THE IMPACT OF OPTIMIZING THE NUMBER OF POINTS OF ALS DATA SET ON THE ACCURACY OF THE GENERATED DTM

Wioleta Błaszczak-Bąk

Institute of Geodesy
University of Warmia and Mazury in Olsztyn

A b s t r a c t

Airborne laser scanning technology delivers the result of the survey in the form of a point cloud. In order to construct a digital terrain model, it is necessary to perform filtration, which consists in separating data reflecting the relief features from the data reflecting situational details. In view of the very large amount of data in the survey data set, as well as the time consumption and difficulty in automatic filtration of the point cloud, it is possible to apply an optimization algorithm reducing the size of the point cloud while deriving a digital terrain model. This study presents the stages of compiling an airborne laser scanning point cloud using filtration and optimization. The filtration was carried out using the adaptive TIN model and the method of robust moving surfaces, while optimization was carried out with the application of an already existing algorithm to reduce the size of the survey data set. The effect of reducing the size of the data set on the accuracy of the generated DTM was tested and empirical and numerical tests have been performed.

Introduction

Airborne Laser Scanning (ALS) is a survey technology used for quickly gathering a set of coordinates concerning spatial field points of the surveyed area. The most important parameters describing the data obtained from the airborne laser scanning include: scan angle, flying height, scan rate and point density (BRIESE, PFEIFER 2001). Very frequently, the point density obtained is not adjusted for the needs of the planned compilations. We usually deal with a very high (too high) number of points, therefore it is necessary to reduce it to ensure effective, less labour-consuming and less time-consuming generation of a digital terrain model (DTM) (BŁASZCZAK, KAMIŃSKI 2007, BŁASZCZAK-BĄK et al. 2010b).

The method of deriving a DTM from the ALS point cloud consists of a preliminary processing stage and a main processing stage. The preliminary

processing stage consists of filtration (F), i.e. separating the data reflecting the relief features from the data representing situational details, while the main processing stage consists of DTM construction. In view of the fact that the point cloud makes the so-called large set of data an optimization algorithm is added to the methodology of DTM generation at the stage of preliminary processing (BŁASZCZAK, KAMIŃSKI 2007, BŁASZCZAK-BĄK et al. 2010a). The purpose of the optimization algorithm (O) is to reduce the size of the survey data set before (O – F) or after (F – O) filtration.

A modified methodology of ALS point cloud compilation can be therefore presented in the following diagram (Fig. 1).
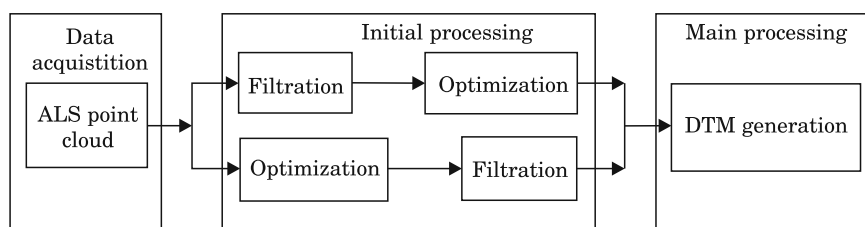

Fig. 1. A general scheme of the modified methodology for processing the ALS point cloud

The application of the O – F variant facilitates operation of the data filtration algorithm, shortens filtration time and influences the effective i.e. faster preparation of data for DTM construction without losing information required for proper implementation of the task. On the other hand, the application of the F – O variant makes it possible to reduce the data set obtained after filtration, i.e. directly used for construction of a DTM (BŁASZ-CZAK-BĄK et al. 2010a, BŁASZCZAK-BĄK et al. 2011).

The study presents the stages of DTM compilation with the application of the O – F variant, F – O variant and F variant for comparison. The optimization was carried out with the use of an optimization algorithm. The algorithm uses the Visvalingam – Whyatt (V – W) cartographic generalization method (VISVALINGAM, WHYATT 1992). The filtration was carried out with the use of the adaptive TIN (Triangular Irregular Network) model (AXELSSON 2000) and the method of robust moving surfaces (ELMQVIST 2002, SHUT 1976, HUBER 1981). The effect of reducing the size of the data set on the accuracy of the generated DTM was tested.

