Data Number Reduction in Measurement Results Set Using Optimization Algorithm

Wioleta BŁASZCZAK and Waldemar KAMIŃSKI, Poland

Key words: optimization, initial processing, large data set.

SUMMARY

Rapidly developed automated data gathering techniques and methods like LIDAR (Light Detection And Ranging), GPS, integrated barymetric measurements consisted of GPS and digital sounder, are often used for obtaining observations for creation Digital Terrain Model (DTM). Methods mentioned above produce *large data sets*, which may cause problems for users at the stage of main processing, in which DTM is generated. Thus, there is a need to resample and reorganize the data sets before processing by using optimization algorithm.

Optimization algorithm allows to make more effective main processing, what results in shorter time of studies and drop of the expenses of completed studies and analyses in GIS. In this paper authors presented solutions of application of the optimization algorithm, in which known generalization methods were used to reduce the size of data sets. Optimum sets was used then for creation DTM.

STRESZCZENIE

Szeroko rozwinięte automatyczne metody pozyskiwania danych takie jak skanning laserowy, GPS, zintegrowane pomiary batymetryczne składające się z zestawu odbiornika GPS i cyfrowej echosondy ultradźwiękowej, najczęściej wykorzystywane są do gromadzenia obserwacji w celu budowy Digital Terrain Model (DTM). Wymienione metody pozyskiwania danych dostarczają tzw. *dużych zbiorów danych*, mogą niekiedy wywoływać problem ich opracowania w etapie przetwarzania głównego, w którym następuje generowanie DTM. Uzasadnionym wydaje się zatem, zastosowanie w etapie przetwarzania wstępnego odpowiedniego algorytmu zmniejszającego liczebność zbioru.

Zmniejszenie liczby danych w oryginalnym zbiorze wyników pomiaru pozwoli na efektywne przetwarzanie główne i wpłynie na skrócenie czasu opracowania, a tym samym i obniżenie kosztów kompleksowych opracowań oraz analiz w systemach informacji przestrzennej. W pracy zaprezentowano wyniki przetwarzania dużego zbioru danych algorytmem optymalizacji, wykorzystującym metody generalizacji kartograficznej. Zoptymalizowany zbiór danych posłużył w dalszej kolejności do budowy DTM.

Data Number Reduction in Measurement Results Set Using Optimization Algorithm

Wioleta BŁASZCZAK and Waldemar KAMIŃSKI, Poland

1. INTRODUCTION

Development of data collection methods, involving a decrease in time intensity of observations gathering and increase number of measurement results leading as a consequence to increasing the capacity of data collection media, can generate the so-called ***large data sets.*** Generally, the process of obtaining and processing data in the GIS can be presented in the form of the following chart (figure 1):

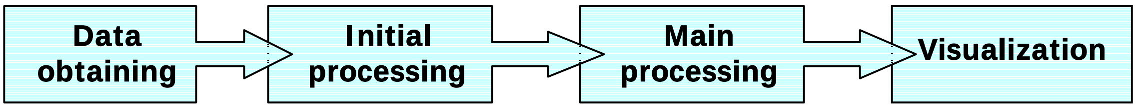

Figure 1. General chart of measurement data sets processing

Large data sets can cause difficulties at the stage of main processing. To avoid that inconvenience, the number of elements in the set can be decreased so that the processed set consists of observation results that are necessary for further use only.
Optimization of data numbers in a large data set can be effected, among others, by means of reduction. The reduction in the number of observations in the original set of measurement results is possible through application of different data elimination methods, including optimization algorithm application during initial data processing [1], [2], [3]. In the proposed algorithm data reduction occurs with application of known generalization methods. One of them is the Douglas-Peucker method (*'D-P')* [4].

The proposed measurement results set optimization algorithm is intended mainly for reduction of ***large data sets*** prior to importing them by the software. The algorithm are applied at the initial processing stage. That cause main processing more effective and results in shorter time of elaboration with data, and consequently cost of completely elaboration and analyses in GIS will decrease[2].

The basic criteria of the proposed ***large data sets*** optimization algorithm is decreasing the number of set elements through reduction without loss of data necessary for correct performance of land survey task.

2. MAIN STAGES OF OPTIMIZATION ALGOTITHM

Algorithm for reduction of the number of sets proposed in the paper by [1], [3] consists of the following steps, which are presented on figure 2.

