

# Standard deviation as the optimization criterion in the OptD method and its influence on the generated DTM

Wioleta Błaszczak-Bąk<sup>1</sup>, Anna Sobieraj-Żłobińska<sup>2,\*</sup>

<sup>1</sup>University of Warmia and Mazury in Olsztyn, Faculty of Geodesy, Geospatial and Civil Engineering, Institute of Geodesy, Oczapowski St. 1, Olsztyn, Poland

<sup>2</sup>Gdańsk University of Technology, Faculty of Civil and Environmental Engineering, Department of Geodesy, Narutowicza St. 11/12, Gdańsk, Poland

**Abstract.** Reduction of the measurement dataset is one of the current issues related to constantly developing technologies that provide large datasets, e.g. laser scanning. It could seem that presence and evolution of processors computer, increase of hard drive capacity etc. is the solution for development of such large datasets. And in fact it is, however, the “lighter” datasets are easier to work with. Additionally, reduced datasets can be exchange/transfer/download faster via internet or cloud stored. Therefore the issue of data reduction algorithms/methods is continuously relevant. In this paper authors presented the results of the study whether the standard deviation of measurement data can be used as optimization criterion in the process of dataset reduction conducted by means of the OptD method. The OptD is based on the cartographic generalization methods. In iterative process irrelevant points are being removed and those that characteristic are being preserved, what in results means more points in complex fragments of scanned object/surface and less in flat/uncomplicated area. Obtained reduced datasets were then the basis for DTMs generation. For DTMs assessment RMSE was calculated.

## 1 Introduction

DTM (Digital Terrain Model) is one of the most popular model representing the surrounding world. It is used in many fields of science and economy to study the state of existing or to simulate future events [1, 2]. It can be developed, inter alia, on the basis of LiDAR (Light Detection and Ranging), which in a relatively short time provides a large amount of measurement data. However, the measurement data itself is insufficient. Additionally, measurements should include a measure of their reliability or uncertainty of measurement [3]. The uncertainty of measurement data is discussed in the Guide to the expression of uncertainty in measurement developed by the International Organization for Standardization (ISO) in consultation with a number of global scientific and technical organizations [4, 5]. Uncertainty of measurement is expressed e.g.: by measurement error, which represents the difference between the measured value  $x$  and the real  $x_0$ . The measurement error defined in this way is not a measure of the accuracy of the measurement method, because similar measurements made with another instrument at another time and place will give another values [6]. Thus,  $x_i$  is a random value and its scattering is characterized by a parameter called the standard deviation estimator SD.

$$SD = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \quad (1)$$

where:  $x_i$  - observed values,  $\bar{x}$  - mean value of these observations,  $n$  - number of observations.

SD can be identified with uncertainty of measurement if any of the values  $x_i$  is taken as result of measurement.

An important aspect in the development of LiDAR data for DTM generation is an evaluation of the accuracy of the obtained product [7-9]. The accuracy of the DTM is a function of a number of variables such as roughness of the terrain surface, interpolation function, interpolation methods, and the three attributes (accuracy, density and distribution) of the source data [10, 11]. Therefore, the measurement uncertainty parameter, i.e. SD of source data has an impact of the generated DTM.

In this paper focus is on the influence of the source data, in particular, whether the standard deviation estimator of ALS (Airborne Laser Scanning) data can be used as optimization criterion in dataset reduction and whether using SD has an impact on the generated DTM. Reduction was performed by means of the Optimum Dataset (OptD) method [12], which allows to preserve points representing characteristics elements in reduced dataset [13, 14, 15]. DTM was generated on the basis of the original dataset (after its filtration) as well as from the datasets obtained after processing by the OptD method.

