

A COMPARISON OF GEOSTATISTICS AND FUZZY APPLICATION FOR DIGITAL ELEVATION MODEL

R. Sunila ^{a*}, and K.Kollo ^b

^aLaboratory of Geoinformation and Positioning Technology, Department of Surveying, Helsinki University of Technology, P.O.BOX 1200, FIN-02015, Espoo, Finland – Rangsim.Sunila@tkk.fi

^bDepartment of Geodesy, Estonian Land Board, Mustamäe tee 51, 10602, Tallinn, Estonia – Karin.Kollo@maaamet.ee

KEY WORDS: Geodesy, Geostatistical analysis, DEM, Kriging, Fuzzy Logic

ABSTRACT:

This paper is continuation work of a research started in 2005. In the previous research, variety of fuzzy digital elevation models was constructed and the best fit model was selected in order to present an alternative in modelling height information. The previous study revealed that fuzzy digital elevation model gave satisfied result compared with the result from TIN (Triangulated Irregular Network). Fuzzy concepts and theories deal with ideas based on possibility approach. This fact leads us to the idea of attempting to apply geostatistical method in Digital Elevation Model as geostatistical theories are based on probability concepts. Kriging is one of the most well-known geostatistical methods and it is said to be an optimal interpolation. Therefore, kriging was selected for this study. There are various kriging techniques but in this study, ordinary kriging is our main concern. Kriging model construction and interpolation are our main focus. Our assumption was simple that we would like to compare resulting maps produced from different techniques such as fuzzy application, geostatistics and TIN. The resulting maps from fuzzy modelling and TIN were ready from our previous research. In this paper, we created a kriged interpolation map in order to include it to the final comparison. In additional, we discuss the strengths and weaknesses of different methods for better understanding of model selection.

1. INTRODUCTION

The models based on strong mathematical background are behind the ability to form pictures of geographical information. In modern geodesy, existent computer-based programs are used as well as created suitable models for data computation. Digital Elevation Data are essential products from geodetic measurements. To achieve high quality data of vertical network needs effort because the method, called levelling is very time consuming and expensive. There are various applications to present topographic information and DEM (Digital Elevation Model) is one of those well-known models in geodesy and cartography.

DEM is often based on GRID (regular raster model) or TIN (Triangulated Irregular Network). Seemingly, TIN is more preferably. Nevertheless, TIN has some disadvantages such as each prediction depends on only three data; it makes no use of data further away, and there is no measure of error. The resulting surface has abrupt changes in gradient at the margins of the triangles so the resulting surface is discontinuous which make a map with non-smooth isolines (Webster and Oliver, 2001). As a result, alternate models based on possibility theories were constructed in order to introduce new approach in viewing coordinate information. Fuzzy concepts were then brought into focus.

Fuzzy approach is based on the premise that key elements in human thinking are not just numbers but can be approximated to tables of fuzzy sets, or in other words, classes of objects in which the transition from membership to non-membership is gradual rather than abrupt (Sankar and Dwijesh, 1985). Unlike crisp sets that allow only true or false values, fuzzy sets allow

membership functions with unclear defined boundaries. The grade of membership is expressed in the scale of 0 to 1 and it is a continuous function (Sunila et al, 2004). To whatever extent, the use of fuzzy methods simplifies the mathematical models, which are usually used in geodetic applications. These mathematical models may be very simplified mathematical polynomials up to very high degree, which cannot be computed and visualised with a simple method (Kollo and Sunila, 2005).

Geostatistics is a subject concerned with spatial data. That is, each data value is associated with a location and there is at least an implied connection between the location and the data value (Jesus, 2003). There are various methods in geostatistics. All of those have different approaches to fit with various kinds of data and environment of model design. The geostatistical methods also provide error measures. Unlike TIN model, geostatistical model for example, kriging provides great flexibility of interpolation that yields smoothly surface. Where fuzzy model based on possibility theories may not be suitable for solving all problematic cases, probability theories then arise here to provide better alternative for reasonable modelling.