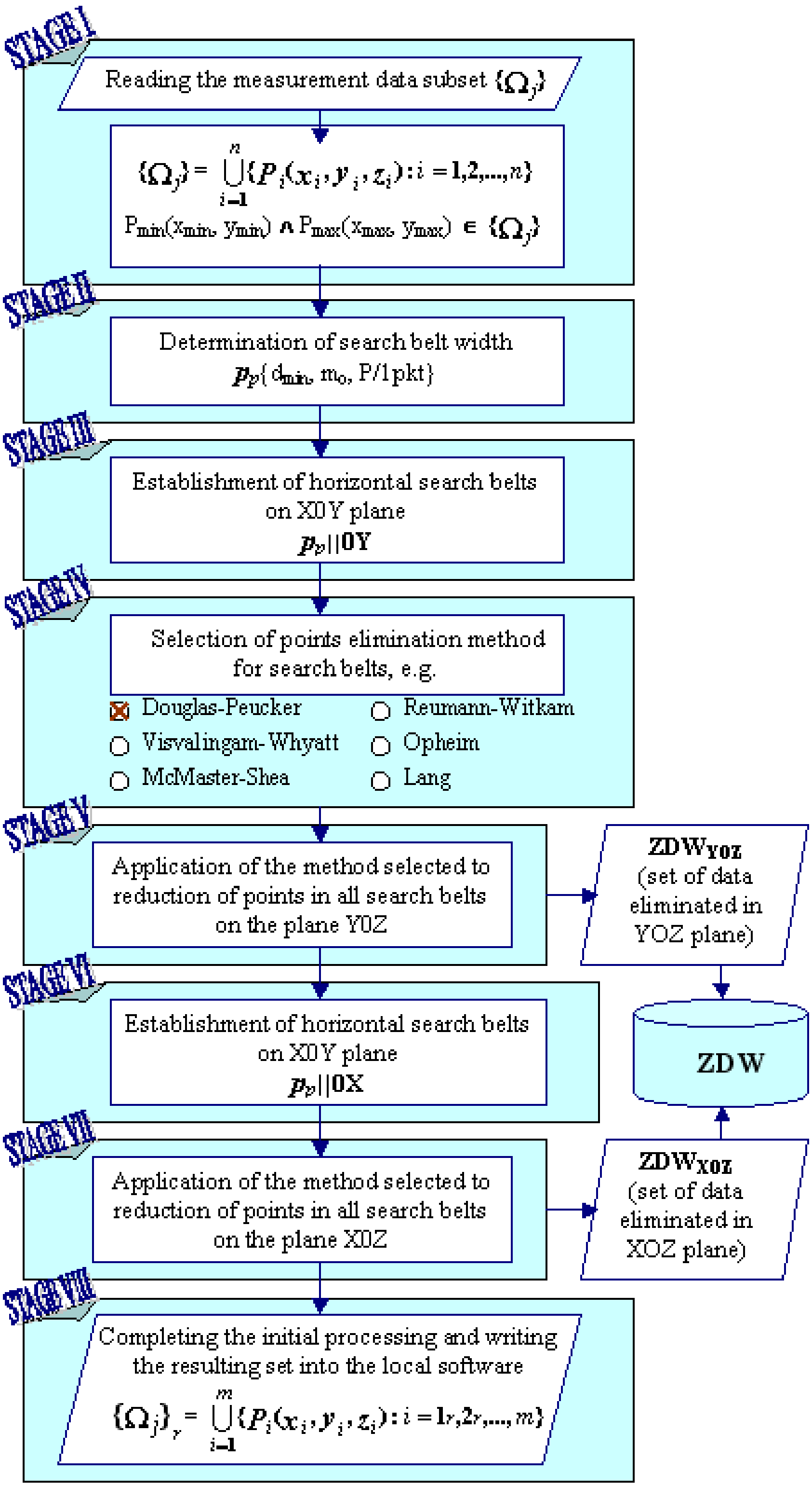

Figure 2. Main steps of algorithm for reduction of large data sets components number

Algorithm for reduction of the number of sets feeding the spatial information systems' databases proposed in the paper by *Błaszczak 2005*, consists of the following stage:

Stage I - Reading of the subset $\{\Omega_j\}$, where j=1,2,...,N and means the number of subset in which the reduction process will be carried out, is imported. At that stage the points of subset with the extreme coordinates $P_{min}(x_{min}, y_{min})$, $P_{max}(x_{max}, y_{max})$ must be defined. Points P_{min} and P_{max} are necessary for determination of the processing area extent.
Stage II – Determination of the width of the search belt within which the number reduction takes place. The width of the search belt p_p depends, among others, on:

- ✓ d_{min} – minimum distance between set points situated on the plane X0Y,
- ✓ m_o – accuracy of measurement points position,
- ✓ P/1pkt – processing area P per one measurement point.