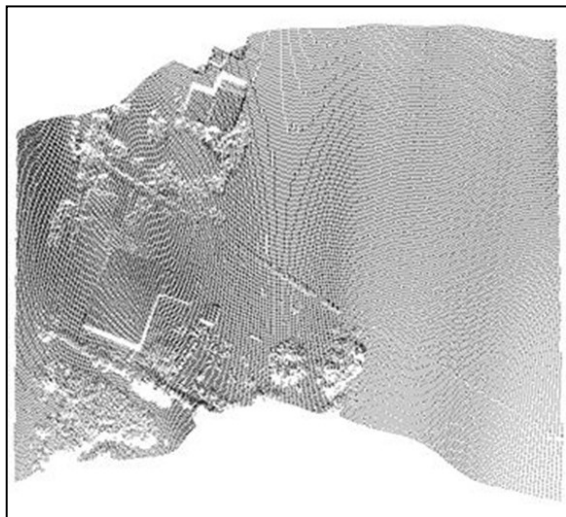
\* Corresponding author: [annsobie@pg.edu.pl](mailto:annsobie@pg.edu.pl)

The study area is a fragment of the national road No. 16, Sielska street in Olsztyn, located in the Warmia-Mazury. The measurement was conducted by Airborne Laser Scanning. ALS was made by Visimind Ltd. For this study, part of this measurement was selected. Laser scanning angle was 60 degrees, with a frequency of 10,000 Hz scanning. Scanning was performed from a helicopter with speed of 50 km/h at an altitude of 70 m.

Within analyzed area control points were set up. They were measured by means of Leica Viva NetRover receiver. Observation time in each measurement point was 12 seconds, weather conditions were satisfactory. Corrections VRS from ASG EUPOS were applied. The result of the measurement was the dataset consisting of 204 points.

## 2 Data reduction and DTM generation

The dataset obtained from ALS used in this study contains 74785 points. There are buildings, street, trees within this area. The ALS dataset is presented in Fig. 1.



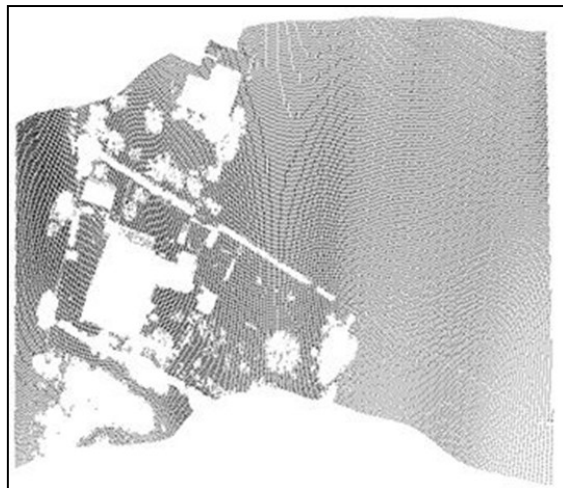
**Fig. 1.** Top view of LiDAR point cloud (source: CloudCompare v.2.6.0).

The selected fragments were filtered by using 'adaptive TIN model' method [16] in proprietary software. As a result of the filtration, there are two datasets: the dataset of points showing the topography (topographic surface dataset - TSset) and a dataset of points showing the details points. The topographic surface dataset for TSset consists of 37481 points. It is presented in Fig. 2.

The datasets after filtration contain only points representing the terrain. Next, they were processed by means of the OptD method in order to reduce them. The OptD puts an emphasis on fact, that the reduction is conducted in a way, that the information necessary for the proper performance of a task is not lost. In this case, points representing characteristics of terrain are preserved due to further DTMs generation. To achieved that, the dataset is divided into measurements strips and within each strip linear object generalization method is used. In this paper Douglas - Peucker (D-P) method [17] was used in OptD-