Optimization algorithm

The density of the ALS point cloud obtained is usually between a few and a few dozen points per 1 m^2 of the surveyed area. Such a data set makes the stage of main processing more difficult. Construction of a DTM requires a set

of data containing significant information about the area, while the points which do not contribute any significant information for DTM generation are unnecessary. Therefore, the point cloud can be reduced at the stage of preliminary processing by applying an algorithm optimizing the size of the survey result data set (BŁASZCZAK, KAMIŃSKI 2007, BŁASZCZAK-BĄK et al. 2010).

The optimization algorithm consists of the following stages:

– Defining survey strips in the XY plane, parallel to the Y axis.

– Selecting the method of cartographic generalization to reduce the size of the survey data set.

– Application of the chosen generalization method in each strip (in the YZ plane).

An important phase is the choice of the method in the survey strip, as well as selection of appropriate tolerance ranges in the method. The choice of tolerance ranges determines the degree of the ALS set reduction (number of removed points).

The V – W generalization algorithm was selected in the study which creates a generalized line of triangles from the nearest points. The surface of the calculated areas of triangles is compared to the area of the tolerance triangle, the size of which determines how many points are to be removed. The size of the tolerance triangle area is established by the user on the basis of statistical characteristics of the survey result data set. It is possible, for example, to assume that its value equals the area of the equilateral triangle with a side corresponding to a minimum distance between the points of the data set. If the area of the triangle defined on the basis of the survey results exceeds the area of the tolerance triangle, then the second point of the analysed triangle is maintained, otherwise it is removed.

ALS point cloud filtration

The literature contains many reports of filtration methods. The filtration methods presented in the literature are based, among others, on morphological filters (ZHANG 2003, VOSSELMAN 2001), gradient methods (HYYPPÄ et al. 2002, WACK, WIMMER 2002), modelling of active surfaces (ELMQVIST 2002), spline interpolation (BROVELLI et al. 2002) or the adaptive TIN model method (AXELSSON 2000).

The study applied two methods: an adaptive TIN model method and the robust moving surfaces method.

Adaptive TIN model method (ATIN)

The algorithm of the adaptive TIN model method is run iteratively and can be presented in the following way (AXELSSON 2000, BŁASZCZAK-BAK et al. 2010a):

– creating a GRID (Regular Raster Grid) net on the basis of the ALS point cloud, and selecting, in each mesh of the net, a point of the lowest altitude;

– carrying out triangulation of selected points (usually by Delauney's method);

– calculating, in each of the triangles obtained, the distances and angles between points situated within its vertical projection, and on their basis, the medians of the distances and medians of the angles;

– establishing new points belonging to the topographic area based on the criteria assumed (i.e. values of calculated angles and distances lower than the respective median);

– iterative return to the triangulation stage from selected points, until all points have been checked or the established threshold value of the analysed parameters is exceeded.

Robust moving surface (MS)

While carrying out filtration using the moving surface method, we can use a polynomial of the first degree in the following form:

$$z(x, y) = a + bx + cy \qquad (1)$$

where:
x, y, z are coordinates of points obtained from ALS survey, a, b, c are polynomial coefficients. The present study applies a polynomial of the first degree for analysis purposes, therefore, the equation of adjustments can be written in the following form:

$$V_i = a + bx_i + cy_i - z_i \qquad (2)$$

where:
$i = 1, 2,..., k$ (k is the number of observations in analyzing data set).

Equation (2) can also be written as:

$$\mathbf{V} = \mathbf{A}\hat{\mathbf{X}} - \mathbf{z} \qquad (3)$$

where:
$\mathbf{V} = [V_1, V_2,..., V_k]^T$ residuals vector, $\hat{\mathbf{X}} = [\hat{a}, \hat{b}, \hat{c}]^T$ is estimated polynomial coefficients vector, $\mathbf{z} = [z_1, z_2,..., z_k]^T$ is a vector of heights of points obtained from ALS, $\mathbf{A}$ is a matrix of coefficients.

