Derivation of Quality Parameters for Terrain Models: Examples from Seafloor and Landscape Mapping

Jan T. Bjørke[1,2] and Stein Nilsen[1]

[1] Norwegian Defence Research Establishment,
PO Box 115, N-3191 Horten, Norway
jtb@ffi.no, sni@ffi.no
[2] Norwegian University of Life Sciences,
PO Box 5003, 1432 Ås, Norway

Abstract. Random measurement errors in digital terrain models are important to their quality. The present paper describes an untypical use of variograms to compute this kind of measurement errors. After subtracting a low frequency drift from depth/height measurements, a variogram analysis of the residuals is used to derive the random measurement errors of the observations. Numerical experiments on lasercanning and multibeam echo sounder data show how the semivariogram clearly identifies the amount of noise in the original data. The method can run in real time and does not require additional measurements like ground control patches or overlapping survey strips.

Keywords: Drift, global estimate, subdivision, nugget effect, error analysis.

1 Introduction

Multibeam echo sounders and airborne laserscanners have become very important instruments for the acquisition of digital seafloor and landscape data. These collection methods are characterized by high sampling density, and random errors in the dataset are frequently visible in contour maps and shaded relief maps. The rough contour lines and the relief model in Figure 1 demonstrates the effect of noisy depth data in a seafloor model.

The aim of this study is to demonstrate a fast and robust method for computing the random measurement errors in digital terrain models. When the noise component is estimated, it can be smoothed, but this kind of improvement of the model is outside the scope of the present study.

The computation of random measurement errors in digital terrain models can be based on least-squares matching (LSM) of a number of datasets or comparison of the terrain model to measurements we regard as true values. The collection of the true terrain model requires that we apply a surveying method which gives higher accuracy than the model we will investigate. The derivation of this model, therefore, can be time consuming and expensive; particularly in seafloor mapping projects. Maas [Maa02] computes systematic and random errors from LSM of a number of overlapping survey strips using a method that does not require knowledge of the true terrain model.

The novelty of our method is that we can derive information about the random error component from a single survey line. The clue is to exploit the high sampling rate of modern measurement systems. As long as the data density is high enough, our method for noise computation is independent of the technology used for the acquisition. Therefore, we have a uniform method which can be used on seafloor as well as landscape models. The procedure we use applies de-trending from average interpolating subdivision, and then statistical analysis of the residuals; see [BN05].

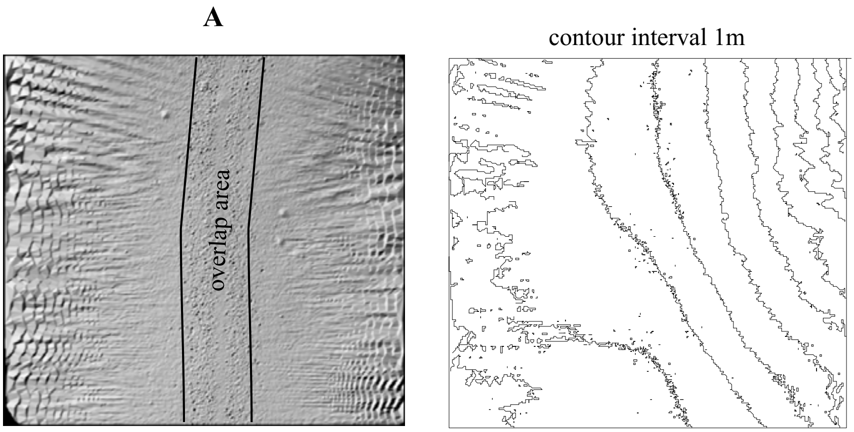

Fig. 1. The seafloor data illustrated as shaded relief and contour maps. The overlap area between the two transducers of EM3000 is indicated in the relief map.

2 The methodology

We consider terrain surfaces as realizations of random functions $Z = Z(x, \omega)$, where $x \in D \subset \mathbf{R}^2$. The letter ω denotes the particular realization, or outcome. For a fixed value $x', Z(x', \omega)$ is a random variable. For a fixed $\omega' \in \Omega$, $Z(x, \omega')$ is a deterministic function of x, called a realization of the random function.

Our first goal is to construct a trend, and by *trend* we understand a slowly varying component of a data set or random function that prevents us from using the hypothesis of stationarity. The trend is an approximation to the *drift*, denoted $m(x)$, which is defined as the expected value of a random function:

$$m(x) = E(Z) = \int_{\Omega} Z(x, \omega)\, dP(\omega),$$

Fig. 2. Landscape data from Toten illustrated as shaded relief maps.

where $P(\omega)$ is the distribution function of ω.