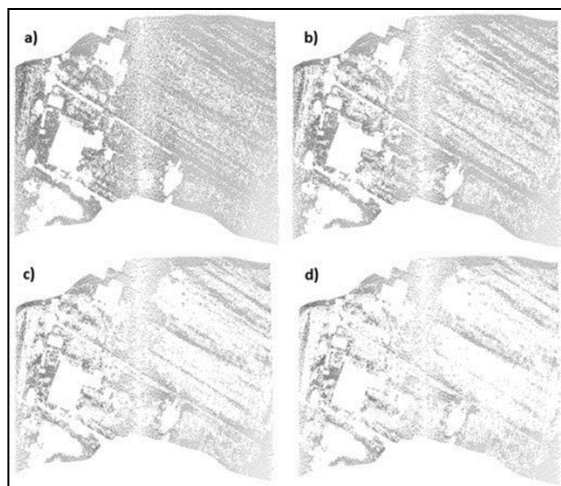
single. The D-P method is based on tolerance distance. The calculations are done in a vertical plane, what allows to accurately check each elevation. The degree of reduction depends on the value of mentioned parameters. They are automatically changed during processing to meet the criterion or criteria set by user. In results, the optimal datasets are obtained.

As optimization criteria in the OptD method parameters like: the number of points in reduced dataset (M) and the percentage of points to be in the dataset after processing (p%) were used and tested so far [13-17]. In this paper it was decided to use the standard deviation estimator (SD) of ALS data.



**Fig. 2.** LiDAR point cloud after filtration - TSset (source: CloudCompare v.2.6.0).

The first calculation was made with assumption, that the difference in SD calculated for original and reduced dataset cannot increase more than  $SD_{TSset} - SD_{OptDset} < 0.200$  m. Then the difference of SDs was increased by 0.200 m. In this way, four datasets were determined. The datasets after processing by means of the OptD-single method were named OptDset1, OptDset2, OptDset3, OptDset4 respectively.



**Fig. 3.** ALS point cloud after processing by the OptD-single method: a) OptDset1, b) OptDset2, c) OptDset3, d) OptDset4 (source: CloudCompare v.2.6.0).

In order to compare the results of using OptD-single method, the parameters that describe the TSset were determined. Statistics of TSset are presented in Table 1.

**Table 1.** Statistics of TSset.

Parameter	value [m]
Number of points (N)	37481
Max height [m]	366.361
Min height [m]	348.610
Mean height [m]	358.912
SD[m]	3.760
Point distance	0.340

The OptD-single method selected the optimum solutions for all OptDsets. Statistics of OptDsets, as well as the difference of standard deviation estimator for TSset and for OptDsets are presented in Table 2. For SD OptDset calculation mean height is taken from TSset.

SD calculated for TSset and OptDsets concerns measured height:

$$SD = \sqrt{\frac{\sum(z_i - \bar{z})^2}{n-1}} \quad (2)$$

where:  $z_i$  ( $i = 1, 2, \dots, n$ ) – the height of subsequent points in the dataset,  $\bar{z}$  – mean height in TSset,  $n$  – number of points in dataset (original – N or reduced - M).

The value of SD increases with the decrease in the size of the dataset. In the range from 25% to 70% of reduced points in the TSset, the value of the difference between SDs of TSset and OptDsets ranges from 0.160 m to 0.752 m, these are insignificant values for data representing the surface area.

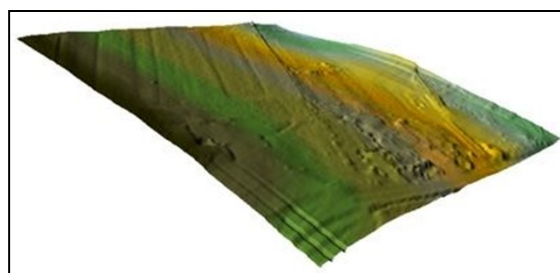
On the basis of the reduced datasets, DTMs were generated using the kriging method in the Surfer with various grid sizes (0.5m and 1m). In Fig. 4 and Fig. 5 the DTM generated from the TSset (DTM TSset) is shown. Application of the OptD method gave a few solutions, which were used for DTM generation. In Fig. 6 and Fig. 7 four DTMs based on these OptDsets are presented.

To test, how the reduction of measurement data influence the DTMs generation, some cross-sections in random localization were performed. It was made for DTMs with grid 0.5m and 1m. those differences are from range 0.035m -0.149m. For others cross-section results are similar. Thus, it can be assumed, that the reduction did not distorted the generated surface.