The aims of this paper were to present alternative in modelling digital elevation data using geostatistical method, such as kriging, to compare results from using different methods in DEM and provide possibility of choices to get height data for use in different lower accuracy geodetic and cartographic applications. The expected results will be the knowledge of using geospatial methods in geodesy, geostatistical models and the analysis of the use of geostatistical methods as alternative for providing height information. The results from kriging technique from this research and fuzzy DEM and TIN from the previous research will be compared and discussed.

* Corresponding author

2. KRIGING

2.1 Theoretical background

Kriging is a term coined by G. Matheron in 1963 after the name of D.G. Krige. Kriging is based on a statistical model of a phenomenon instead of an interpolating function. It uses a model for a spatial continuity in the interpolation of unknown values based on values at neighbouring points (Sunila et al., 2004). Kriging is called an optimal method because the interpolation weights are chosen to provide for the value at a given point the Best Linear Unbiased Estimate (BLUE) (Jesus, 2003).

There are several kriging techniques for different purposes such as ordinary kriging, simple kriging, universal kriging, indicator kriging, cokriging, point kriging, block kriging, disjunctive kriging, bayesian kriging and so on. In this research, the authors would like to focus on ordinary kriging as it is by far the most common type of kriging in practice.

Ordinary kriging is a variation of the interpolation technique, one that implicitly estimates the first order component of the data and compensates for this accordingly. This technique enables interpolation without the necessity of explicitly knowing the first order component of the data a priori (GIS dictionary, 1999). The basic equation used in ordinary kriging is as follows: *(Note: variables denoted in the equation are mentioned for the first time and they will be used throughout the text.)*

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (1)$$

where n = number of sample points
 λ_i = weights of each sample point
 $z(x_i)$ = values of the points

When the estimate is unbiased, the weights are made to sum to 1 or $\sum_{i=1}^n \lambda_i = 1$. The prediction variance is given by:

$$\sigma^2(x_0) = \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) + \phi \quad (2)$$

where σ^2 = variance
 $\gamma(x_i, x_0)$ = semivariance between sample point x_i and unvisited point x_0
 ϕ = Lagrange multiplier

2.2 Variogram Models

A variogram is a geostatistical technique which can be used to examine the spatial continuity of a regionalized variable and how this continuity changes as a function of distance and direction. Computation of a variogram involves plotting the relationship between the semivariance ($\gamma(h)$) and the lag distance (h) (Iacozza and Barber, 1999). The variogram is an essential step on the way to determining optimal weights for interpolation (Burrough and McDonnell, 1998).

The most commonly used variogram models are spherical, exponential, linear and Gaussian. Other models are such as Pentaspherical model and Whittle's elementary correlation and Pure nugget which they are left out in this research. The four

commonly used models that were mentioned earlier are described and explained in the question (3), (4), (5), and (6).

2.2.1 Spherical model: The spherical function is one of the most frequently used models in geostatistics (Webster and Oliver, 2001). The spherical model is good choice when the nugget variance is important but not too large, and there is a clear range and sill (Burrough and McDonnell, 1998).

$$\gamma(h) = \begin{cases} c_0 + c_1 \left\{ \frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right\} & \text{for } 0 < h < a, \\ c_0 + c_1 & \text{for } h \geq a \end{cases} \quad (3)$$

$$\gamma(0) = 0$$

where $\gamma(h)$ = semivariance
 h = lag
 a = range
 c_0 = nugget variance
 $c_0 + c_1$ = sill

2.2.2 Exponential model: The exponential model is a good choice when there is a clear nugget and sill, but only a gradual approach to the range.

$$\gamma(h) = c_0 + c_1 \left\{ 1 - \exp\left(-\frac{h}{a}\right) \right\} \quad (4)$$

2.2.3 Linear model: It is a non-transitive variogram as no sill within the area sampled and typical attributes are varying at all scales.

$$\gamma(h) = c_0 + bh \quad (5)$$

where b = the slope of the line

2.2.4 Gaussian model: If the variance is very smooth and the nugget variance is very small compared to the spatially dependent random variation, then the variogram can often best fitted with Gaussian model (Burrough and McDonnell, 1998).

$$\gamma(h) = c_0 + c_1 \left\{ 1 - \exp\left(-\frac{h^2}{a^2}\right) \right\} \quad (6)$$

2.3 Variance and Standard deviation

To measure the variability or how spread out the data values are, the variance is computed as the summation of the square of the difference of each data value from its mean and divide by number of points or number of points - 1 for unbiased estimate of variance. The formula for computing variance is:

$$S^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1} \quad (7)$$

where S^2 = variance
 x_j = data value
 \bar{x} = mean of the data values

The standard deviation (S) is simply calculated by the square root of the variance.

$$S = \frac{\sqrt{(x - \bar{x})^2}}{n - 1} \quad (8)$$

In the unbiased estimator case, one can say that the minimum value of mean square error is the variance. Hence, the standard deviation is the minimum of the root mean square error.