Solving equation (3) with the least squares method, yields the estimated parameters vector:

$$\hat{\mathbf{X}}^{LS} = (\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P}\mathbf{z} \tag{4}$$

where:

$\mathbf{P} = \mathrm{Diag}(p_1,..., p_k)$ diagonal weight matrix, in which, $p_i = \dfrac{1}{D_{P_iP_0}}$, $D_{P_iP_0}$ is a distance between P_i and the point with interpolated height P_0, $i=1,...,k$, k is the number of points being processed. Estimated parameters can be derived by means of robust estimation (HUBER 1981, KRAUS, PFEIFER 1998). In robust estimation the weights of observations p_i, which standardized residuals $\bar{V}_i = \dfrac{V_i}{m_{V_i}}$ (where $m_{V_i} = \sqrt{\mathbf{P}^{-1} - \mathbf{A}(\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^T}$ is mean error of residuum) are outside of a certain range $<-g; g>$ are damped by dampening function. In presented research, the one sided form of weight function according to Huber's method is used (HUBER 1981). The function is written as:

$$w_i(V_i) = \begin{cases} 1 & \text{when } |\bar{V}_i| \le g \\ \dfrac{g}{\bar{V}_i}\, sng(V_i) & \text{when } |\bar{V}_i| > g \end{cases} \tag{5}$$

Calculations in robust estimation are carried out in an iterative process, in which equivalent weights $\bar{p}_i$ are determined from the following relation:

$$\bar{p}_i = p_i w_i (\bar{V}_i) \tag{6}$$

Polynomial parameters resistant to outliers (objects belonging to the area coverage) are obtained by performing the following iterative procedure:

$$\hat{\mathbf{X}} = (\mathbf{A}^T\bar{\mathbf{P}}^{l-1}\mathbf{A})^{-1}\mathbf{A}^T\bar{\mathbf{P}}^{l-1}\mathbf{z} \tag{7}$$

where:
$\bar{\mathbf{P}} = \mathrm{Diag}(\bar{p}_1, \bar{p}_2,..., \bar{p}_l)$ is an equivalent weight matrix, ($l = 1, 2,...,$ is the number of iterations). Results obtained by the least squares method: $\hat{\mathbf{X}}^0 = \hat{\mathbf{X}}^{LS}$ usually provide a starting point for an iterative process.

Practical tests

The modified methodology of ALS point cloud processing was applied to reduce the number of observations in a point cloud from surveying a section of Olsztyn. The tested large dataset comes from Visimind surveying system, which comprises of Riegl LMS-Q240 laser, GPS Topcon, IMU, and digital cameras. The laser scanning angle equaled 60 degrees, with a resolution of 10 000 Hz. The scan was made during a helicopter flight with the speed of around 50 km/h at the height of around 70 m. The subset of ALS data used for digital tests contains 80589 points. Point density is about 25 points per 1 m^2. The subset is presented in Fig. 2.

a   *b*
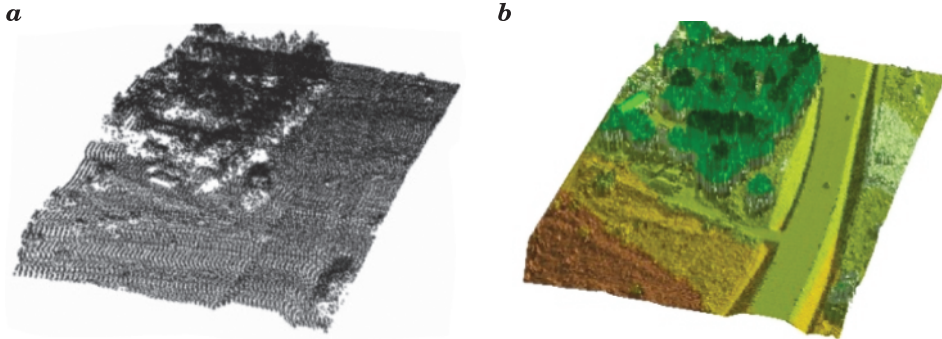

Fig. 2. Perspective presentation the original ALS point cloud: a) points, b) TIN model

Variant O – F

The size of the presented data set was reduced by applying an optimization algorithm. The optimization algorithm used the V – W method. During the reduction, the assumed width of search belt was 2 m and the tolerance triangle surface was 0.05 m^2. The width of search belt has been establish on results of earlier studies conducted with similar datasets. Area of tolerance triangle is an area of equilateral triangle with side equals mean distance between points from dataset (0,12 m). Point density after optimization is about 12 points per 1 m^2. The optimized ALS data set is presented in Fig. 3.