Stage III - Establishment of horizontal search belts in X0Y plane p_p (belts parallel to 0Y axis).
Stage IV - Selection of the points elimination method in search belts [4], [5], [6], [7], [8], [9]. The choice of the method depends, among others, on the purpose of the work and character of the object surveyed. In this paper Douglas-Peucker cartographic generalization method [4] was applied to reduce set points. The Douglas-Peucker method is one of the basic cartographic generalization methods for linear objects and as a consequence, it is applied in GIS software very frequently. At this stage the tolerance range should also be defined. The tolerance range is necessary in the number reduction process and it depends, among others, on accuracy of observations. The elimination process is implemented in Y0Z plane.
Stage V - Application of the method selected for reduction of points in all search belts in Y0Z plane. The elimination process is stopped after all the search belts have been tested.
Points eliminated from individual search belts are placed in the ZDW_{YOZ} (set of data eliminated in plane YOZ.)
Stage VI - Establishment of vertical search belts in X0Y plane (belts parallel of 0X axis).
Stage VII – Application of the method of points reduction selected at ***stage IV*** for reduction of points in all vertical search belts. Verification of points is done on plane X0Z. The points eliminated are recorded in ZDW_{XOZ} (set of data eliminated in plane XOZ). After testing all vertical search belts the elimination process is completed. The eliminated points recorded in sets ZDW_{YOZ} and ZDW_{XOZ} are finally recorded in set ZDW.
Stage VIII represents completing the initial processing and writing the resulting set $\{\Omega_j\}_r$ = $\bigcup_{i=1}^{m}\{P_i(x_i, y_i, z_i): i = 1_r, 2_r, ..., m\}$, where m>n, mark r means set or points after reduction $1_r, 2_r$,...m are points of the set after reduction, into the local software.

3. EXAMPLE – PRACTICAL APPLICATION

The proposed measurement data number optimization algorithm was applied to processing of fragment of the set of results from measurement of Świnoujście-Szczecin waterway bottom. The processing involved the set of 60857 x and y coordinates as well as depths z, determined using the multibeam probe integrated with DGPS. The tested area was 3462 m^2.
The following search belt width variants were used for reduction of the bottom measurement set: ***variant I*** 0,012m, ***variant II*** 0,024m, ***variant III*** 0,037m, ***variant IV*** 0,049m, ***variant V*** 0,050m, ***variant VI*** 0,100m. Tolerance range O_i values presented in TABLE 1 were applied for the number reduction by means of the D-P method.

TABLE 1

TOLERANCE RANGE O_i [m]													
O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9	O_{10}	O_{11}	O_{12}	O_{13}	O_{14}
0,0005	0,001	0,002	0,003	0,004	0,006	0,008	0,012	0,017	0,024	0,034	0,048	0,059	0,068

The proposed algorithm produced sets with different numbers of components is presented in TABLE 2.

TABLE 2

Tolerance range	60857 points in the oryginal set					
	variant I 'D-P'	*variant II* 'D-P'	*variant III* 'D-P'	*variant IV* 'D-P'	*variant V* 'D-P'	*variant VI* 'D-P'
O_1	58625	59264	59218	58907	59055	58878
O_2	57791	59107	59122	58889	59019	58871
O_3	55752	58767	59008	58842	58983	58854
O_4	53561	58396	58896	58834	58913	58838
O_5	51049	57890	58782	58755	58858	58826
O_6	45126	56862	58455	58631	58736	58784
O_7	39148	55314	58085	58474	58569	58732
O_8	29852	49113	54017	54735	54767	54729
O_9	24858	43047	52202	54114	54054	54701
O_{10}	21158	31636	44903	49791	49485	50707
O_{11}	18703	22779	34842	43593	43644	47114
O_{12}	17077	16507	23833	33567	33935	43693
O_{13}	16143	14080	18288	26210	26485	40385
O_{14}	15621	12680	14445	21415	21605	36955

From among many search belt width variants and tolerance range values applied for the purpose of this study, in this paper only the solution that satisfied the criteria of optimization is presented. The set of data which number is close to a half of the original set element, and the surface developed from the points of the set has the smallest average error was considered as the optimum solution. The data set chosen in that way was used for development of the digital terrain model.

