NEW OPTIMUM DATASET METHOD IN LiDAR PROCESSING

Wioleta BŁASZCZAK-BĄK

Institute of Geodesy, Faculty of Geodesy, Geospatial and Civil Engineering, University of Warmia and Mazury in Olsztyn, Oczapowski 1 Street, 10-719 Olsztyn, Poland

*Corresponding author's e-mail: wioleta.blaszczak@uwm.edu.pl

ARTICLE INFO

ABSTRACT

Until now, the optimization of a large dataset acquired by means of the laser scanning technology was understood as reducing the number of data and finding a satisfactory solution. Generating Digital Terrain Model on the basis of the reduced dataset does not always lead to desired results or previously planned goals. Therefore, it is important that the algorithm which reduces large dataset, could find the optimal solution for creating the model. The objective of this paper is to develop and test a new OptD method (Optimum Dataset) in the processing of Airborne Laser Scanning point cloud. The algorithm of this method can reduce the dataset in terms of number of measuring points for a given criterion, such as e.g. mean error of the Digital Terrain Model.

1. INTRODUCTION

Until now, the concept of optimization, understood as determination of the best/optimal solution for a particular criterion, functioned in many fields of science, technology, economy, etc. In surveying the optimization can be understood as:

(a) Reduction of the number of points in the LiDAR dataset (Light Detection and Ranging) in particular in ALS dataset (Airborne Laser Scanning), presented by e.g.: Błaszczak (2006), Błaszczak et al. (2011a, 2011b). Reduction results in decreasing the number of the dataset, the remaining points are actual measuring points.

(b) Generation of datasets as GRID, presented by e.g.: Bauer-Marschallinger et al. (2014). Generation also results in reduction of the number of the dataset, in reduced set there are new points (interpolated) instead of points with the original coordinates.

(c) Effective planning of a point cloud processing by the choice of the appropriate filtration methods presented by e.g: Sithole and Vosselman (2005), Tóvári and Pfeifer (2005), Hebel and Stilla (2008). In these papers there is information about applied point cloud filtration methods, for example: morphological filters, linear prediction, iterative algorithms fitting modeled surface into point cloud, adaptive TIN modelling. Effective LiDAR point clouds processing are also presented by Reitberger et al. (2009), Vosselman (2008), Saeedi et al. (2009), Vosselman and Maas (2010). These papers are proof that method which would properly work for all types of terrain has been not developed so far. Therefore, work is still continuing on improving the methodology for ALS point cloud processing.

However, it should be noted, that the concept of optimization is much wider.

The optimization process should find optimum understood as the best solution according to the criteria preset by the user. It is particularly important during the development of point clouds for DTM (Digital Terrain Model) generation. It is important that the set of data points representing the terrain meets certain criteria that will lead to the planned DTM's accuracy. In the existing literature dataset for DTM generation was chosen on the basis of a number of tests and repeated calculations, until a satisfactory solution was achieved. User decided on the selection criteria, but has had no direct impact on the result of the calculation. Complex calculation was repeatedly performed with changed values of criteria until a satisfactory result (not necessarily optimal) was reached.

