DOUGLAS-PEUCKER AND VISVALINGAM-WHYATT METHODS IN THE PROCESS OF LARGE SETS OF OBSERVATION RESULTS REDUCTION

Wioleta Błaszczak, Waldemar Kamiński
Institute of Geodesy
University of Varmia and Mazury
Email: wioleta.blaszczak@uwm.edu.pl

Abstract: Geodetic surveys may procuce large amount of data which results in *large data sets*. Such data sets may cause problems for users with loading them to suitable software or making next stage of processing more difficult. In this paper authors propose modification of data processing algorithm, particularly relating to large data sets. Modification executes in stage of preliminary processing and rest on reduction of basic data set numbers. For reduction of data numbers two known generalisation methods has been used, namely: Douglas-Peucker method, Visvalingam-Whyatt method. Theorethical considerations are completed by practical example of application. Reached results encourage to further and detailed theoretical and empirical research.

1. Introduction

Sets accumulated as a result of geodesic survey automation, including *large data sets*, can sometimes cause a problem with processing at the stage of main processing. Generally, the process of obtaining and processing data in the SIS can be presented in the form of the following chart (fig. 1):

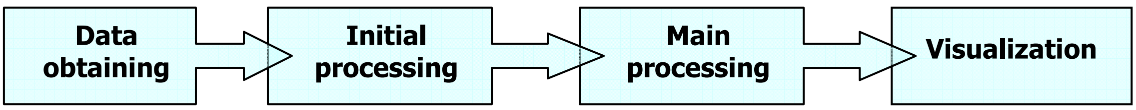

Figure 1: General chart of measurement data sets processing

Application of the existing algorithms decreasing the numbers of data in the sets seems almost a necessity considering the current status of measurement techniques development. As *large data sets* may hinder main processing or make it impossible, decreasing the numbers of data should occur, according to the authors, at the stage of initial processing. The reduction in the number of data can be achieved by applying algorithms of cartographic generalization. Two from among many known methods of cartographic generalization were applied for achievement of the goal of this study (reduction of the data set) Douglas-Peucker method [3] is the first of them and Visvalingam-Whyatt method [4] the other. Both algorithms were tested on an example of real bathymetric measurement results for Świnoujście-Szczecin

Navigation Channel. Application of Douglas-Peucker (further referred to as D-P) Visvalingam-Whyatt (further referred to as V-W) methods allowed decreasing the number of data and facilitated further processing of observations remaining after reduction. The results of algorithm operation were presented in the form of digital terrain models. The course of contour lines obtained from all points of the set as well as of the sets obtained after reduction of the number of observations was presented also. The results obtained encourage conducting further theoretical and empirical tests.

2. Proposal of initial data processing algorithm modification

Initial measurement data processing usually involves preparation of input data for spatial systems. During the initial processing process elimination of observations with gross errors takes place. The qualitative and quantitative characteristics of the set are also obtained. The process also aims at adjustment of the size and initial structure of primary data to the requirements of specific software. Measurement data processing generally takes place according to the process presented in fig. 1. Modification of that process presented in fig. 2 represents taking a *large set of data* into account at the stage of initial processing.