2.4 Cross validation

Cross validation is a model evaluation method for checking validity of the spatial interpolation method used. There are several validation techniques e.g. holdout set method, 5-fold, 10-fold, K-fold and N-fold. In this study, we select N-fold or also known as leave-one-out cross validation method. In N-fold cross-validation, the data set is split into N subsets of roughly equal size. The classification algorithm is then tested N times, each time training with N-1 of the subsets and testing with the remaining subset (Weiss and Kulikowski, 1991).

3. CASE STUDY

In this research the coordinates and absolute heights of initial points are used. The area named Rastila in the eastern suburb of Helsinki was chosen. The whole area is about 2 square kilometres, where about 2000 laser scanning points are situated. Some points were eliminated due to errors in data collecting process. Summary statistics of the sample points are shown in Table 1.

Number of observations	Min	Max	Mean
1985	0	18	4.926

Table 1. Summary statistics of the sample points in Rastila

The plotting of sample points is shown in figure 1. The colour range presents the different heights from 0 to 18 metres.

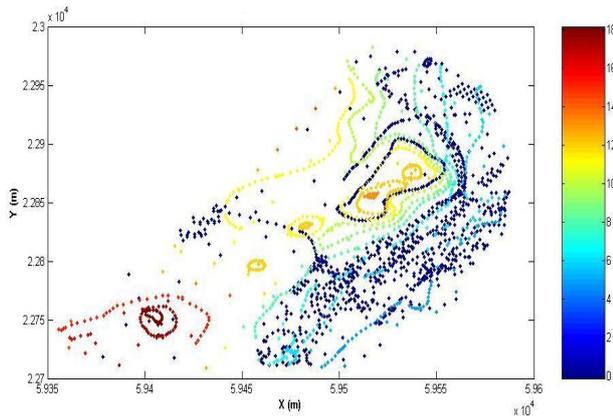


Figure 1. Plotted sample points of the study area in Rastila

The histogram in figure 2 shows a frequency of different heights in our sample points.

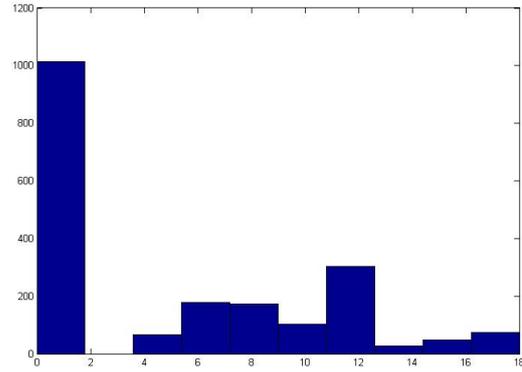


Figure 2. Histogram of the elevation sample points in the study area

3.1 Modelling the variogram

The elevation data we used in this study seem to be no directional differences so the separate estimates can be average over all directions and yield the isotropic variogram. As it was mentioned in section 2.2 that there are four most commonly used variogram models, therefore in this study, the authors limited the scope of study to focus on analysing those variogram models which are spherical, exponential, linear and Gaussian. It is questionable that what model is suitable for modelling our sample data. In order to solve this question, we decided to construct potential variogram models based on different parameters and compare them then, select the best-fit one for our analysis.

The model parameters with different ratio of nugget, sill, length, and range that are used in computation of candidate variogram models are shown in table 2.

