Non-stationary modelling and simulation of LIDAR DEM uncertainty

Juha Oksanen* and **Tapani Sarjakoski**
Department of Geoinformatics and Cartography
Finnish Geodetic Institute
Masala, Finland
*juha.oksanen@fgi.fi

Abstract— **Appropriate modelling and simulation of digital elevation model (DEM) uncertainty is among the most long-lasting of topics in geographical information science, because DEMs and terrain analysis are widely used in tasks with high societal impact. Decisions based on the analysis are expected to be of better quality if the uncertainty of the analysis results is taken into account. Despite the long research history of the topic, a few big challenges have decelerated the final break-through of the uncertainty-aware terrain analysis. Firstly, the utilisation of Monte Carlo simulation, which is the most flexible method for investigating the propagation of uncertainty in terrain analysis, is time-consuming. Moreover, the use of massive high-resolution DEMs based on airborne light detection and ranging (LIDAR) has made the performance issue even worse. Secondly, mainstream uncertainty-aware terrain analysis is done by applying stationary models of DEM uncertainty, even though the number of experiments has proven that the uncertainty would be modelled more realistically as a non-stationary stochastic process. The paper demonstrates how the process convolution method can be applied in a realistic and efficient non-stationary simulation of LIDAR DEM uncertainty.**

Keywords: process convolution, unconditional simulation

I. INTRODUCTION

The aim of the uncertainty-aware terrain analysis is to improve the quality of decisions that are based on the analysis results. Uncertainty-awareness helps to minimise the risk of false decisions in issues related to the environment and natural resources (Brown and Heuvelink 2007). The importance of analysis taking into account the imperfections of the input DEMs and the used models has been acknowledged in the geospatial research community and studied for decades (e.g. Fisher 1991, Kyriakidis et al. 1999, Oksanen and Sarjakoski 2006). Previous research related to uncertainty-aware terrain analysis has mostly focused on the use of the Monte Carlo method and stationary Gaussian error models (e.g. Hunter and Goodchild 1997, Holmes et al. 2000, Oksanen and Sarjakoski 2005), which means that the modelled uncertainty has been homogeneous in space (Goovaerts 1997). With the Monte Carlo method, the idea is to run the terrain analysis task a number of times using the sum of the DEM and a single realisation of the DEM uncertainty model, and, finally, to make a summary of the results (e.g. Fisher 1991). However, stationary modelling of uncertainty over large regions has been known to be a gross generalisation of a complex reality (e.g. Fisher 1998, Carlisle 2005, Oksanen and Sarjakoski 2006). Also, the appropriateness of the commonly used method, in which the realisation of the DEM error model is added to the DEM as part of the simulation process, has been questioned (Hengl et al. 2008) and use of the method needs to be justified.

Two important factors appear to explain the lack of scientific knowledge about the use of LIDAR DEMs in uncertainty-aware terrain analysis. Firstly, the common belief has been that the high quality of LIDAR DEMs (e.g. Hodgson et al. 2005, Barber and Shortridge 2005, Vaze and Teng 2007) will make the uncertainty-aware terrain analysis unnecessary. Secondly, uncertainty propagation studies have typically made use of simulation methods, such as simulated annealing and sequential Gaussian simulation (Goovaerts 1997), that are unsuitable for massive data sets because of their poor scalability. Nevertheless, in threshold-based terrain analysis, such as flood area analysis and drainage basin delineation, even a small uncertainty in the DEM can result in significant "leaking" in the model at critical circumstances. Therefore, the research related to uncertainty-aware terrain analysis needs to be updated to accommodate wide-spread new data products (Fisher and Tate 2006) and to reflect the current knowledge about DEM uncertainty.

The paper demonstrates how location-dependent uncertainty of the LIDAR DEM can be modelled and simulated using a process convolution method, as well as how the uncertainty propagates in a catchment delineation based on LIDAR DEM.