Therefore, there is a need to have a better, than developed so far, solutions or methods for decreasing the number of the LiDAR point clouds. Such methods should be automated and applying them should enables to achieve an optimal set for specific goal or task. It was a reason for developing a new OptD method, presented in this paper. The use of the OptD method allows select the optimal dataset, which will be reduced in terms of number of measuring points for a given criterion. Mentioned criterion is also related to the issues of the quality of the DTM, generated on the basis of the optimal solution.
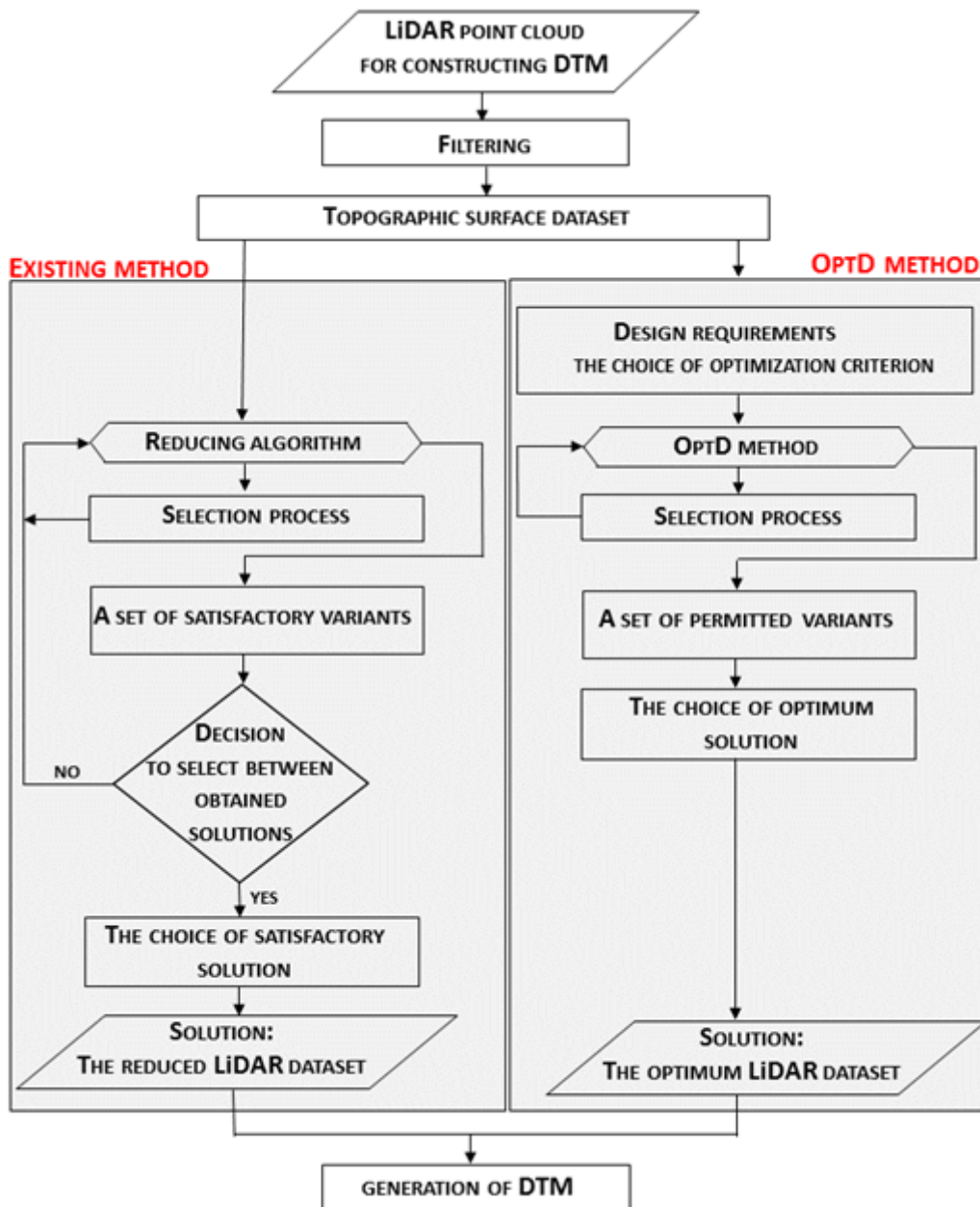
Fig. 1   Options of LiDAR point cloud development (source: own study).

2.   THE OPTD METHOD IN LiDAR PROCESSING

The development of ALS point clouds for the construction of DTM is performed in stages. The first step (pre-processing) comprises a gross error elimination, as well as filtration and optimization. The second stage (initial processing) is the generation of DTM.

Earlier studies in reducing the number of LiDAR point clouds presented in Błaszczak et al. (2011a, 2011b) indicate that the reduction of the LiDAR dataset can be made before or after filtration. Although total time of point cloud development for DTM generation in variant reduction - filtration is shorter than in the variant filtration - reduction, it is important to perform filtration first in order to apply the reduction criterion which takes into account the

average error of DTM. After filtration a set of points representing the terrain will be received and on such basis DTM will be generated.

The reduction of the number of measurement points in the LiDAR point cloud can be carried out in two options:
(a) satisfaction option (existing approach),
(b) optimum option (a new approach).

In the first option the existing algorithm was applied. It was proposed in earlier papers by Błaszczak (2006), Błaszczak-Bak (2012), Błaszczak et al. (2015). In the second option **Opt**imum **D**ataset (**OptD**) method was used.

