropy occurs in data, each direction should be analysed separately, and it is necessary to construct kriging maps in *x*-, *y*- and *z*-directions separately. Construction of 2D or 3D maps requires the simultaneous analysis of *x*-, *y*-, and *z*-directions, what can be solved using the multivariate approach to modelling. Nevertheless, the multivariate approach is rather not simplistic and; therefore, this topic was left to further research. Since it is rather difficult to imagine separate kriging maps, an approximate 3D model is presented in Figure 9. However, for the more accurate model, both *x*- and *y*-directions should be analysed simultaneously.

The accuracy of kriging predictions depends on data configuration in the investigated area. A good measure of the spatial configuration of the data points is the KSE. In general, the KSE will increase as the variance of the data increases or data become more redundant. The KSE will decrease as data are closer to the location of the estimation. Therefore, in order to use all the information from the data, the algorithm for imputation of missing values based on the kriging as well is also proposed. The effect of imputed values is not only in a reduction of the KSE but also in predictive performance of fitted models. However, the KSE is essentially independent of data values used in the estimation. It is purely the function of the spatial distribution and the data configuration, and the version of the kriging used. The only link between the KSE and data values is through the semivariogram. Using the same variogram and the same data points will give the same KSE, regardless of the values of the data points. Hence, better fitted semivariogram model leads to lower KSE and better kriging estimates. In our case, the Gaussian model had the smallest RMSE in all directions (4.9 cm, 17.3 cm, and 1.8 cm in the *x*-, *y*-, and *z*-direction, respectively).

The limited number of previous studies related by both, the topic and the location, can be found. The results can be compared only to findings published in [18, 19], which provides approaches for the modelling of the same locality as is presented in this paper. The improvement of the recent study is the development of suitable imputation method that allows improving analysis in a case of missing data. The retrospective overview of land-sliding processes in the area is provided by [36]. Although this paper mentions landslide analysed in this study, it misses statistical aspect that is provided by the authors. Other studies [37–39] may then advise on imputation method but its results are either general or in the different field of study, so the direct comparison is not possible.

5 Conclusions

Kriging methods are usually used for modelling of mineral deposits. In the paper, the universal kriging is presented in the context of landslide modelling. We discuss main issues of the kriging such as isotropy and stationarity of studied process, the accuracy of predictions and a problem of missing values. When a studied process is not stationary, and a polynomial trend occurs in a data, it is appropriate to use the universal kriging. In the case of presence of anisotropy (landslide varies in different directions), it is necessary to construct kriging maps in *x*-, *y*- and *z*-directions separately. The accuracy of kriging predictions depends on data configuration in the investigated area. In order to use all the information from the data, the algorithm for imputation of missing values based on kriging as well is also proposed. The methodology was verified on landslide modelling in Halenkovice, Czech Republic. The resulting kriging maps confirmed the fact that the biggest shifts occurred at the lower part of the slope, in so-called accumulation zone. The results are also satisfactory from the statistical point of view. The most important benefit of our contribution is that kriging methods can be successfully utilised also for the landslide modelling.

References

[1] Van Westen C.J., van Asch T.W.J., Soeters R., Landslide hazard and risk zonation – why is it still so difficult? Bull Eng Geol Environ, 2006, 65, 167–184

[2] Mergili M., Marchesini I., Rossi M., Guzzetti F., Fellin W., Spatially distributed three-dimensional slope stability modelling in a raster GIS. Geomorphology, 2014, 206, 178–195

[3] Naydenova V., Roumenina E., Monitoring the mining effect at drainage basin level using geoinformation techniques. Cent Eur J Geosci, 2009, 1, 318–339

[4] Catani F., Casagli N., Ermini L., Righini G., Menduni G., Landslide hazard and risk mapping at catchment scale in the Arno River basin. Landslides, 2005, 2, 329–342

[5] Iverson R.M., Reid M.E., Iverson N.R., LaHusen R.G., Logan M., Mann J.E., Brien D.L., Acute sensitivity of landslide rates to initial soil porosity. Science, 2000, 290, 513–516

[6] Larsen M.C., Simon A., A rainfall intensity-duration threshold for landslides in a humid-tropical environment, Puerto Rico. Geogr Ann Ser a-Physical Geogr, 1993, 75, 13–23

[7] Hong Y., Adler R., Huffman G., Use of satellite remote sensing data in the mapping of global landslide susceptibility. Nat Hazards, 2007, 43, 245–256

[8] Pardeshi S.D., Autade S.E., Pardeshi S.S., Landslide hazard assessment: recent trends and techniques. Springerplus, 2013, 2, 1–11

[9] Mongillo M., Choosing basis functions and shape parameters for radial basis function methods. SIAM Undergrad Res Online, 2011, 4, 190–209

[10] Apaydin H., Kemal Sonmez F., Yildirim Y.E., Spatial interpolation techniques for climate data in the GAP region in Turkey. Clim Res, 2004, 28, 31–40

[11] Li J., Heap A.D., A Review of Spatial Interpolation Methods for Environmental Scientists. Aust Geol Surv Organ, 2008, GeoCat# 68, 154