II. METHODS

A. Non-Stationary Unconditional Simulation of LIDAR DEM Uncertainty

The convolution-based simulation method is suitable for uncertainty-aware terrain analysis due to its flexibility in modelling spatial dependence, as well as its performance on large data sets (Higdon et al. 1999). Process convolution can be considered as a special case of the spatial moving average method (Swall 1999, Kern 2000, Higdon 2002). It is possible to create the stationary Gaussian process $z(s)$ with pre-defined spatial autocorrelation, defined by the covariogram $c(s)$, over spatial domain S by convolving a Gaussian random process $x(s)$, $s \in S$ with a unimodal and symmetric kernel $k(s)$ (Higdon 2002). The solution for defining the appropriate kernel k lies in the fact that convolving a dense Gaussian noise with an appropriately scaled kernel yields a zero-mean Gaussian random field with a correlation structure that is the same as the convolution of the kernel with itself (Kern 2000). According to convolution theorem, the Fourier transform of the convolution of two functions is proportional to the product of the Fourier transformations of the individual functions. Thus, if the covariogram *c(s)* has been defined, the kernel that satisfies $k(s)*k(s) \propto c(s)$ may be obtained by identifying the Fourier transform $F[c(s)]$ of $c(s)$, and calculating the inverse Fourier transform of the square root of $F[c(s)]$ when k is continuous and integrable (Kern 2000) (Table I). To extend the stationary simulation to non-stationary simulation, we let the parameters of spatial autocorrelation model change over space (Figure 1). Thus, each location s in the S has its own covariogram model and kernel. In addition, the process convolution needs to be done for all s. Methods for representing and modelling local uncertainty model parameters are introduced in the following chapter.

B. *Location Dependent Parameters of DEM Uncertainty*

The challenge in non-stationary modelling and simulation of DEM uncertainty is how we define and model the location-dependent uncertainty model parameters. Such parameters could be, for example, local semivariogram parameters having an influence limited to an area where the stationary simulation is reasonable. For building a stochastic model of DEM uncertainty, it is useful to express the uncertainty by applying a regionalised variable theory (Burrough and McDonnell 1998). Thus, the outcome $z(s)$ of the stochastic process at location s is the sum of a structural component with a constant drift $m(s)$, a random but spatially correlated component $\varepsilon'(s)$, and a spatially uncorrelated Gaussian noise term ε'' (equation (1)).

$$z(s) = m(s) + \varepsilon'(s) + \varepsilon'' \qquad (1)$$

The original definitions of the variables may be relaxed for the purposes of DEM uncertainty modelling. For example, while the drift m(s) is strictly the systematic component of the uncertainty in the spatial domain, it may also contain systematic effects of the uncertainty in other, non-spatial, domains. An example of the effect is the possible correlation between DEM uncertainty and slope (Hodgson et al. 2005, Bater and Coops 2009).

TABLE I. MATLAB-CODE USED FOR CONVERTING COVARIOGRAM MATRIX C TO CONVOLUTION KERNEL K FOR PROCESS CONVOLUTION.

```
function k = createKernel(c)
%PURPOSE: Conversion of covariogram c to kernel k for
%         process convolution
%---------------------------------------------------
%USAGE: k = createKernel(Covariogram)
%where: [Covariogram] is the covariogram matrix.
%---------------------------------------------------
%EXAMPLE: E = createKernel(c);
%
%Juha Oksanen, Finnish Geodetic Institute, 2010.

%% Convert covariogram c to convolution kernel k
k = fftshift(ifft2(sqrt(real(abs(fft2(c))))));
end
```

It is possible to take location-dependent DEM uncertainty into account in a number of ways. Firstly, based on expert knowledge, the study area may be divided into regions, in which the uncertainty is assumed to be homogeneous. While the method is simple, appropriate delineation of the regions is challenging and discontinuities of the uncertainty model parameters along the region boundaries are unsatisfying for a number of global terrain analysis tasks. Furthermore, a region delineation approach generalises the spatial variation of the uncertainty model parameters. Secondly, it is possible to generate a spatially varying uncertainty model from the geometrical basis if the reference data covers the whole extent of the study area (Oksanen and Sarjakoski 2006). In such case, the assumption is that the model parameters change regularly between the locations where reference data and experimental uncertainty model parameters are available. Thirdly, it is possible to generate a simple or multivariate regression model between the uncertainty model parameters as the response variables and number of predictors (Carlisle 2005), such as slope and LIDAR point density. In the third option, the complexity of the model may be chosen based on cost-efficiency. In the simplest case, only one of the uncertainty model parameters (e.g. sill) changes in space. In a more complex non-stationary uncertainty model, the number of parameters, such as semivariogram model, sill, range, anisotropy ratio, angle of anisotropy, and nugget variance, may change locally (Figure 1).