Variogram model	range	nugget	sill	length
Exponential	0.5	0.10847	1.1232	0.18064
Linear	0.5	0.50754	1.806	-
Gaussian	0.5	0.4546	1.0447	0.22811
Spherical	0.75	0.35265	1.0666	0.48576
Exponential	0.75	0.13161	1.1388	0.19144
Linear	0.75	0.58671	1.5711	-
Gaussian	0.75	0.48089	1.0728	0.24717
Spherical	0.95	0.54285	1.9597	1.8513
Exponential	0.95	0.36872	1.4082	0.40794
Linear	0.95	0.55676	1.6424	-
Gaussian	0.95	0.57831	1.1922	0.34484

Table 2. Parameters of variogram models

It is noticeable that during the fitting model computational process, the spherical model at range 0.5 gave error due to the spherical curve certainly not fit into the model so it was left out of our interest.

3.2 Selecting the best fit model

To select the suitable model for modelling our observations, the standard deviation or the sum of square error guided us how well the model with specified parameter values displayed fits the empirical semivariogram. Using parameters in section 3.1,

the resulting standard deviation for each estimated model is shown in table 3.

Variogram model	Std
Exponential (0.5)	3.281
Linear (0.5)	3.367
Gaussian (0.5)	3.431
Spherical (0.75)	3.307
Exponential (0.75)	3.272
Linear (0.75)	3.397
Gaussian (0.75)	3.439
Spherical (0.95)	3.381
Exponential (0.95)	3.302
Linear (0.95)	3.386
Gaussian (0.95)	3.466

Table 3. The standard deviation values calculated from the varied variogram models

From table 3, it can be seen that the best-fit variogram model for our study is the exponential variogram model with the range 0.75, nugget = 0.13161, sill = 1.1388 and length = 0.19144. The chosen variogram model, exponential, was then drawn in figure 3.

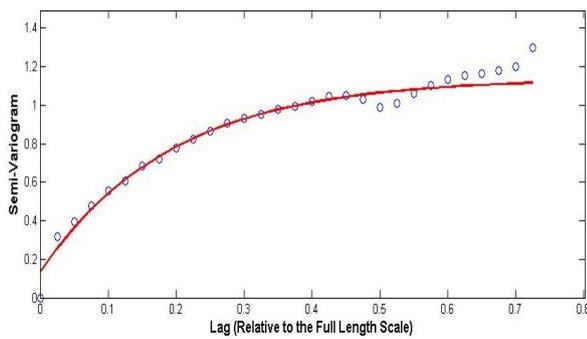


Figure 3: The exponential variogram model

After choosing variogram model, kriging method was implemented to estimate and interpolate the data. As it was stated in section 2.1 that ordinary kriging is our focus in this study, the ordinary kriging technique was then use in computation for estimation and interpolation. The search radius is 0.3 where the minimum number of kriging points is 10 and maximum is 30. As a result, kriging interpolation result from the chosen exponential model is displayed with krig map in figure 4.

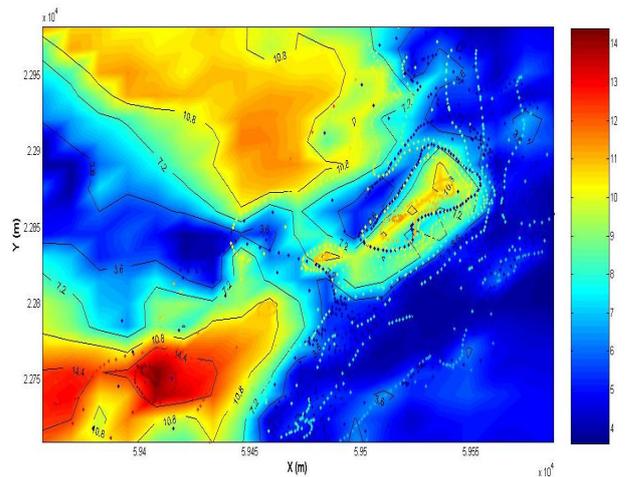


Figure 4. Kriging interpolation map of the study area

The variance map displays the krig variance normalised by the variance from the covariance function. The resulting variance map is shown in figure 5.