Options of LiDAR point cloud development are presented in Figure 1.

In previous studies we find a satisfactory option. LiDAR point cloud is filtered, and in result we obtain a set of points presenting the terrain. Then, the chosen fragment of point cloud is processed by means of reduction algorithm. During reduction a set of points that meets specific criteria (search strip width, the tolerance of chosen generalization method) is selected.

The selection process is conducted several times, until satisfactory variants are achieved. Satisfactory variants are the sets of points with such number of measurement points, which enables to build the DTM. The user decides which set of data will be used in the further development. If none of the datasets, chosen during selection, do not meet user's expectations, then original set of points is processed by means of a reduction algorithm again. If decision on a satirfactory solution has been made, then processing has been stopped, and on that basis of chosen dataset, DTM is generated.

In the new approach, there is also pre-filtration, which leads to selection a set of points representing the terrain. However, the process of selecting a dataset, on the basis of which the DTM will be build, starts with setting out the optimization criterion e.g: mean error, minimum number of points. Depending on the number of adopted criteria, optimizations can be divided into single optimization and polioptimalization (multi objective) (Cempel, 2000).

Single optimization is an optimization, in which only one criterion is required to achieve for assessing condition as ideal state. On the other hand, polioptimalization makes possible to reach the desired solution after meeting a number of criteria for assessment. Among the methods of multi-criteria optimization techniques we can distinguish Pareto approaches (Marcinkowski, 2008).

The criterion for optimization is the basic concept of optimizing, by means of which a comparison of the respective solutions is being made. The criterion expressed in the language of mathematics is called the objective function.

The criterion for optimization is always selected in the initial phase of processing. The criterion may be selected among quality parameters of DTM, it can be a combination of many parameters. The decision problem in this case is the mathematical formula of DTM quality expressed by parameters.

The selection of criteria is then followed by processing using the optimization algorithm of OptD method until a set of optimal number (in the sense of adopted criteria) of LiDAR points is obtained.

OptD method can be conducted in two variants:

(a) OptD method with single objective optimization called **OptD-single**,

(b) OptD method with multi objective optimization called **OptD-multi**.

If OptD-single method is chosen, then a set which strictly is fulfilling one condition is sought. If there is a decision on processing using OptD-multi, then in result several sets will be obtained, among which the best one should be selected. It can be assumed, that this is the set of decision D and a set of measurable criterion K. In the set D the "best" decision can be determined, in terms of the considered set of criterion.

The optimum solution is chosen from the obtained set of feasible solutions. The successful dataset is the basis for generating DTM with established design requirements.

The proposed OptD method in new optimal variant can be used not only on the data obtained from ALS, but also for TLS (Terrain Laser Scanning) and other large datasets, e.g. from sonar measurements. In this study, it was decided to use the OptD-single method. Multi-criteria optimization will be tested in author's future works.

3. OptD-SINGLE METHOD

The OptD-single method enables the reduction of big dataset by means of one optimization criterion

The algorithm of OptD method consists of the following steps:

step 1: Loading the N points of the original LiDAR dataset.

step 2: Establishing optimization criterion (f), e.g. number of points in the set, the mean error of DTM.

step 3: Determination of the XYZ coordinate system. The aim of this step is to 'fitting' the measurement lines, that are a direct result of the measurement, into coordinate system so that the measurement lines are approximately parallel to the X or Y axis (the coordinates of points have the certain regularity and are arranged in a measurement lines).

step 4: Projection of LiDAR data points onto a plane X0Y.

step 5: The choice of initial width of belts (L). Choosing the appropriate belts width some parameters (depends on the user) can be taken into consideration: the average distance between points in the measurement set, as well as the distance between the belts, which arose directly from the type of measurement (here: LiDAR) and they are a consequence of the principle of LiDAR measurement. Another way of belt's width determination is in an iterative process and change e.g. at a fixed interval.

step 6: The division of area covered by points on the test belts (nL).

step 7: Selection of measurement points for each measurement belt.

step 8: Projection on the plane Y0Z LiDAR data points for each measurement belt.

step 9: Selecting the method of cartographic line generalization, e.g. the Douglas-Peucker