The optimized ALS point cloud was then filtrated using the ATIN method and the MS method. The ATIN method assumed that the size of the grid in which the height of the lower value was searched was 10 m × 10 m. Triangulation of selected points was made by Delauney's method in 84 iterations. In the MS method, a 10 m × 10 m square was used for calculation purposes, while the acceptable value of estimated adjustment was 3 m.
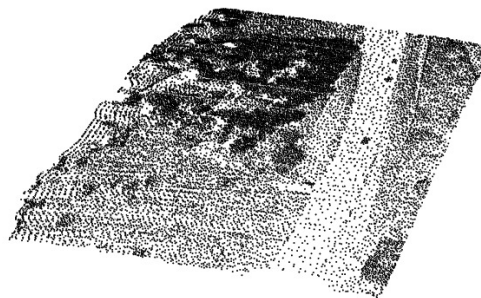
Fig. 3. Perspective presentation the ALS data set after optimization

Results of processing with the use of variant O – F are presented in Fig. 4 and Fig. 5.
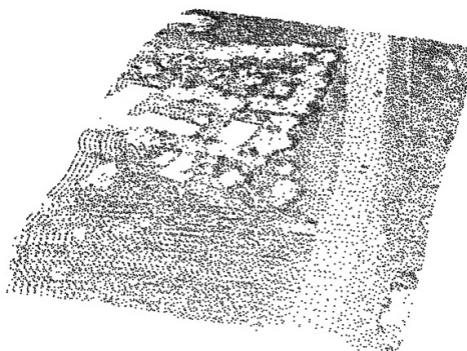

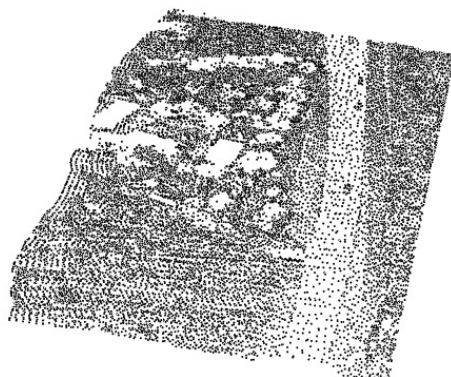Fig. 4. The optimized data set after ATIN model method filtration

Fig. 5. The optimized data subset after MS method filtration

Variant F – O

The original point cloud has been subjected to filtration by ATIN method and MS method. For filtration the same parameters as in variant O – F were applied.

In ATIN method triangulation of selected points was made in 172 iterations. Filtered data sets are shown in Fig. 6 and Fig. 7.
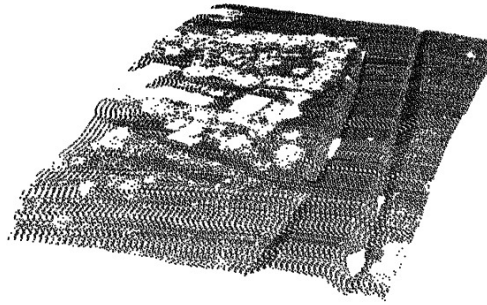

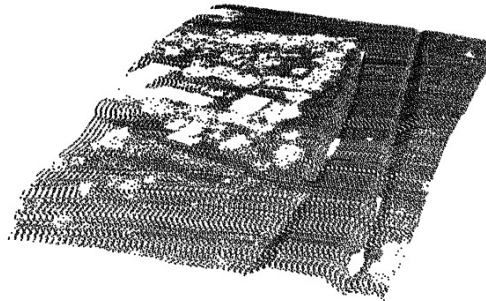Fig. 6. The original data set after ATIN model method filtration