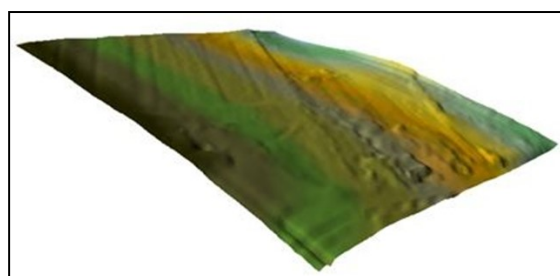
In order to determine the accuracy of generated DTM OptDsets within tested area 204 control points were measured by means of GNSS technique. Those points were used to calculate RMSE of generated models.

**Table 2.** Statistics of OptDsets.

Parameter	OptDset 1	OptDset 2	OptDset 3	OptDset 4
Mean height in OptDset [m]	358.799	358.652	358.480	358.458
Mean point distance in OptDset [m]	0.397	0.491	0.632	0.744
Number of terrain points in OptDset (M)	41488	28955	19610	15136
SD OptDset [m]	3.920	4.121	4.313	4.512
The absolute value of differences of SD TSset and SD OptDset [m]	0.160	0.361	0.553	0.752



**Fig. 4.** DTM TSset, grid size=0.5m (source: own study in Surfer v.9).



**Fig. 5.** DTM TSset, grid size=1m (source: own study in Surfer v.9).

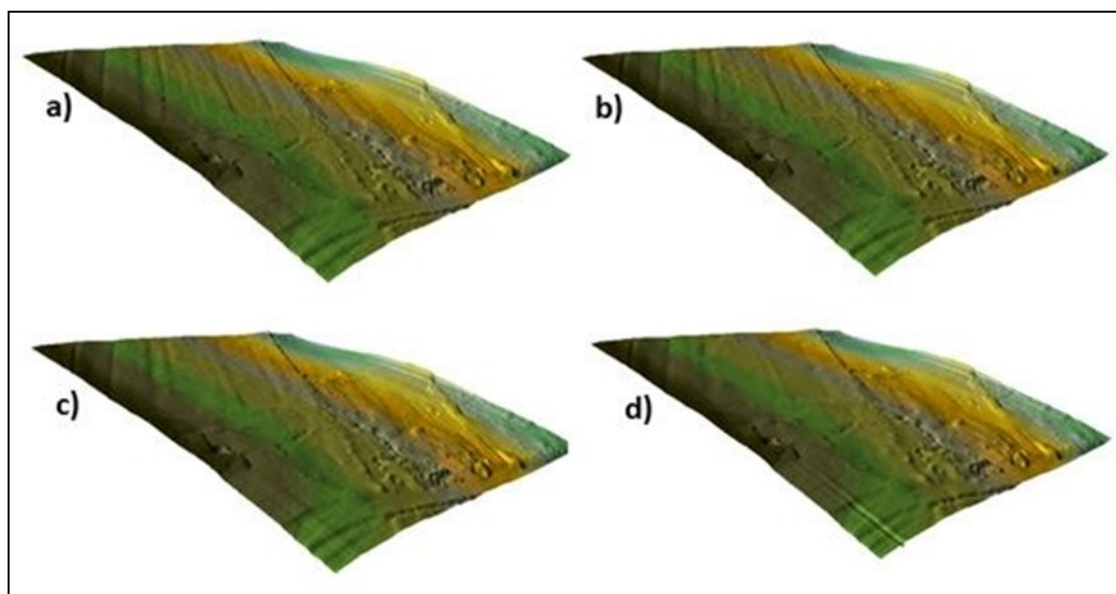


Fig. 6. DTMs based on OptDsets, grid size = 0.5m: a) DTM1, b) DTM2, c) DTM3, d) DTM4 (source: own study in Surfer v.9).