(D-P) (Douglas and Peucker, 1973), Visvalingam-Whyatt (Visvalingam and Whyatt, 1992).

step 10: Using the selected method of generalization in the plane Y0Z. Choosing the tolerance parameters in the selected method of generalization. For the method of D-P it is a distance of tolerance. The initial value of the section is defined by the user, the following values are determined in an iterative process, in which there is increase or decrease at a fixed interval.

step 11: Obtaining the reduced data set with the number of M, where M<N.

step 12: Verification, whether obtained in Step 10 set fits the specified criterion optimization. If so, the reduction process is completed, and the obtained set from Step 10 is the optimal dataset. If not, the steps 9-12 are repeated, wherein in step 8 the value of tolerance parameter is changed. If repeating steps 9-12 do not give a solution, there is need to back to Step 6 and change the width of measuring belt.

The algorithm of OptD-single method in the form of flow chart is presented in Figure 2.

Presented algorithm of OptD method was tested on real ALS data.

4.   TESTS FOR OptD METHOD

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

ALS measurement enabled the acquisition of point clouds. A fragment of the original dataset ALS (144500 points), which was used as a study area of this research, is presented in Figures 3 and 4.

The selected fragment was filtered by using 'adaptive TIN model' method (Axelsson, 2000) in own software. As a result of the filtration, there are two sets of data: a) the set of points showing the topography (topographic surface dataset - **TSset**) (108 313 points), b) a set of points showing the details points (36187 points).

Point cloud after filtration, which will be used to generate the DTM is shown in Figure 5.
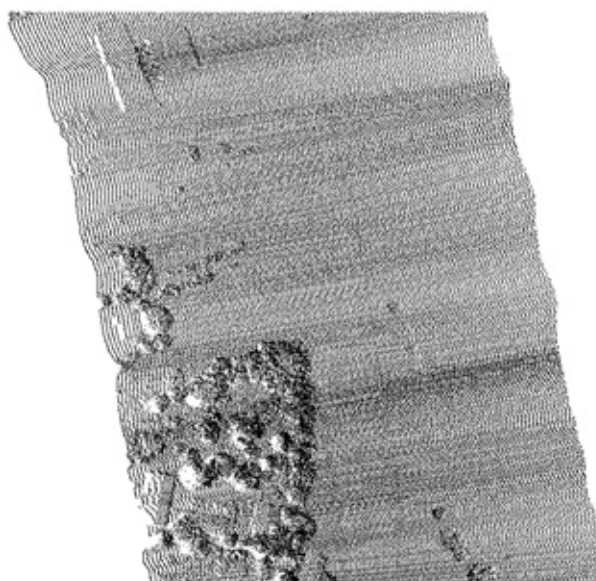
Point cloud after filtration comprising only ground points was optimized by OptD-single method.