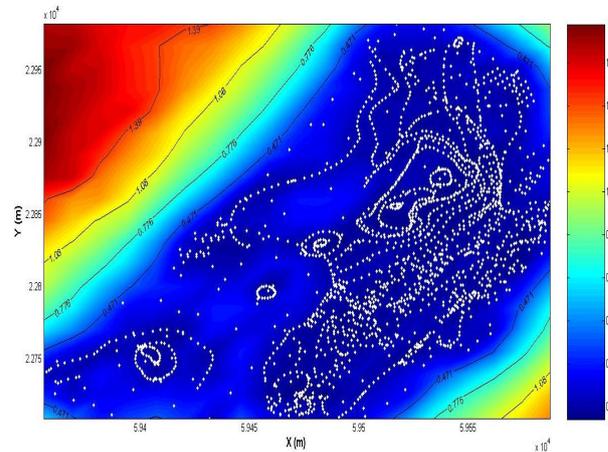


Figure 5. Variance map

3.2.1 The leave-one-out cross validation technique is used to present observation values and prediction values which are plotted in cross validation graph, figure 6. The validation reveals also the standard deviation or the sum of square error.

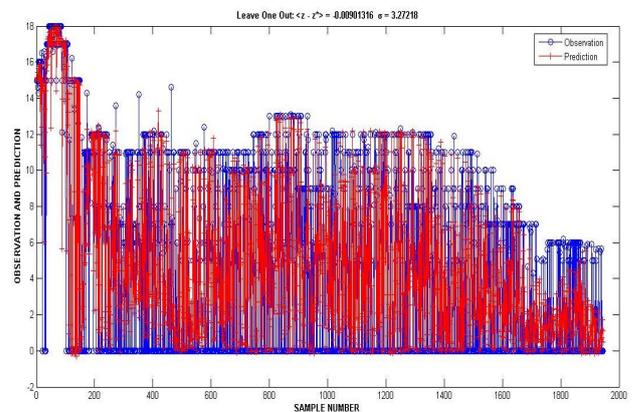


Figure 6. Cross validation of exponential variogram model

3.3 Comparison with models from the previous study

From the previous study (Kollo and Sunila, 2005), three different Fuzzy models were computed: firstly, the constant grid partition method with Gaussian membership function with number of membership functions 12, named as Model 1. Secondly, the linear grid partition method, trapezoidal membership function with number of membership functions 9, named as Model 2. Third, the Subclustering method with the range of influence being 0.2, named as Model 3. The RMSE values for Fuzzy DEM models were 4.45, 4.38 and 4.21 for Models 1, 2 and 3 respectively. From these three models the Model 3 (Subclustering method) was chosen for final comparison with TIN model.

From this research, ordinary kriging using exponential variogram model gives us least RMSE, 3.27. The result is then compared with the results from the previous fuzzy DEM study. The comparison of RMSE of the both methods is then shown in table 4.

Method/model	RMSE
Fuzzy/Subclustering	4.21
Ordinary kriging/Exponential	3.27

Table 4. The RMSE comparison between fuzzy model and kriging model.

The results from table 4 are obviously shown that kriging model gives us less RMSE value than fuzzy model. It can be explained that because our data contain sufficient points that benefit for geostatistical modelling in this case, kriging.

It should be mentioned here that the resulting map from TIN model was constructed by using ArcGIS as a completed product. TIN map is, however, not in our main area of interest in this research. It is, therefore, used as a comparison for map visualisation to give us better viewing of different results obtained from various modelling methods.

The visualisation of the resulting interpolation maps constructed from three different methods based on TIN, fuzzy and ordinary kriging is shown in figure 7.