Fig. 7. The original data set after MS method filtration

Filtered data sets were optimized using V – W method. During the reduction, the assumed width of search belt was 2 m and the tolerance triangle surface was 0.05 m^2 (as in O – F variant). Results are presented in Fig. 8 and Fig. 9.
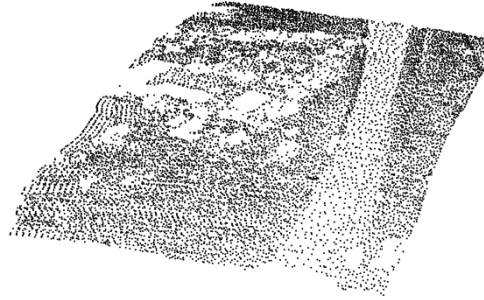
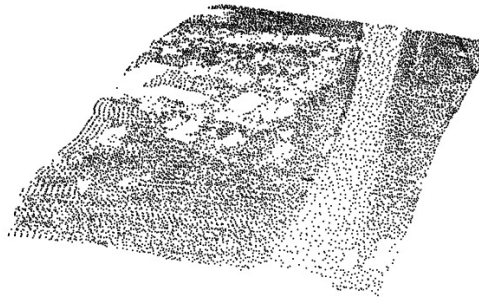Fig. 8. The data set after ATIN method filtration and after optimization

Fig. 9. The original data set after MS method filtration and after optimization

Analysis of results

The table below presents the number of points obtained after O – F variant and F – O variant.

In the case of F – O variant the application of the optimization algorithm has reduced the size of the data set after filtering by ATIN and MS to about 24%. The optimization was performed on the terrain points, which were directly used for DTM generation. In the case of O – F variant optimization was performed on the original (whole) data set (points showing the terrain and points showing situational details) and the point cloud is reduced to about 46%. Then the optimized sets were filtered by ATIN and MS method to obtain sets of points showing the area with the number of 15 515 points and 17 447 points, respectively.

Four data sets were obtained for such optimization and filtration parameters. For variant F – O there are very similar numbers of points (12 075 and 12 436 for ATIN and MS), while for variant O – F the number of points in the collection varies by about two thousand points.

The number of points in data subsets

Variant F – O			Variant O – F		
Total number of points in original ALS data set 80587					
Filtration	number of terrain points in ALS data subset after ATIN method filtration	48550	total number of points in ALS data subset after optimization	37080	Optimization
	number of terrain points in ALS data subset after MS method filtration	49765			
Optimization	number of terrain points in ALS data subset filtered by means of ATIN method after optimization	12075	number of terrain points in optimized ALS data subset after ATIN method filtration	15515	Filtration
	number of terrain points in ALS data subset filtered by means of MS method after optimization	12436	number of terrain points in optimized ALS data subset after MS method filtration	17447	

DTM generation

Calculation results of two variants were used to generate DTMs.