The solution that satisfies the optimization criteria (marked green in TABLE 2) is the result of processing with optimization algorithm applying the following parameter: ***variant VI*** O_{14}. The average error of GRID type surface generated from points of the optimized set was m = 0,0178m.

4. DTM GENERATION

The selected optimal data set was used for generating the *DTM'D-P'* (figure 3b), which presents the fragment of Świnoujście-Szczecin waterway bottom after reduction of the number of points in the actual measurement results set. The generated model was compared to the *DTM'R'* (figure 3a), that represents the fragment of real form of the bottom. The DTM model was generated using the Surfer v. 8 software. On the basis of practical experience of the Maritime Office in Szczecin the GRID mesh size of 2m was assumed for all models generated. The DTM nodes were determined by Kriging method.

Additionally, for fragments of the sets of points, the system of contour lines presented in figure 3c (*‘D-P’*) and 3d (*‘R’*) was generated.

Comparing the course of contour lines presented in figure 3c we notice that their layout does not deviate significantly from the system of contour lines presented in figure 3d. Only in a few places we can see minor changes in the layout of contour lines. Based on the graphic presentations we can conclude that despite decreasing the number of elements in the set we do not observe significant disturbances in the DTM structure.

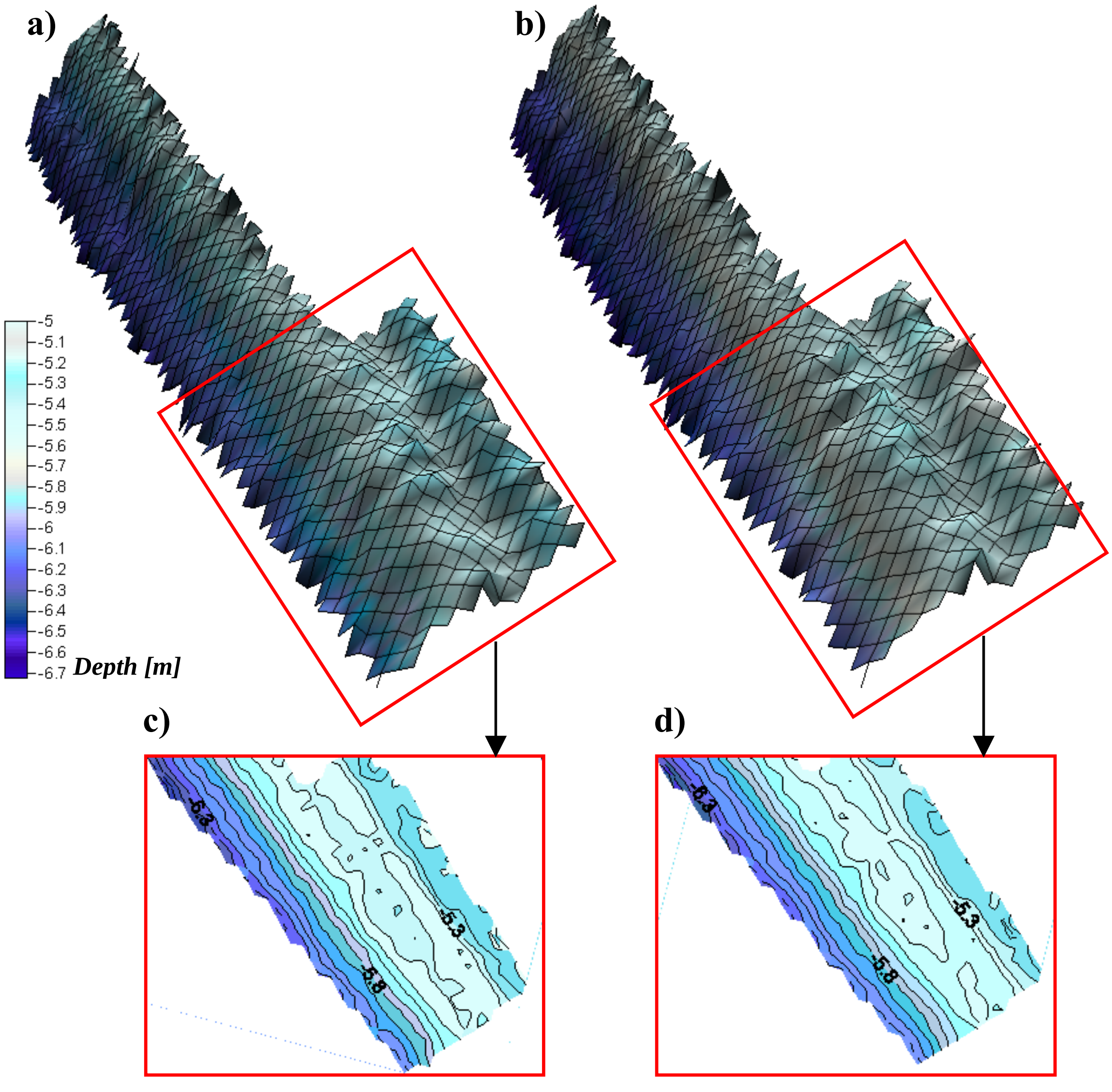

Figure 3. a) DTM ‘D-P’(36955 points; the fragment of Swinoujscie-Szczecin channel)
b) DTM ‘R’(60857 points; the fragment of Swinoujscie-Szczecin channel),
c) Contour model of the analyzed area (36955 points),
d) Contour model of the analyzed area (60857 points)

5. CONCLUSION

The initial processing stage with the proposed algorithm would allow to reduce the actual set of data by about 39%. Identification of the optimum solution was possible through application of the optimization criterion. The choice of that criterion depends mainly on the user and purpose of the work. In the presented figures we see very similar digital models of the bottom. As a consequence, it can be concluded that the selected reduced data set can be used in further detailed studies and analyses.