Let $f(x) = Z(x, \omega')$ denote the exact terrain surface we are trying to model. We can write:

$$f(x) = m(x) + z(x),$$

where m is the drift we defined earlier, and z a realization of a zero mean random function. The values we observe can be seen as samples of another surface

$$\tilde{f}(x) = m(x) + z(x) + w(x),$$

where w is assumed to be a zero mean white noise component due to measurement errors. Of course, we only know $\tilde{f}$ at the observed points, and the question is how well we can estimate the drift from a discrete set of noisy observations. The procedure we will apply uses spatial averaging over subsets of D as proposed by [BN05].

We start the trend computation by dividing the geographical area into grid blocks, making sure a suitable number of measurements fall within each block; numerical test show that an average of ca. 10 values per block gives good results. Based on these values, we estimate the average depth a_i inside each block, and we use these values to approximate the drift by generating the sum

$$m(x) = \sum_{j=1}^{n} a_j \, \phi_j(x), \tag{1}$$

where n is the number of blocks. Each function ϕ_j is continuous and has integral 1 over block j and integral zero over all other blocks. The method we use to construct these functions is based on average interpolating subdivision; for an introduction see Sweldens and Schröder, [SS96]. Note that we are not interpolating the observations – the surface we construct has the same average values over the grid blocks as the original estimated values, and does not in general pass through any of the measured data points.

When we have a model of the drift, we can compute the residuals between the drift and the observations as

$$r(x) = \tilde{f}(x) - m(x) = z(x) + w(x). \tag{2}$$

The two components $z(x)$ and $w(x)$ cannot be separated from these equations, but from the semivariogram we can derive $w(x)$. In the semivariogram $w(x)$ appears as the so-called nugget effect (see e.g. [Ole99] or [Cre93]) The semivariogram is defined as

$$2\gamma(h) = E[\{r(x) - r(x+h)\}^2],$$

where x and $x + h$ are any two locations in the sample domain.

3 The experiment

The data we use are collected with the EM3000 dual head multibeam echo sounder and airborne laser scanner equipment. The test areas are shown in Figures 1 and 2. Buildings, trees and forest areas in the landscape models are filtered out. This kind of filtering is not a trivial task. The methodology used is based on filtering out the measurements which deviate more than a certain threshold from the trend surface generated by the applied subdivision algorithm.

The two transducers of the EM3000 are installed so that they have a small overlap area. The overlap area is indicated in the relief map in Figure 1. The multibeam data are from the Oslo fjord and the laser data from the Toten region in the South-East part of Norway.

Table 1. Statistics for the test areas. The standard deviation of the random error, termed σ, is computed from the semivariogram of the residuals between the drift and the observed depth/height values. σ is computed for the whole area.

Area	No. of datapoints	Size of area m	No. of points per m^2	Size of block m	Noise σ m	sea depth m
Seafloor						
A	50044	160.0	2.0	5.0	±0.059	27.4
Landscape						
B	125981	427.8	0.7	6.7	±0.062	
C	199902	427.8	1.1	6.7	±0.067	

Table 1 gives statistics for the different survey strips. The seafloor model consists of one test area of size 160x160m. The landscape model is composed of two areas termed Toten B and C. These areas are each of size 430x430m, approximately. The number of data points inside each of the test areas varies from fifty- to two hundred thousand points with a point density of 2.0 points per m^2 for the seafloor model and 0.7 to 1.1 points per m^2 for the land scape models. The grid of rectangular blocks we have used for the computation of average values, varies in size from 5x5m to 7x7m, approximately.

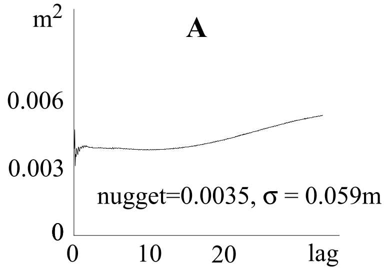

Fig. 3. Analysis of the seafloor data; semivariogram of the residuals.

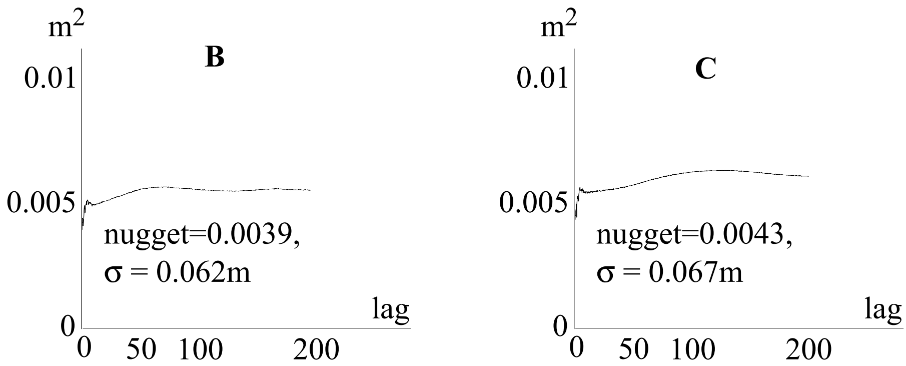

Fig. 4. Analysis of the land scape data from Toten; semivariogram of the residuals.