The O – F variant generated DTM$^{ATIN\ O\text{-}F}$ and DTM$^{MS\ O\text{-}F}$ and the F – O variant DTM$^{ATIN\ F\text{-}O}$ and DTM$^{MS\ F\text{-}O}$. To compare the DTMs generated from the two datasets: O – F and F – O, there were also two DTMs generated from dataset obtained by using standard methodology for ALS point cloud processing (called as F variant): DTM$^{ATIN\ F}$ and DTM$^{MS\ F}$. The DTM was generated by the Kriging method (SLUITER 2008), assuming a grid mesh of 1 m. DTMs generated are presented in Fig. 10, Fig. 11, Fig. 12, Fig. 13, Fig. 14 and Fig. 15.
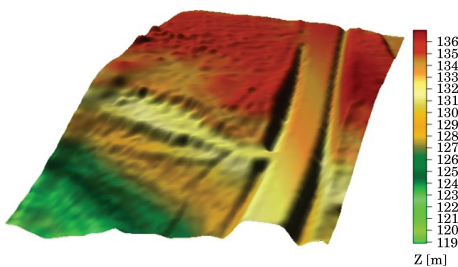
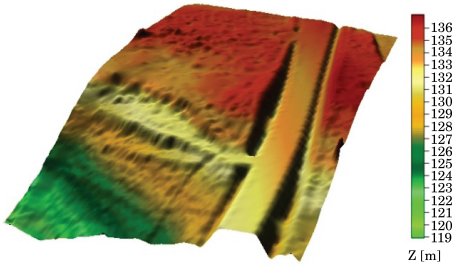Fig. 10. DTM$^{ATIN\ O\text{-}F}$
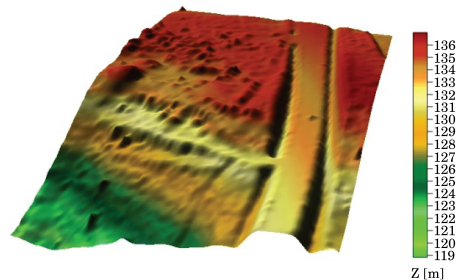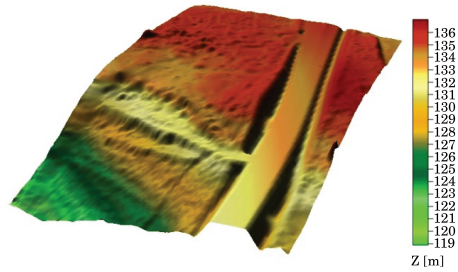

Fig. 11. DTM$^{MS\ O\text{-}F}$

Fig. 12. DTM$^{\text{ATIN F-O}}$
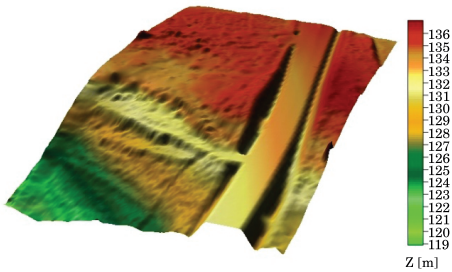

Fig. 13. DTM$^{\text{MS F-O}}$
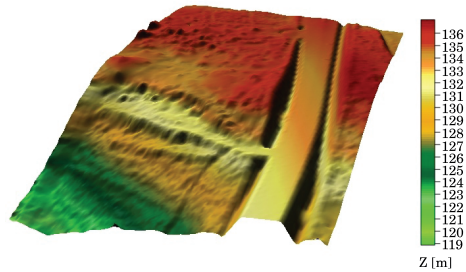

Fig. 14. DTM$^{\text{ATIN F}}$

Fig. 15. DTM$^{\text{MS F}}$

The generated DTMs represent the same area preserving the general character of the terrain. However, they differ in the level of detail. DTM$^{\text{ATIN}}$ has a more smoothed form when compared to DTM$^{\text{MS}}$, which was affected by the number of points in the filtrated data sets.

Assessment of DTM accuracy

Generated DTMs were subject to statistical comparison. The following parameters were used for DTM comparison (OKSANEN, SARJAKOSKI 2005, HEJMANOWSKA et al. 2008):

a) mean error:

$$m_0 = \sqrt{\frac{\Sigma(z_{\text{mean}} - z_i)^2}{k - 1}} \tag{8}$$

where:

z_{mean} is a mean height calculated from heights of both DTMs, z_i ($i = 1, 2, ..., k$) are heights of the point assumed for creating DTM, k is the size of the subset used for DTM construction,

b) range $R = z_{\text{max}} - z_{\text{min}}$, where z_{max} is the maximum height and z_{min} is the minimum height,

c) mean value of height difference:

$$\Delta h_{\text{mean}} = \frac{\Sigma_{i=1}^{k}(\tau)}{k} \qquad (9)$$

where:

$\tau = \tau_1, \; \tau_2, \; \tau_3 \text{ or } \tau_4$

$\tau_1 = Z_{\text{DTM}_{\text{ATIN O-F}}} - Z_{\text{DTM}_{\text{ATIN F}}}, \; \tau_2 = Z_{\text{DTM}_{\text{ATIN F-O}}} - Z_{\text{DTM}_{\text{ATIN F}}},$