4.1. *PROCESSING BY ALGORITHM OF OptD-SINGLE METHOD*

TSset with the number N = 108 313 points has been processed by OptD methods, that begins with the determination of optimization criterion. In this work,
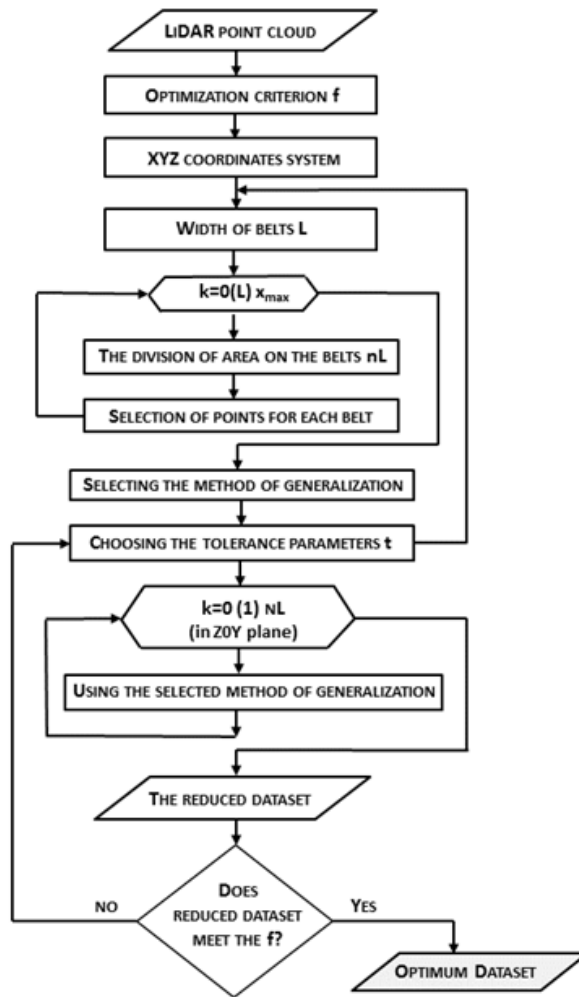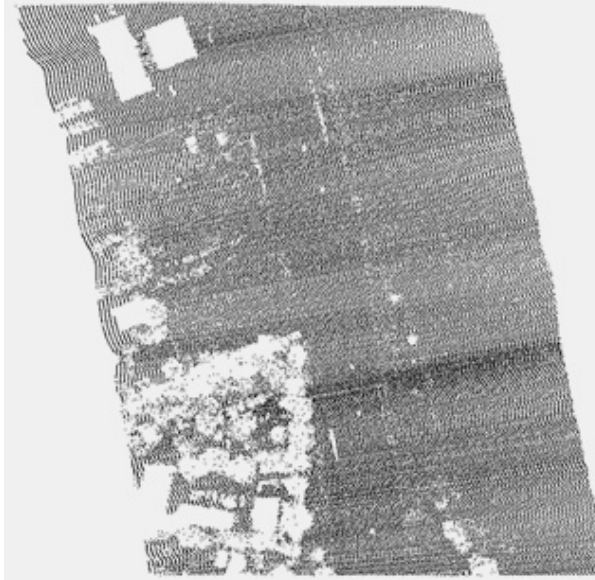

Fig. 2   The algorithm of   OptD-single method (source: own study).

Fig. 3   Top view of the original fragment of the ALS point clouds top view (source: own in the CloudCompare v. 2.6.0).

Fig. 4 Side view of the original fragment of the ALS point clouds (source: own in the CloudCompare v. 2.6.0).

Fig. 5 Point cloud after filtration (source: own in the own software).

in practical tests it was decided to perform an OptD-single optimization method. Multi-criteria optimization will be tested in author's future works.

In DTM generation there are very important parameters for assessing the quality of the model as: the mean error, range, mean value of height difference, the root-mean-square error, coefficient of determination.

In this work as overriding optimization criterion DTM mean error was adopted:

$$m_0 = \sqrt{\frac{(z_{mean} - z_i)^2}{M - 1}} \qquad (3)$$

where: z_{mean} is a mean height calculated from heights **TSset**, z_i (i=1,2…, M), z_i are heights of the point assumed for creating DTM, M is the size of the set used for DTM construction.

It should be understood that m_0 is calculated for the set on the basis of which GRID dataset will be generated, so calculation of m_0 is conducted before generating grid points. The error is calculated for the actual data from the set containing points representing the TSset.

Before the processing by means of the proposed method OptD had begun, the mean error m_0 was calculated for TSset, m_0 = 0.795 m. This value allowed to establish the value of optimization criterion, which was taken as m_0 = 0.895m (about 10 cm larger, assuming error positional accuracy for the details of 1st class precision).