III. CASE STUDY: MODELLING AND SIMULATION OF LIDAR DEM UNCERTAINTY IN THE VALUE-PROJECT

The VALUE-project aims at updating the Finnish drainage basin register. As part of the project, research on

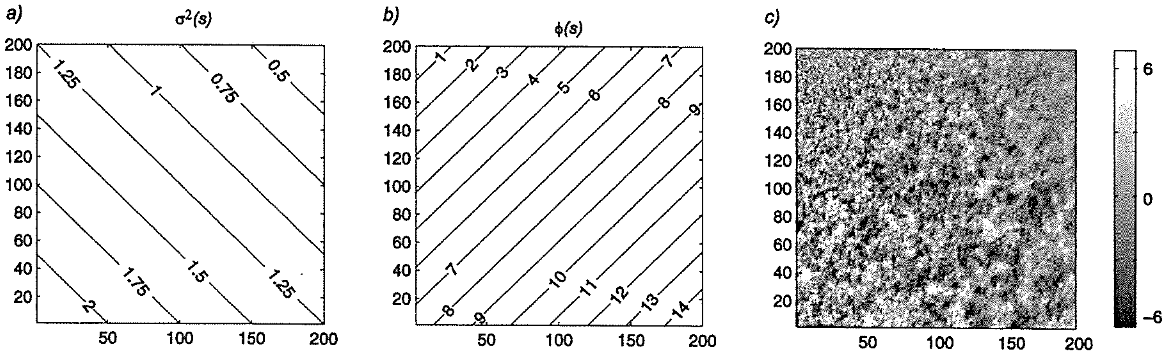

Figure 1. Realisation (c) of 200x200 unit non-stationary random field, in which sill (a) and range (b) changes over space.

uncertainty-aware terrain analysis has been carried out. The aim was to make use of a location-dependent, non-stationary error model for elevation data. The only location dependent error model parameter in the demonstration was variance; this was due to the simplicity of the modelling process as well as the efficiency of the simulation.

The section demonstrates how a simple non-stationary model of uncertainty can be derived for the forthcoming nationwide LIDAR DEM of Finland, NLS KM2 (NLS 2010), and applied in uncertainty-aware catchment delineation using the Monte Carlo method. The production steps for NLS KM2 are data collection, automatic point cloud classification (Axelsson 2000), manual correction of classification blunders in a stereo-workstation, and, finally, a surface approximation from the ground points to a 2 m grid using a finite element method (Ebner et al. 1980, NLS 2010). The study area is the Lakistonjoki drainage basin in southern Finland, from which the Finnish Geodetic Institute (FGI) has collected reference data for constructing the Nuuksio test environment (Sarjakoski et al. 2007) (Table II).

TABLE II. ELEVATION DATASETS USED IN THE STUDY.

Materials	Elevation data sets	
	DEM	*Reference data*
Producer	National Land Survey of Finland	Finnish Geodetic Institute
Data	NLS KM2 - LIDAR DEM in 2 m grid (NLS 2010)	2% random sample of the LIDAR ground points (Axelsson 2000)
Size	9 km * 12.3 km (27649206 cells)	n = 2112274
Scanning date	May 11th 2008	a) May 14th – 15th 2006 b) June 2007
Flying altitude	2000 m	a) 1100 m b) 1000 m
Average point density	1.1 points / m^2	a) 4.5 points / m^2 b) 8.9 points / m^2

A. Model Calibration and Validation

Construction of the DEM uncertainty model was split into two steps. In the first step, the elevation uncertainty of NLS KM2 (standard deviation of height differences) was calculated for each of the 1% slope bins derived from the same DEM. Regression analysis of the DEM error and slope with a high coefficient of determination indicated that most of the variation in local variance can be explained by slope (Figure 2).

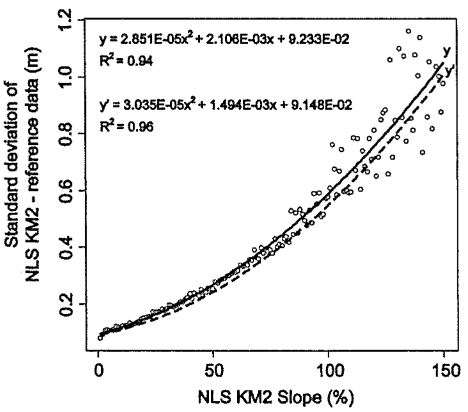

Figure 2. Observations and the regression model (y, solid line) of NLS KM2 slope and elevation uncertainty. Model validation (y') with independent data set is shown with the dashed line.