Figures 3 and 4 show the semivariogram of the seafloor and the landscape models, respectively. Since the variograms soon reach the sill, the drift is well eliminated from the residuals. The nugget effect $\gamma(0)$ is derived from the variograms, and σ computed as $\sigma = \sqrt{\gamma(0)}$. Figure 5 shows how the standard deviation of the residuals $r(x)$ varies along the survey strip.

4 Discussion

Maas [Maa02] gives an estimate of about 40mm for the precision of the individual laserscanner point height. The average precision we have derived for our land-

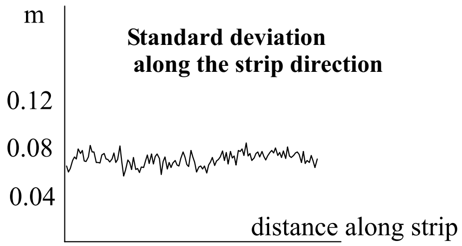

Fig. 5. Analysis of the seafloor data; σ of the residuals $r(x)$ along the strip.

scape models is 65mm. Since Maas [Maa02] computes the noise component from least-squares matching where the height shift between the overlapping survey strips is introduced, the effect of systematic errors to high degree is eliminated in his investigation. In our case the landscape model is composed of several overlapping laserscanner strips, but we have not evaluated the accuracy of the calibration. Calibration errors, therefore, can have an effect on the noise component we have derived. Maas [Maa02] applies interpolation in a triangular mesh when comparing the overlapping strips. This kind of interpolation has a noise reduction effect, which leads to an optimistic estimate of the random error. The previous moments considered, we can expect that our average precision value for the laserscanner models is more pessimistic than the average value given by Maas [Maa02].

We observe that the noise component in the seafloor model is 59mm, i.e., at the same level as for the laser scanner data. Although multibeam echo sounders and airborne laserscanning are different measurement techniques, our investigation indicates that data collected with the two techniques can be merged into a terrain model with homogenous accuracy. This is important for mapping in coastal zones.

Figure 5 shows how the standard deviation σ_r of the residuals $r(x)$ varies along the survey strip. We cannot observe any systematic change of σ_r. Since the semivariogram of the residuals $r(x)$ shows how well the drift computation works, we can regard the variation of $z(x)$ as minor to the variation of $w(x)$. Therefore, we make the assumption that there is no systematic change of the random error component in the strip direction.

5 Conclusions

We have demonstrated a method for the derivation of the random measurement errors in digital terrain models. The experiments on one seafloor model and two landscape models demonstrate how well the method works.

The procedure for computing the random error of the observed depth/height values is as follows: (1) partition the area into rectangular blocks of appropriate size; (2) compute average values for the blocks; (3) compute the drift $m(x)$ by applying average interpolating subdivision, see [BN05]; (4) compute the residuals from $r(x) = \tilde{f}(x) - m(x)$; (5) compute the experimental semivariogram of the residuals; (6) fit a variogram model to the experimental semivariogram and derive the nugget effect, i.e., compute the noise component of the observations as $\sigma_w = \sqrt{\gamma(0)}$.

An attractive property of the presented method is that it requires only one set of data. Therefore, the terrain model can be analyzed in real time, i.e., the noise component can be continuously monitored during the data collection. This is very attractive in seafloor mapping.

Acknowledgments

This research is supported by the FFI project SWASI. We are very grateful to the company Blom Geomatics A/S that provided us with the laser data set. The EM3000 data set is collected by FFI. Our gratitude goes also to Leif Erik Blankenberg, Sveinung Himle, and Håkon Andersen for preparing the laser data set and Kjersti Hovemoen, Henrik Halvorsen, Bård Henriksen, and Arnfinn Karlsen for collecting the EM3000 data. Their support is highly appreciated.

References

[BN05] Jan T. Bjørke and S. Nilsen. Trend extraction using average interpolating subdivision. *To appear in Mathematical Geology*, 2005.

[Cre93] Noel A. C. Cressie. *Statistics for Spatial Data, Revised Ed.* Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1993. 900 pp.

[Maa02] Hans-Gerd Maas. Methods for measuring height and planimetry discrepances in airborne laserscanner data. *Photogrammetric Engineering* & *Remote Sensing*, 68(9):933–940, 2002.

[Ole99] R. A. Olea. *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers, Boston, 1999. 303 pp.

[SS96] Wim Sweldens and Peter Schröder. Building your own wavelets at home. In *Wavelets in Computer Graphics*, pages 15–87. ACM SIGGRAPH Course Notes, 1996.