In the course of ALS point cloud processing by OptD-single method the following initial parameters were used: optimization criterion f{m_0 = 0.895}, width of measurement belts L = 0.600 m (the average distance between all points in the ALS set), tolerance in the D-P method t = 0.250 m (suggesting to the average distance between the points within the
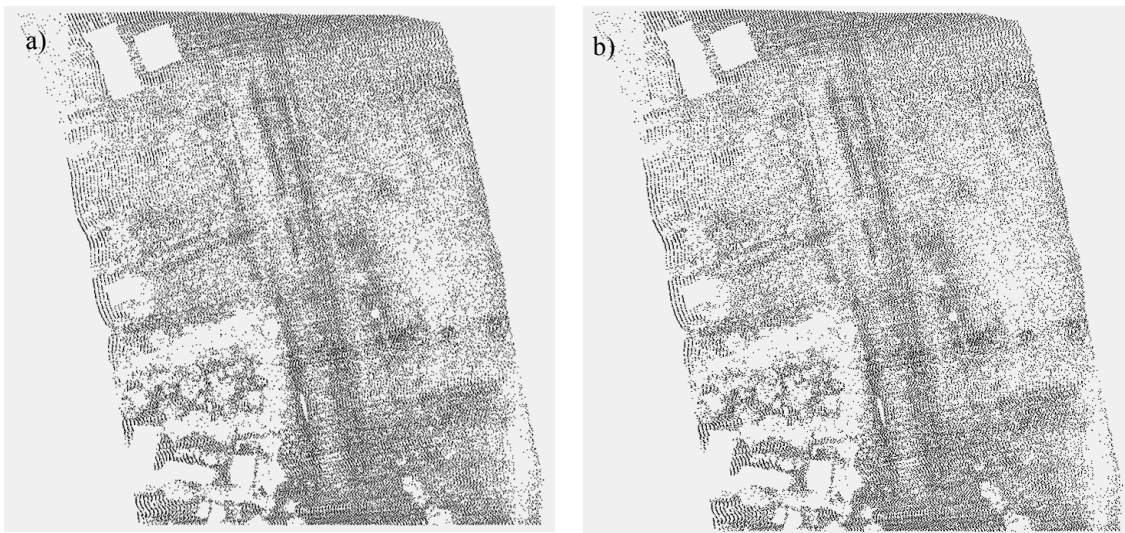
Fig. 6 a) TSset after OptD method. b) TSset after reduction with existing option (source: own in the CloudCompare v. 2.6.0).

Table 1 Comparison of the OptD method and existing option.

Total number of points in original ALS dataset **144500**			
Number of terrain points in TSset **108313**			
z_{mean} = 134.753 m			
Range (R) = 7.47m			
m_0 = 0.795 m			
OptD method		**existing algorithm of reduction**	
Number of terrain points in optimum ALS dataset	**58551**	Number of terrain points in satisfactory ALS dataset	**57123**
Total operation time [sec.]	**184**	Total operation time [sec.]	**540**
m_0	**0.895m**	m_0	**0.901m**
Zmean	**134.74m**	Zmean	**134.74m**
R	**7.47m**	R	**7.47m**

measurement belts of the original ALS data set). In the course of running the OptD method, the output tolerance 0.250 m was changed to 0.235 m (during iteration) in order to complete the established optimization criterion.

Application of the OptD method selected the optimum solution, which is presented in Figure 6a. For comparison in Figure 6b the result of a reduction for the satisfactory option is presented. The same width of measurement belts and the same tolerance in D-P method were used for optimum option.

Obtained two datasets were compared in next subsection.

4.2. COMPARISON OF THE RESULTS

Conducted tests resulted in two datasets: the first one was obtained by means of OptD-single method, the second by applying existing reduction algorithm. Results of processing in optimum and existing options are presented in Table 1.

Table 1 shows that in the optimum option a set of points to generate DTM is larger of 1428 points
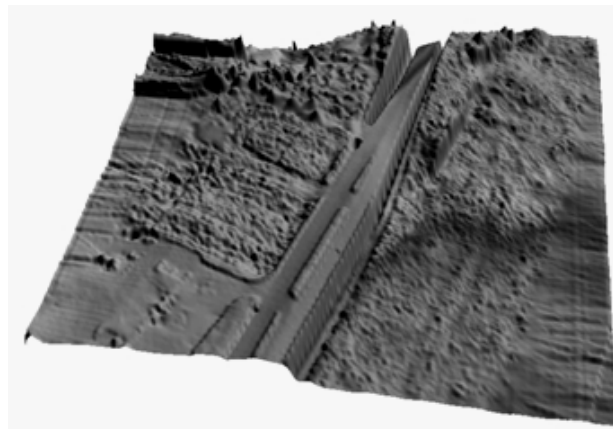
Fig. 7 DTM for all points in TSset (source: own in Surfer v.8).
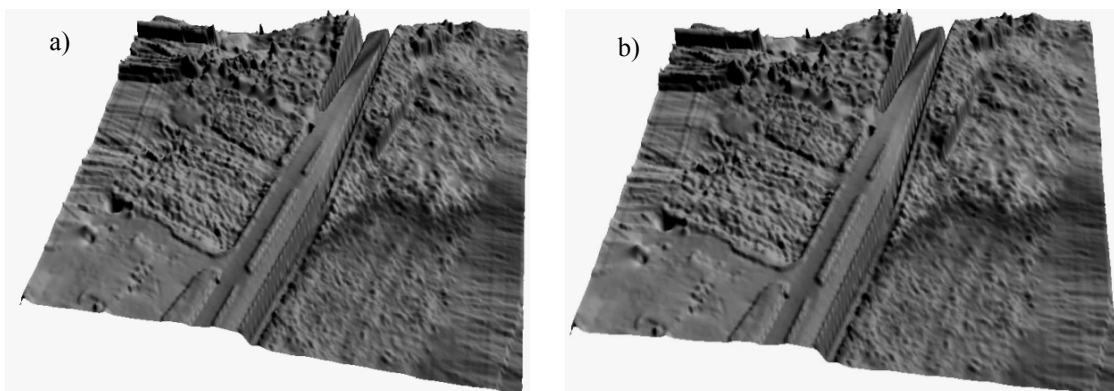
Fig. 8 a) DTM after OptD method. b) DTM after reduction with satisfactory option (source: own in Surfer v.8).