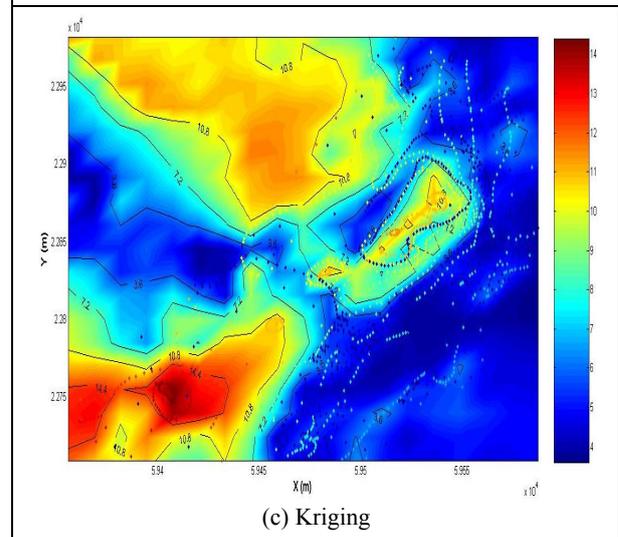
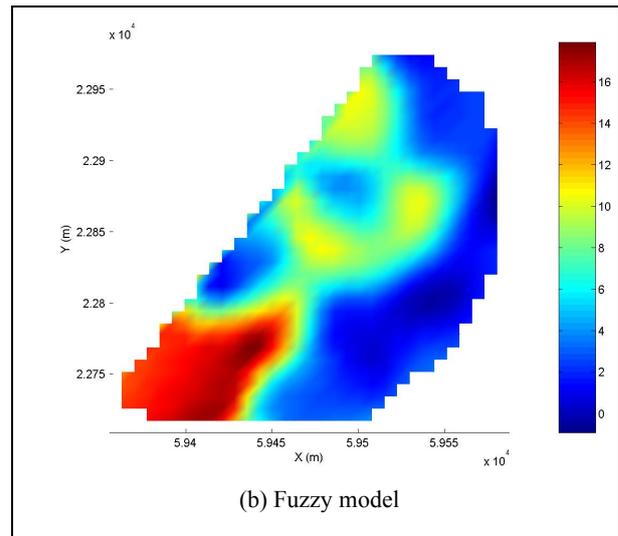
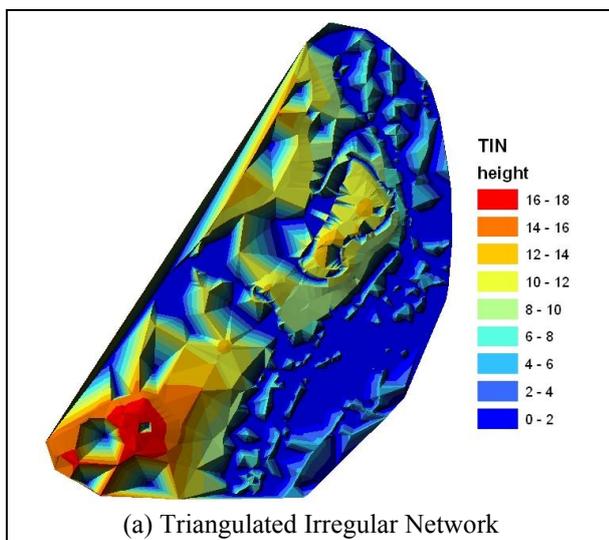


Figure 7. Comparison of resulting maps from different techniques used in DEM: TIN, Fuzzy model and Kriging

4. DISCUSSION

The study provided different alternatives to model height surface. Three different kinds of methods, Triangulated Irregular Network, Fuzzy theories and techniques and geostatistical method: ordinary kriging. From this research, advantages and disadvantages from these techniques were listed out and gathered for the discussion.

Triangulated Irregular Network (TIN) has its advantage that the technique is simple so it does not require much time for constructing the TIN map. However, some difficulties are found with TIN. Although the surface of resulting map is continuous, the map is presented with non-smooth isolines. TIN uses only three data so no further away data are included. TIN contains no measure of error.

Fuzzy technique, as the theories use concepts of possibility, allow us to visual and model the data in a new approach. The advantages of fuzzy in DEM are denseness of dataset is not necessary. Based on the data we have, fuzzy approach gave us very promising results. The computational algorithm does not require much work but can be time consuming. A disadvantage

of fuzzy DEM that we come across is the resulting map has no continuous extrapolation or it is vague.

Kriging technique based on probability theories provides us many advantages in modelling elevation data. It is possible to use kriging technique to calculate missing locations. The result gave us an estimate of potential error. Kriging interpolation is smooth and visualisation of the surface is good. Fortunately, our sample data are efficient; the resulting map looks very reliable. The disadvantages of kriging are, for example, fitting variogram models can be time consuming because it requires skill and judgement. The computational process is also demanding. Kriging will not work well if number of sample points are too less. Fitting model can be difficult if behaviour of data points are extreme noisy. Table 5 concludes advantages and disadvantages of both fuzzy and kriging methods.