The procedure proposed in the paper requires further theoretical and empirical studies, in particular, studies conducted on actual sets of measurement results.

RERERENCES:

Błaszczak W., 2005: The measurement data preliminary processing in study of large information set. Technical Sciences. UWM Olsztyn.(in Polish).

Błaszczak W., Kamiński W., 2005: Line generalization using the Visvalingam-Whyatt method in the process of large data sets reduction. Scientific paperr of the Institute of Mining of Wrocław University of Technology. Geoinformation for everybody. XIX Autumn School of Geodesy. (in Polish)

Błaszczak W., Kamiński W, 2006.: Douglas-Peucker and Visvalingam-Whyatt methods in the process of large sets of observatin results reduction. 3rd IAG Symposium on Geodesy for Geotechnical and Structural Engineering and 12th FIG Symposium on Deformation Measurement. Baden-Austria. CD-R.

Douglas D. H., Peucker T. K., 1973: Algorithms for the Reduction of the Number of Points Required to Represent a Digitised Line or its Caricature. The Canadian Cartographer, 10(2).

Lang T., 1969: *Rules for the Robot Droughtsmen*. The Geographical Magazine, 42 (1).

McMaster R. B., Shea K. S., 1992: *Generalization in digital Cartography*. Assoc. Of Amer. Geographers, Washington.

Opheim H., 1981: *Fast reduction of a digitized curve*. Geo-Processing.

Reumann K., Witkam A. P. M., 1979: *Optimizing curve segmentation in computr graphics*. International Computing Symposium.

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

BIOGRAPHICAL NOTES

Ph. D. Wioleta Błaszczak – Assistant Porfessor at Institute of Geodesy
Professor Waldemar Kamiński – Vice Director at Institute of Geodesy

CONTACTS

Ph. D. Wioleta Błaszczak
Institute of Geodesy, Faculty of Geodesy and Land Management,
University of Warmia and Mazury in Olsztyn
Oczapowskiego Street 1, 10-957 Olsztyn
Olsztyn
POLAND
Tel. + 48 895234870
Fax. + 48 895134768
Email: wioleta.blaszczak@uwm.edu.pl

Professor Waldemar Kamiński, Ph. D.
Institute of Geodesy, Faculty of Geodesy and Land Management,
University of Warmia and Mazury in Olsztyn
Oczapowskiego Street 1, 10-957 Olsztyn
Olsztyn
POLAND
Tel. + 48 895234873
Fax. + 48 895134768
Email: waldemar.kaminski@uwm.edu.pl