[12] Bugya T., Fábián S., Görcs N., Kovács I., Radvánszky B., Surface changes on a landslide affected high bluff in Dunaszekcső (Hungary). Open Geosci, 2011, 3, 119–128

[13] Park I., Lee J., Saro L., Ensemble of ground subsidence hazard maps using fuzzy logic. Cent Eur J Geosci, 2014, 6, 207–218

[14] Ziary Y., Safari H., To Compare Two Interpolation Methods: IDW, Kriging for Providing Properties (Area) Surface Interpolation Map Land Price. District 5, Municipality of Tehran area 1. In Proceedings of FIG Working Week. Hong Kong, 2007, 13

[15] Rusu C., Rusu V., Radial Basis Functions Versus Geostatistics in Spatial Interpolations. In Artificial Intelligence in Theory and Practice. Volume 217. Springer US, 2006, 119–128

[16] Armstrong M., Basic Linear Geostatistics. Berlin: Springer-Verlag, 1998.

[17] Hatvani I.G., Magyar N., Zessner M., Kovács J., Blaschke A.P., The Water Framework Directive: Can more information be extracted from groundwater data? A case study of Seewinkel, Burgenland, eastern Austria. Hydrogeol J, 2014, 22, 779–794

[18] Marek L., Tuček P., Marek J., Pászto V., Stochastic approach for determining landslide activity. In Advances in Geoinformation Technologies 2010. Edited by Horák J, Halounová L, Hlásný T, Kusendová D, Voženílek V. Ostrava: VŠB-TUO, 2010

[19] Marek L., Tuček P., Pászto V., Marek J., Hypotheses testing in regression models and their using in landslide detection. In GIS Ostrava 2010. Ostrava: VŠB-TUO, 2010, 10

[20] Prusova M., Orlikova L., Hanzlova M., Comparison of geostatistical approaches dealing with the distribution of snow. Open Geosci, 2012, 4, 603–613

[21] Isaaks E.H., Srivastava R.M., An Introduction to Applied Geostatistics. New York: Oxford University Press, 1989

[22] Cressie N.A.C., Statistics for Spatial Data. New York: John Wiley & Sons, 1993

[23] Cressie N., Kriging Nonstationary Data. J Am Stat Assoc, 1986, 81, 625–634.

[24] Boisvert J.B., Manchuk J.G., Deutsch C.V., Kriging in the presence of locally varying anisotropy using non-euclidean distances. Cent Comput Geostatistics, 2009, 41, 585–601

[25] Chen S., Zhang J., Lu X., Chou K., Oates W., Schmidfetzer R., Chang Y., Calculation of rose diagrams. Acta Mater, 2007, 55, 243–250

[26] Bagheripour P., Asoodeh M., Nazarpour A., Vertical resolution enhancement of petrophysical Nuclear Magnetic Resonance (NMR) log using ordinary kriging. Cent Eur J Geosci, 2013, 5, 450–456

[27] Kovács J., Korponai J., Székely Kovács I., Hatvani I.G., Introducing sampling frequency estimation using variograms in water research with the example of nutrient loads in the Kis-Balaton Water Protection System (W Hungary). Ecol Eng, 2012, 42, 237–243

[28] Matheron G., Principles of geostatistics. Econ Geol, 1963, 58, 1246–1266

[29] Morgan C.J., Theoretical and practical aspects of variography: In particular, estimation and modelling of semi-variograms over areas of limited and clustered or widely spaced data in a two-dimensional South African gold mining context. PhD thesis, University of the Witwatersrand, Johannesburg, 2011

[30] Freedman D., Pisani R., Purves R., Statistics. New York: W.W. Norton & Company, 2007

[31] Stone M., Cross-Validatory Choice and Assessment of Statistical Predictions. J R Stat Soc, 1974, 36, 111–147

[32] Yao Y., Ye L., Tang H., Tang P., Wang D., Si H., Hu W., Van Ranst E., Cropland soil organic matter content change in Northeast China, 1985-2005. Open Geosci, 2015, 7, 234–243

[33] Šámalíková M., Locker J., Pospíšil P., Geology for Civil Engineers. Brno: VUT Brno, 1995 (in Czech)

[34] Marek L., Miřijovský J., Tuček P., Pászto V., Surveying and analysis of the landslides using UAS remote sensing. In SGEM2014 Conference Proceedings, 2014, 825–832

[35] Ribeiro P.J., Diggle P.J., geoR: Analysis of Geostatistical Data, R Package Version 1.7-5.1. 2015

[36] Bíl M., Krejčí O., Bílová M., Kubeček J., Sedoník J., Krejčí V., A chronology of landsliding and its Impacts on the Village of Halenkovice, Outer Western Carpathians. Geografie, 2014, 4, 342–363

[37] Gelfand A.E., Zhu L., Carlin B.P., On the change of support problem for spatio-temporal data. Biostatistics, 2001, 2, 31–45

[38] Schneider T., Analysis of incomplete climate data: Estimation of Mean Values and covariance matrices and imputation of Missing values. J Clim, 2001, 14, 853–871

[39] Korup O., Stolle A., Landslide prediction from machine learning. Geol Today, 2014, 30, 26–33