$\tau_3 = Z_{\text{DTM}_{\text{MS O-F}}} - Z_{\text{DTM}_{\text{MS F}}}, \; \tau_4 = Z_{\text{DTM}_{\text{MS F-O}}} - Z_{\text{DTM}_{\text{MS F}}}$

d) root-mean-square error (RMSE), which describes the absolute altitude accuracy of DTM:

$$\text{RMSE} = \sqrt{\frac{\Sigma_{i=1}^{k}(\rho)^2}{k}} \qquad (10)$$

where:

$\rho = \rho_1, \; \rho_2, \; \rho_3 \text{ or } \rho_4$

$\rho_1 = \tau_1 - \Delta h_{\text{mean}}, \; \rho_2 = \tau_2 - \Delta h_{\text{mean}}, \; \rho_3 = \tau_3 - \Delta h_{\text{mean}}, \; \rho_4 = \tau_4 - \Delta h_{\text{mean}}$

e) coefficient of determination, which is the measure of model adjustment (the closer to 1, the better the match of the model to another model):

$$D_1^2 = \frac{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{ATIN O-F}}} - Z_{\text{mean}})^2}{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{ATIN F}}} - Z_{\text{mean}})^2} \text{ and } D_2^2 = \frac{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{ATIN F-O}}} - Z_{\text{mean}})^2}{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{ATIN F}}} - Z_{\text{mean}})^2}$$

$$D_3^2 = \frac{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{MS O-F}}} - Z_{\text{mean}})^2}{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{MS F}}} - Z_{\text{mean}})^2} \text{ and } D_4^2 = \frac{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{MS F-O}}} - Z_{\text{mean}})^2}{\Sigma_{i=1}^{k}(Z_{\text{DTM}_{\text{MS F}}} - Z_{\text{mean}})^2} \qquad (11)$$

The results of received parameters for all variants are presented in Table 2:

Table 2

The parameters values for all variants

Parameters	DTM$^{\text{ATIN O-F}}$	DTM$^{\text{ATIN F-O}}$	DTM$^{\text{ATIN F}}$	DTM$^{\text{MS O-F}}$	DTM$^{\text{MS F-O}}$	DTM$^{\text{MS F}}$
m_0	3.64 m	3.65 m	3.63 m	3.71 m	3.75 m	3.70 m
R	18.09 m	18.10 m	18.02 m	18.18 m	18.20 m	18.15 m
Δh_{mean}	0.05 m	0.05 m	–	0.05 m	0.05 m	–
RMSE	0.04 m	0.03 m	–	0.05 m	0.04 m	–
D^2	1.01	0.99	–	0.97	0.98	–

Parameters calculated for the purpose of assessing the accuracy of the DTM obtained reveal a similar character. The mean error is higher for DTM$^{\text{MS}}$, which is influenced by a range which is also higher in case of this surface. The

RMSE error has a very similar value in the case of all surfaces. On the other hand, the determination coefficient calculated indicates that the surfaces are well matched.

Conclusion

In this paper the effect of reducing the size of the data set on the accuracy of the generated DTM was tested. The presented research proposes a modified methodology of analysing the ALS point cloud in variant O – F and variant F – O. Filtration in both variants was executed with the application of the ATIN method and MS method. Optimization was made by using the V – W method. From the data sets after optimization and filtration four DTMs were generated: DTM$^{ATIN\ O-F}$, DTM$^{ATIN\ F-O}$ and DTM$^{MS\ O-F}$, DTM$^{MS\ F-O}$. To compare the DTMs generated from the variants: O – F and F – O, there were also two DTMs generated from dataset obtained by using standard methodology for ALS point cloud processing: DTM$^{ATIN\ F}$ and DTM$^{MS\ F}$. On the basis of the analyses presented above, the following conclusions are:

a) parameters for assessing the accuracy of the generated DTM are similar,

b) optimization does not adversely affect the accuracy of the generated DTM,

c) optimization reduces the number of points of an ALS point cloud and streamlines the methodology of processing a set of measurement for generating DTM,

d) generated DTMs exhibit a similar nature terrain.