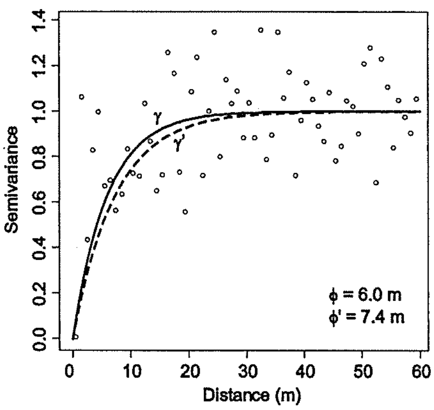

Figure 3. Experimental scaled semivariances and the exponential semivariogram model (γ, solid line) of NLS KM2 elevation uncertainty. Model validation (γ') with independent data set is shown with the dashed line. In the exponential model, the

In the second step, the slope effect was removed from the observations and a global semivariogram analysis was performed for the random subsample (n=10112) of the reference data to detect the shape and range of the spatial autocorrelation function. According to the analysis, an exponential semivariogram model having a practical range of 18 m was selected (Figure 3). Both steps could be done simultaneously in a single semivariogram analysis. However, when variance is the only location dependent variable, the simulation is more efficient since the solution of the convolution kernel is global and a local variation of the variance can be added to the realisation after the convolution.

For validation, another independent 2% random sample (n=2147885) and a subsample (n=9876) of the FGI's LIDAR ground points were taken and a single realisation of the uncertainty model was analysed. The regression and the semivariogram model parameters from the validation data set were close to the results from the calibration step (Figures 2 and 3).

B. Result and Conclusions

The result of the uncertainty-aware delineation of the Lakistonjoki catchment with 1000 simulation runs reveals that the observed small uncertainty of the NLS KM2 (LIDAR DEM) induces a variation of a few hundred meters in the location of the catchment boundaries (Figure 4).

According to the result, uncertainty-aware terrain analysis seems to offer valuable additional information, even with LIDAR DEMs. Furthermore, use of non-stationary modelling and simulation resolves the limitations of global uncertainty-model parameters. While the preliminary results of the method appear to be promising, conclusions about the influence of LIDAR DEM uncertainty must be drawn using caution. Firstly, in the case study the accuracy of the reference data was not an order of magnitude higher than the DEM's accuracy. Therefore, the uncertainty model combined the uncertainties of both data sets. Secondly, an unconditional simulation with process convolution makes summing of the

DEM and a single realisation of the uncertainty model necessary (indirect simulation), which is assumed to result in the final surface being too noisy (Hengl et al. 2008). The essential question is: Do we assume the DEM to be one equi-probable image of the random process representing the topography? Or, do we assume it to be a generalised representation of the topography? The first interpretation makes a direct geostatistical simulation of the DEM necessary, while the second interpretation, which was used in the case study, allows for the use of indirect simulation. As a consequence, the more the DEMs are based on direct measurements the less valid the indirect modelling and simulation of DEM uncertainty becomes. Simultaneously, the role of the error's spatial autocorrelation induced by surface approximation diminishes. Finally, the unexpectedly large width of the uncertain catchment boundary raises a question about the appropriateness of the modelling method and the analysis scale in general. Since conventional terrain analysis does not consider the meaningful scale of the phenomenon, there is a possibility that the analysis reveals spatial patterns that are irrelevant in reality.

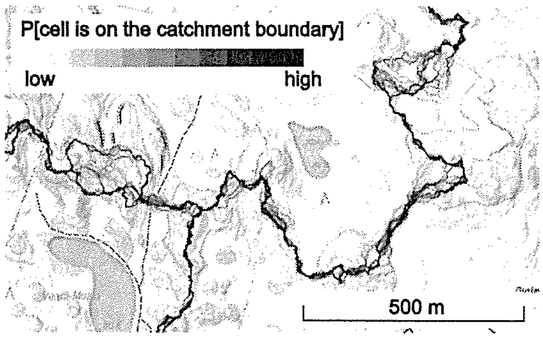

Figure 4. Detail of the Lakistonjoki's uncertainty-aware catchment delineation based on a non-stationary model of the NLS KM2

ACKNOWLEDGEMENTS

VALUE is a co-operative project of the Finnish Geodetic Institute and the Finnish Environment Institute. The project is funded by the Finnish Ministry of Agriculture and Forestry.

REFERENCES

Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. In Beek, K.J. and M. Molenaar (Eds) *International archives of photogrammetry and remote sensing, Vol. 33, Part B4/1* (pp. 110–117). Amsterdam, The Netherlands: GITC.

Barber, C.P. and A. Shortridge (2005). Lidar elevation data for surface hydrologic modeling: Resolution and representation issues. *Cartography and Geographic Information Science.* 32(4), 401-410.

Bater, C.W. and N.C. Coops (2009). Evaluating error associated with lidar-derived DEM interpolation. *Computers and Geosciences.* 35(2), 289-300.