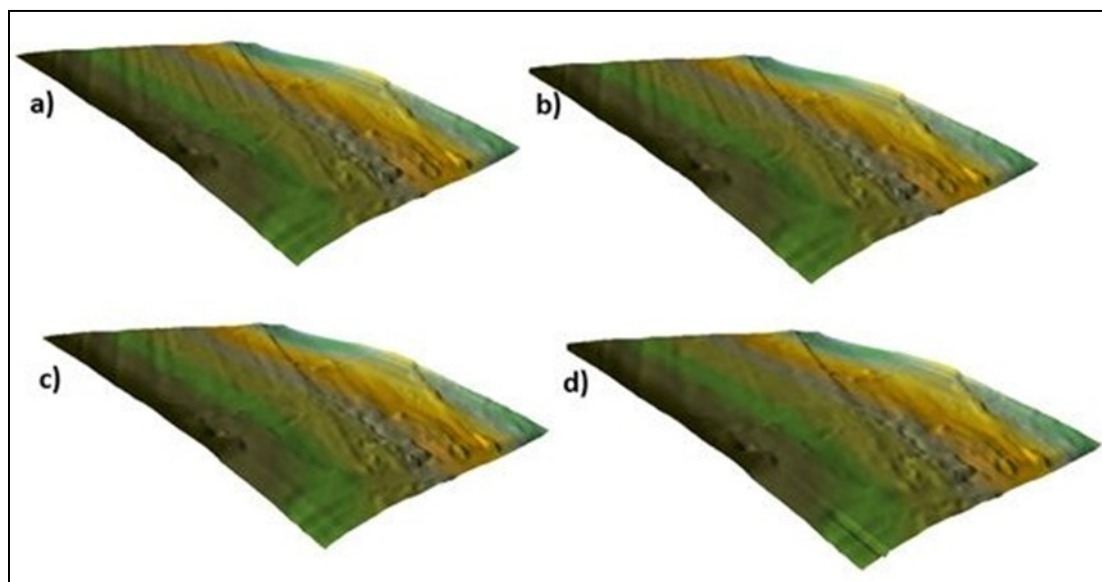


Fig. 7. DTMs based on OptDsets, grid size = 1m: a) DTM1, b) DTM2, c) DTM3, d) DTM4 (source: own study in Surfer v.9).

Table 3: Characteristics of DTM.

grid size 0.5m					
	DTM TSset	DTM1	DTM2	DTM3	DTM4
Zmean [m]	357.988	357.989	357.989	357.990	357.994
RMSE	0.430	0.520	0.605	0.743	0.896
grid size 1m					
	DTM TSset	DTM1	DTM2	DTM3	DTM4
Zmean [m]	357.987	357.988	357.988	357.989	357.992
RMSE	0.510	0.583	0.612	0.782	0.980

**Table 4:** Differences of RMSE

The absolute value of differences [m]	DTM TSset – DTMi (i=1, 2, 3, 4)			
	DTM1	DTM2	DTM3	DTM4
ARMSE for DTMs grid size 0.5m	0.090	0.175	0.313	0.466
ARMSE for DTMs grid size 1m	0.073	0.102	0.272	0.470

For the generated DTM mean height and RMSE were calculated.

The results show that the values of SD increase along with the decrease in the number of points in the ALS dataset.

Presented results show that the more points have been removed in a dataset by the OptD method, then the bigger are the differences between RMSE of DTMs generated on the basis of this dataset. The range of changes is from 0.073 to 0.470 m. Therefore, using DTMs for various purposes is determined by these values.

### 3 Conclusions

This paper presents the use of the standard deviation parameter as the optimization criterion in OptD-single method. It was assumed that the difference in SD calculated for the original and reduced dataset cannot increase more than 0.200 m. Tests were carried out for 4 cases, each time increasing the difference in SD by 0.200 m. Next, on the basis of obtained (reduced) datasets, DTMs were generated. For the DTM assessment parameter like RMSE was used. To calculate the RMSE the control points were measured by GNSS.