Detailed findings can be summarized as follows:

a) mean error for DTMs generated from data sets filtrated with MS method is larger than mean error for DTMs generated from data set filtrated with ATIN method,

b) mean error for DTMs generated from data sets which were subjects only to filtration (ATIN and MS method) is smallest because the range was smallest,

c) mean value of the height difference is the same for all generated DTMs,

d) RMSE is the smallest for DTM$^{ATIN\ F-O}$, in comparison to other DTMs' errors the difference is not larger than 1 cm,

e) coefficient of determination is the best for DTM$^{ATIN\ F-O}$ and DTM$^{ATIN\ O-F}$ and the worst for DTM$^{MS\ F-O}$.

Studies conducted for two variants F – O and O – F show that optimization can be applied during ALS point cloud processing. It does not disturb DTM generation, has an impact on the time of data processing, because smaller dataset means shorter time of processing.

References

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

Błaszczak W., Kamiński W. 2007. *Data number reduction in measurement results set using optimization algorithm*. Proceedings of FIG Working Week, CD.

Błaszczak-Bąk W., Janowski A., Kamiński W., Rapiński J. 2010a. *Proposition of modification of aerial laser survey point cloud processing methodology*. Archives of Geomatics "New technology and instruments In survey".

Błaszczak-Bąk W., Janowski A., Kamiński W., Rapiński J. 2010b. *Modification of Lidar Point Cloud Processing Methodology*. Sydney The XXX FIG General Assembly and Working Week, Sydney 13.05 – 17.05 2010. CD.

Błaszczak-Bąk W., Janowski A., Kamiński W., Rapiński J. 2011. *ALS Data Filtration with Fuzzy Logic*. Journal of the Indian Society of Remote Sensing, 39(4): pp 591–597.

Briese C., Pfeifer N. 2001. *Airborne laser scanning and derivation of digital terrain models*. In Fifth Conference on Optical 3-D Measurement Techniques, Vienna, Austria.

Brovelli M.A., Bannata M., Longoni U.M. 2002. *Managing and processing LIDAR data within GRASS*. Proceedings of the Open source GIS – GRASS user conference, Trento.

Elmqvist M. 2002. *Ground surface estimation from airborne laser scanner data using active shape models*. ISPRS, Commission III, Symposium Photogrammetric Computer Vision, September 9–13, Graz, pp. 114–118.

Hejmanowska B., Drzewiecki W., Kulesza Ł. 2008. *The quality of Digital Terrain Models*. Archives of Photogrammetry and Remote Sensing, 18: 163–175.

Huber P.J. 1881. *Robust Statistics*. John Wiley and Sons.

Hyyppä J., Pyssalo U., Hyyppä H., Samberg A. 2002. *Elevation accuracy of laser scanning – derived digital terrain and target models in forest environment*. International Archives of Photogrammetry and Remote Sensing, XXXIV/ 3A, Graz.

Kraus K., Pfeifer N. 1998. *Determination of terrain models in wooded areas with airborne laser scanner data*. ISPRS Journal of Photogrammetry and Remote Sensing, 53: 193–203.

Oksanen J., Sarjakoski T. 2005. *Error propagation of DEM-basd surface derivatives*. Computers and Geosciences, 31: 1015–1027.

Schut G.H. 1976. *Review of interpolation methods for digital terrain models*. XII th Congress of the International Society for Photogrammetry. Helsinki .

Sluiter R. 2008. *Interpolation methods for climate data: literature review*. De Bilt, Royal Netherlands Meteorological Institute (KNMI).

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

Vosselman G. 2001. *Adjustment and filtering of raw laser altimetry data*. OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models, Stockholm.

Wack R., Wimmer A. 2002. *Digital terrain models from airborne laser scanner data – a grid based approach*. International Archives of Photogrammetry and Remote Sensing, XXXIV/3B, Graz.

Zhang K. 2003. *A progressive morphological filter for removing nonground measurements from airborne LIDAR data*. IEEE Transactions on Geoscience and Remote Sensing, 41(4): 872–882.