than in existing option. However, such a difference in the size of the set is not critical. It is important that in applying the OptD method which exactly meets our expectations written by means of optimization criterion. In contrast, using a method based on the reduction algorithm obtained result is rather random, depending on the input parameters necessary for the proper implementation of the method. With new OptD algorithm the desired number of set will be achieved

faster, without the need of restarting the algorithm with different values of input parameters.

4.3. DTM GENERATION

The entire set of TSset was used to built DTM (Fig. 7). On the basis of datasets obtained for optimal and satisfactory options DTMs were also generated. They are presented in Figures 8a and 8b, respectively. For all models, GRID with 1 m size was adopted.

Table 2 Comparison of generated DTMs.

Parameters:	DTM for TSset	DTM for OptD method	DTM for existing method
RMSE	0.071 m	0.083 m	0.084 m
D^2	-	0.980 m	0.980 m

RMSE (root mean square error) and coefficient of determination D were calculated for the generated DTMs. Coefficient of determination D is the measure of model adjustment (the closer to 1, the better the match of the model is another model). The results are presented in Table 2.

Parameters calculated for the assessing the accuracy of the DTM reveal a similar character. The RMSE is higher for DTM for existing option. The coefficient of determination is better for optimum option.

5. CONCLUSION

In this paper a new OptD method in the processing of LiDAR point cloud was proposed and tested. The algorithm of this method can reduce the dataset in terms of number of measuring points for a given optimization criterion.

Based on the analysis, following general conclusions can be stated:

1. Innovative OptD method is a simple in application method for data reduction, which takes into account optimization criteria.
2. The result of the implementation of the OptD method is an optimal dataset that can be used to generate DTM.
3. The OptD method fulfills all the expectations of reducing the size of the dataset without losing data necessary for the proper DTM generation.
4. The OptD method allows, without user intervention, to choose the most favorable, optimally reduced dataset. It is fully automatic.
5. The algorithm of new method implies, that only once the initial values of the parameters are introduced, the subsequence values of them in following iterations are automatically selected.
6. With new OptD algorithm the desired number of set will be achieved faster, without the need of restarting the algorithm with different values of input parameters.
7. The optimal dataset from optimum option meets the established criteria about mean error.
8. The use of OptD method enables to reduce the number of points in the point cloud with varying degrees of reduction in different areas. Thus, points, which are not essential in the generation process (e.g. modeling or DTM generation) will be removed from the measurement dataset.
9. The OptD method solves the problem of constant interference in the operation of the program (as it was in the previous version), and it allows the use of the algorithm for people who have no experience in the reduction of the set, but they want to obtain the optimally reduced set.

Detailed conclusions are as follows:
1. The coefficient of determination is the same for optimum option and satisfactory option. This prove, that despites the smaller number of 1428 points in the optimum option, the matching of the surface to the DTM generated from all TSset points for each option is the same.
2. RMSE is 1mm smaller for the optimum option.
3. The mean error of the reduced dataset for the optimum option is 0.895m, for satisfactory option - 0.901 m.