The analyzes show that the SD of the dataset can be an optimization criterion in the OptD-single method. The obtained datasets met the optimization criterion, and the DTM generated on their basis also met the accuracy expectations. At control points, the calculated RMSE for DTM with 0.5 m grid increased by a maximum of 0.466 m, while for DTMs with 1m grid size by 0.470 m. Differences in RMSE indicated, that application of DTMs generated on the basis of reduced datasets, is determined by degree of reduction.

### References

1. K. V. Prasad, *Digital Terrain Modeling: Principles and Methodology*. The Photogrammetric Record, 24: 296–297 (2009) DOI: 10.1111/j.1477-9730.2009.00545\_2.x.
2. C. Suchocki, J. Katzer, J. Rapiński, J., *Terrestrial Laser Scanner as a Tool for Assessment of Saturation and Moisture Movement in Building Materials*. Period Polytech Civ Eng, 1–6. (2018) DOI:10.3311/PPci.11406
3. A. Zięba, P. Ramza, *Standard deviation of the mean of autocorrelated observations estimated with the use of the autocorrelation function estimated from the data*, *Metrol. & Meas. Syst.*, **18**, pp. 529-542 (2011)
4. ISO/IEC Guide 98-3:2008, *Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)*

5. *The Expression of Uncertainty and Confidence in Measurement*, M3003 Edition 3 (2012)
6. S. Bell, *A Beginner's Guide to Uncertainty of Measurement* (1999)
7. J. Estornell, L. A. Ruiz, B. Velázquez-Martí, T. Hermosilla, *Analysis of the factors affecting LiDAR DTM accuracy in a steep shrub area*, *International Journal of Digital Earth*, 4:6, 521-538 (2010) DOI: 10.1080/17538947.2010.533201
8. K. Stereńczak, M. Zasada, M. Brach, *The Accuracy Assessment of DTM Generated from LIDAR Data for Forest Area - a Case Study for Scots Pine Stands in Poland*. *Baltic Forestry*, **19**, 2, pp. 252-262 (2013)
9. Z. Li, Q. Zhu, C. M. Gold, *Digital terrain modeling - principles and methodology*. CRC Press, ISBN 978-0-415-32462-5, pp. I-XVI, 1-323 (2005)
10. Z. L. Li, *Sampling Strategy and Accuracy Assessment for Digital Terrain Modelling*. Ph.D. thesis, The University of Glasgow (1990)
11. Z. L. Li, *Variation of the accuracy of digital terrain models with sampling interval*. *Photogrammetric Record*, **14**, 79, 113–128 pp. (1992)
12. W. Błaszczak-Bąk, *New Optimum Dataset method in LiDAR processing*. *Acta Geodynamica et Geomaterialia*, **13**, 4(184), pp. 381-388 (2016) DOI: 10.13168/AGG.2016.0020
13. W. Błaszczak-Bąk, A. Sobieraj-Żłobińska, B. Wiczorek, *The Optimum Dataset method – examples of applications*, *ES5 Web of Conferences*, **26**, 00005 (2018)
14. W. Błaszczak-Bąk, M. Poniewiera, A. Sobieraj-Żłobińska, M. Kowalik, *Reduction of measurement data before Digital Terrain Model generation vs. DTM generalisation*, *Survey Review* (2018) DOI: 10.1080/00396265.2018.1474685
15. W. Błaszczak-Bąk, A. Sobieraj-Żłobińska, M. Kowalik, *The OptD-multi method in LiDAR processing*. *Measurement Science and Technology*. **8**, 7, 075009 (2017) DOI:10.1088/1361-6501/aa7444
16. P. Axelsson, *DEM generation from laser scanner data using adaptive TIN models*, *International Archives of Photogrammetry and Remote Sensing*, Volume XXXIII, Part B4, Amsterdam, Netherlands:111–118 (2000)
17. D. H. Douglas, T. K. Peucker, *Algorithms for the reduction of the number of points required to represent a digitized line or its caricature*. *Canadian Cartographer*, **10**, 2, pp. 112-122 (1973)