Brown, J.D. and G.B.M. Heuvelink (2007). The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. *Computers and Geosciences.* 33(2), 172-190.

Burrough, P. and R. McDonnell (1998). *Principles of geographical information systems - spatial information systems and geostatistics.* Oxford: Oxford University Press.

Carlisle, B.H. (2005). Modelling the spatial distribution of DEM error. *Transactions in GIS.* 9(4), 521–540.

Ebner, H., Hofmann-Wellenhof, B., Reiss, P., and F. Steidler (1980). HIFI - A Minicomputer program package for height interpolation by finite elements. *International Archives of Photogrammetry, Vol. 23, Part B4, Commission IV* (pp. 202–215). Hamburg, Germany: Committee of the 14th International Congress for Photogrammetry.

Fisher, P.F. (1991). First experiments in viewshed uncertainty: The accuracy of the viewshed area. *Photogrammetric Engineering and Remote Sensing.* 57(10), 1321-1327.

Fisher, P.F. (1998). Improved modeling of elevation error with geostatistics. *GeoInformatica.* 2(3), 215-233.

Fisher, P.F. and N.J. Tate (2006). Causes and consequences of error in digital elevation models. *Progress in Physical Geography.* 30(4), 467-489.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* New York: Oxford University Press.

Hengl,T., Bajat, B., Blagojević, D., and H.I. Reuter (2008). Geostatistical modeling of topography using auxiliary maps. *Computers and Geosciences.* 34(12), 1886-1899.

Higdon, D. (2002). Space and space-time modeling using process convolutions. In Anderson, C.W., Barnett, V., Chatwin, P.C., and A.H. El-Shaarawi (Eds) *Quantitative Methods for Current Environmental Issues* (pp. 37–54). London, UK: Springer-Verlag.

Higdon, D.M., Swall, J., and J.C. Kern (1999). Non-stationary spatial modeling. In Bernardo, J.M., Berger, J.O., Dawid, A.P., and A.F.M. Smith (Eds) *Bayesian statistics 6: Proceedings of the 6th Valencia international meeting* (pp. 761–768). Oxford:Oxford University Press.

Hodgson, M.E., Jensen, J., Raber, G., Tullis, J., Davis, B.A., Thompson, G., and K. Schuckman (2005). An evaluation of lidar-derived elevation and terrain slope in leaf-off conditions. *Photogrammetric Engineering and Remote Sensing.* 71(7), 817–823.

Holmes, K.W., Chadwick, O.A., and P.C. Kyriakidis (2000). Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology.* 233(1–4), 154–173.

Hunter, G.J. and M.F. Goodchild (1997). Modeling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis.* 29(1), 35-49.

Kern, J.C. (2000). *Bayesian process-convolution approaches to specifying spatial dependence structure.* Ph.D. Thesis, Duke University, Institute of Statistics and Decision Sciences, Durham, NC.

Kyriakidis, P.C., Shortridge, A.M., and M.F. Goodchild (1999). Geostatistics for conflation and accuracy assessment of digital elevation models. *International Journal of Geographical Information Science.* 13(7), 677–707.

NLS (2010). Uusi valtakunnallinen korkeusmalli laserkeilaamalla (A new coutry-wide DEM by using airborne laser scanning). http://www.maanmittauslaitos.fi/Tietoa_maasta/Ilmakuvaus/Uusi_valtakunnallinen_korkeusmalli_laserkeilaamalla/. (17 March 2010)

Oksanen, J. and T. Sarjakoski (2005). Error propagation analysis of DEM-based drainage basin delineation. *International Journal of Remote Sensing.* 26(14), 3085–3102.

Oksanen, J. and T. Sarjakoski (2006). Uncovering the statistical and spatial characteristics of fine toposcale DEM error. *International Journal of Geographical Information Science.* 20(4), 345–369.

Sarjakoski, T., Sarjakoski, L. T. and R. Kuittinen (2007). Establishing a test environment for ubiquitous geospatial applications. *Proc. of XXIII International Cartographic Conference*, Cartography for everyone and for you, Moscow, Russia, August 4-10, 2007, Theme 13, CD-ROM.

Swall, J.L (1999). *Non-stationary spatial modeling using a process convolution approach.* Ph.D. Thesis, Duke University, Institute of Statistics and Decision Sciences, Durham, NC.

Vaze, J. and J. Teng (2007) High resolution LiDAR DEM – How good is it?. In Oxley, L. and Kulasiri, D. (Eds) *MODSIM 2007 international congress on modelling and simulation* (pp. 692-698). Modelling and Simulation Society of Australia and New Zealand.