	Fuzzy	Kriging
Advantages	<ul style="list-style-type: none"> + Data need not have to be dense. + Algorithm is ready and easy to use. + Visualisation is smooth and interpolation looks real. 	<ul style="list-style-type: none"> + Missing points are possibly calculated. + Potential error measure can be estimated. + Visualisation is smooth and interpolation looks real.
Disadvantages	<ul style="list-style-type: none"> - Possibly require more time for model construction. - The models are dependent on data structure. - The extrapolation is vague. 	<ul style="list-style-type: none"> - Demand time and effort in computational process. - Fitting variogram models can be time consuming - Fitting variogram model requires skill and effort if data points are inadequate. - Does not work well with too few data points. - Can be expensive regarding data collection method.

Table 5. Advantages and disadvantages of fuzzy and kriging methods.

5. CONCLUSIONS

The study aimed to present an alternative method for constructing DEM models to be used in geodetic and cartographic applications. The kriging method from geostatistical approach was chosen to test our hypothesis. Among various kriging methods, the ordinary kriging was chosen. Then, four commonly used variogram models which are spherical, exponential, linear and Gaussian, were brought into our variogram model construction. To try to search for the best-fit variogram model, we generated various models based on different parameters and assumptions to provide wider alternatives in selecting the final model. Next, the best-fit model was selected. Our computations showed that the exponential model was suitable with our data, giving us the less statistical error measure than other models. Then, ordinary kriging technique was applied for the krig interpolation map. The variance map was also constructed as well as cross validation analysis for finalised RMSE value.

We can conclude that among the three interpolation techniques in our research, kriging approach gives us better error measure so the height surface, produced by means of kriging gives also better visualisation of the output. When number of observations

is adequate, kriging technique seems to provide better fit in elevation surface modelling in terms of realistic visualisation of the results than fuzzy approach.

6. FURTHER RESEARCH

The authors have positive opinion that the research is very interesting. Comparison of various methods based on different theories and approaches are useful for model construction. Each method or technique has its own advantages and disadvantages. It gives us alternatives for choosing a suitable model for data analysis. Yet, a best-fit model for one kind of data may not fit with another kind of data. In the future research, the authors aim to study a new model construction or selection of regions with different terrain forms.

7. REFERENCES

Burrough, P. and McDonnell, R., 1998. *Principles of Geographical Information Systems*. Oxford University Press, pp. 132 – 161.

GIS dictionary, 1999. Association for geographic information. <http://www.geo.ed.ac.uk/agidexe/term?292>. (accessed 20 Jan. 2007)

Iacoza, J. and Barber, D., 1999. An Examination of the Distribution of Snow on Sea-Ice. *Atmosphere-Ocean*, 37(1), pp. 21 – 51.

Jesus, R., 2003. Kriging: An Accompanied Example in IDRISI. GIS Centrum University for Oresund Summer University, Sweden.

Kollo, K. and Sunila, R., 2005. Fuzzy Digital Elevation Model. In: *Proceedings of the 4th International Symposium on Spatial Data Quality '2005*, Beijing, China, pp. 144 – 151.

Sankar, K.P. and Dwijish, K.D.M., 1985, *Fuzzy Mathematical Approach to Pattern Recognition*. Wiley Eastern Limited.

Sunila, R., Laine, E. and Kremenova, O., 2004. Fuzzy Modelling and kriging for imprecise soil polygon boundaries. In: *Proceedings 12th International Conference on Geoinformatics – Geospatial Information Research: Bridging the Pacific and Atlantic*, Gävle, Sweden, pp. 489 – 495.

Webster, R. and Oliver, M., 2001. *Geostatistics for Environmental Scientists*. John Wiley & Sons, UK.

Weiss, S. and Kulikowski, C., 1991. *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*, Morgan Kaufmann, CA, USA.

8. ACKNOWLEDGEMENTS

The authors would like to thank Professor Kirsi Virrantaus for supporting our ideas and research, Mr. Matias Hurme from Department of Real Estate, Division of City Mapping, Helsinki City for providing the sample data and relevant information and sincerely thank to the 5th ISDQ committee for giving us an opportunity to present our work. Second author sincerely

thanks Kristjan Jaak Scholarship Foundation and Estonian Land Board for providing the scholarship.