The proposed OptD method enables rapid, comprehensive and fully automated optimization to generate the DTM with an assumed quality. The OptD method allows for obtaining a representative sample of the original data set as an optimal set of LiDAR, the development of a mathematical schema of optimization procedures. The proposed new method provides new knowledge on the reduction of large datasets, so as not to lose the information necessary for the proper performance of a task. With OptD method, preparation of the data for the DTM construction was more accurate and less time-consuming. It allows for the effective DTM generation and reducing the time and cost of LiDAR point cloud processing, what in turn enables to conduct efficient analyses of acquired information resource.

The next step of research in the application of optimization criteria to reduce datasets is the testing of the OptD-mutli method.

REFERENCES

Axelsson, P.: 2000, DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing, XXXIII/4B, Amsterdam.

Bauer-Marschallinger, B., Sabel, D. and Wagner, W.: 2014, Optimisation of global grids for high-resolution remote sensing data. Computers & Geosciences, 72, 84–93. DOI: 10.1016/j.cageo.2014.07.005

Błaszczak, W.: 2006, Optimization of large measurement results sets for building data base of spacial information system. Doctors thesis, University of Warmia and Mazury in Olsztyn.

Błaszczak-Bąk, W.: 2012, The impact of optimizing the number of points of ALS data set on the accuracy of the generated DTM. Technical Sciences, 15(2), 256–278.

Błaszczak-Bąk, W., Janowski, A., Kamiński, W. and Rapiński, J.: 2011a, Optimization algorithm and

filtration using the adaptive TIN model at the stage of initial processing of the ALS point cloud. Canadian Journal of Remote Sensing., No. 37(6), 583–589. DOI: 10.5589/m12-001

Błaszczak-Bąk, W., Janowski, A., Kamiński, W. and Rapiński, J.: 2011b, ALS data filtration with fuzzy logic. Journal of Indian Society of Remote Sensing, 39, 591–597. DOI: 10.1007/s12524-011-0130-2

Błaszczak-Bąk, W., Janowski, A., Kamiński, W. and Rapiński, J.: 2015, Application of the Msplit method for filtering airborne laser scanning datasets to estimate digital terrain models. International Journal of Remote Sensing, 36(09), 2421–2437. DOI: 10.1080/01431161.2015.1041617

Cempel, C.: 2000, Theory of systems engineering. Script. Department of Systems Dynamics and Vibroacoustics. Poznan University of Technology, Poznań.

Douglas, D.H. and Peucker, T.K.: 1973, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Canadian Cartographer, 10(2), 112–122.

Filipowicz, B.: 1998, Mathematical modeling of decision-making issues. Part I. Publishing House AGH, Kraków.

Hebel, M. and Stilla, U.: 2008, Pre-classification of points and segmentation of urban objects by scan line analysis of Airborne LiDAR data. The International Archives of the Photogrammetry and Remote Sensing and Spatial Information Sciences, XXXVII(B3a), 105–110.

Jastriebow, A. and Wciślik, M.: 2004, Optimization- theory, algorithms and their implementation in Matlab. Publishing House PŚ, Kielce.

Marcinkowski, J.: 2008, Optimizing Multi-in: Operations Research. Publishing: Sikora, PWE, Warsaw.

Reitberger, J., Schnör, C.; Krzystek, P., Stilla, U.: 2009, 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), 561–574. DOI:10.1016/j.isprsjprs.2009.04.002

Saeedi, S., Samadzadegan, F. and El-Sheimy, N.: 2009, Object extraction from lidar data using an artificial swarm bee colony clustering algorithm. ISPRS, 37, 3/W4,133–138.

Sithole, G. and Vosselman, G.: 2005, Filtering of airborne laser scanner data based on segmented point clouds. Geo-Information Science, 66–71.

Tovari, D. and Pfeifer, N.: 2005, Segmentation based robust interpolation–a new approach to laser data filtering. IAPRSSIS, 363W19, 79–84.

Visvalingam, M. and Whyatt, J.D.: 1992, Line generalization by repeated elimination of point. Cartographic Information Systems Research Group. University of Hull.

Vosselman, G.: 2008, Analysis of planimetric accuracy of airborne laser scanning surveys. The International Archives of the Photogrammetry and Remote Sensing and Spatial Information Sciences, XXXVII(B3a), 99–104.

Vosselman, G. and Maas, H.: 2010, Airborne and terrestrial laser scanning. Whittles Publishing.