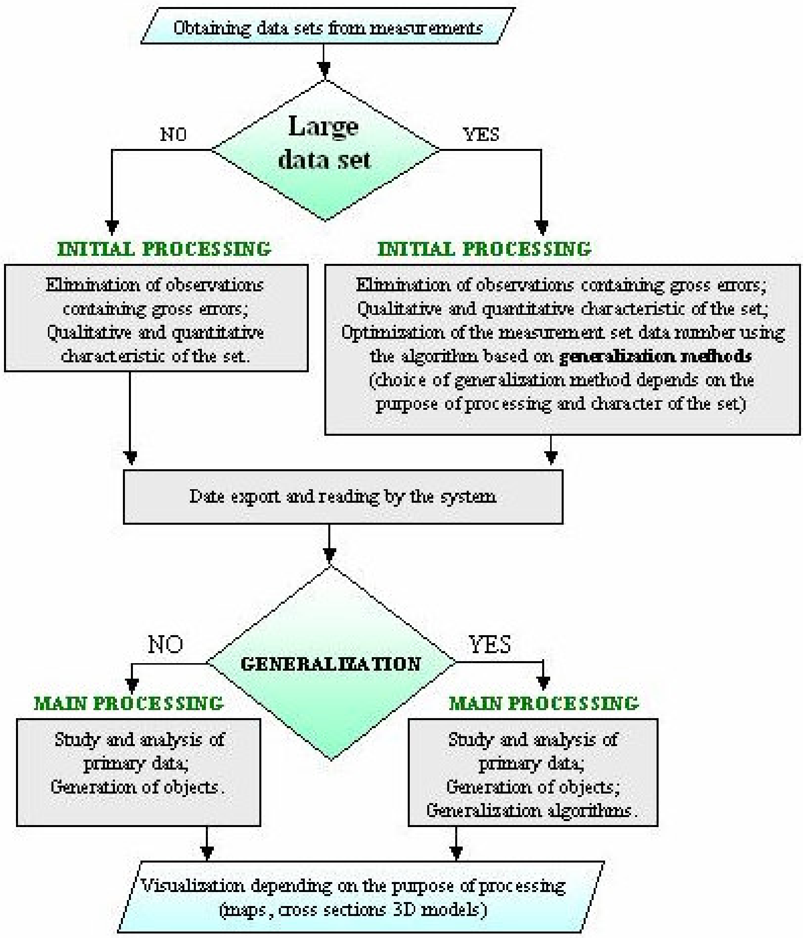

Figure 2: General chart of measurement data sets processing with large data sets and needs for generalization taken into consideration

3. Section Headings

Algorithm for reduction of the number of sets feeding the spatial information systems' databases proposed in the paper by [1], consists of the following steps:

1. Reading of the subset $\{\Omega_j\}$, where j=1,2,...,n means the number of the subset for which the reduction process will be carried out. During that stage the points with extreme coordinates $P_{min}(x_{min}, y_{min})$, $P_{max}(x_{max}, y_{max})$ have to be determined. Points P_{min} and P_{max} are necessary to determine the processing area range.
2. Determining the width of the search belt within which the reduction takes place. The width of the search belt depends, among others, on:
   - minimum distance between the points of the set in X0Y plane,
   - surface of the processed area per one measurement point.
3. Establishment of horizontal search belts in X0Y plane (belts parallel to 0Y axis).
4. Selection of the method for elimination of points in search belts. The elimination process is carried out in Y0Z plane. The method choice depends, among others, on the purpose of processing and character of the processed object. At that stage the tolerance range must also be determined. The tolerance range is necessary in the number reduction process and depends, among others, on the precision of observations.
5. In this study the two following methods were used for reduction of points:
   - Douglas-Peucker method,
   - Visvalingam-Whyatt method.
6. Application of the method selected for reduction of points in all search belts in Y0Z plane. The elimination process stops when all search belts have been tested.
7. Establishment of vertical search belts in X0Y plane (belts parallel of 0X axis).
8. Application of generalization method for reduction of points in all search belts on X0Z plane.
9. Completion of initial processing step and writing the resulting file to the local software.

4. Example of practical application

The proposed algorithm for reduction of the number of measurement data was applied for processing a fragment of the set containing results of Świnoujście-Szczecin channel bottom measurements. The processing involved the set of 60857 pairs of coordinates X and Y as well as depths Z determined using multiple beam sounder integrated with the DGPS. The area tested was 3462 m^2.

D-P and V-W methods were used in the algorithm. The detailed studies and analysis of results obtained were presented in papers by [1],[2]. This paper presents one of many solutions, i.e. the algorithm results based on the following assumptions:

- the search belt width equal to the double value of the least distance between measurement points in X0Y plane (d_{min}=0.006m),
- the tolerance range (section length) in the D-P method equal to 0.008m,
- the tolerance range (area of reference surface) in V-W method equal to 0.00003125m^2 (that value corresponds to a side of equilateral triangle of 0.008m).

The reduced data sets were used for generation of two variants of digital terrain models.

Variant I: The digital terrain model generated of *39148* points of the set obtained after applying the optimization algorithm based on the D-P method – further referred to as the *DTM'D-P'*.

Variant II: The digital terrain model generated of *51062* of the set obtained after applying the optimization algorithm based on the V-W method - further referred to as the *DTM'V-W'*.

Those models were compared to the digital terrain model generated on the basis of the real number of measurements set points, further referred to as the *DTM'R'*. Surfer v. 8 software was used for generation of the digital terrain models. On the basis of practical experience of the Maritime Office in Szczecin the GRID mesh size equal to 2 m was assumed for all the models. The GRID nodal points were determined by kriging.

Figure 3 below presents the *DTM'R'* model while figure 4 presents the contour model of the analyzed area.

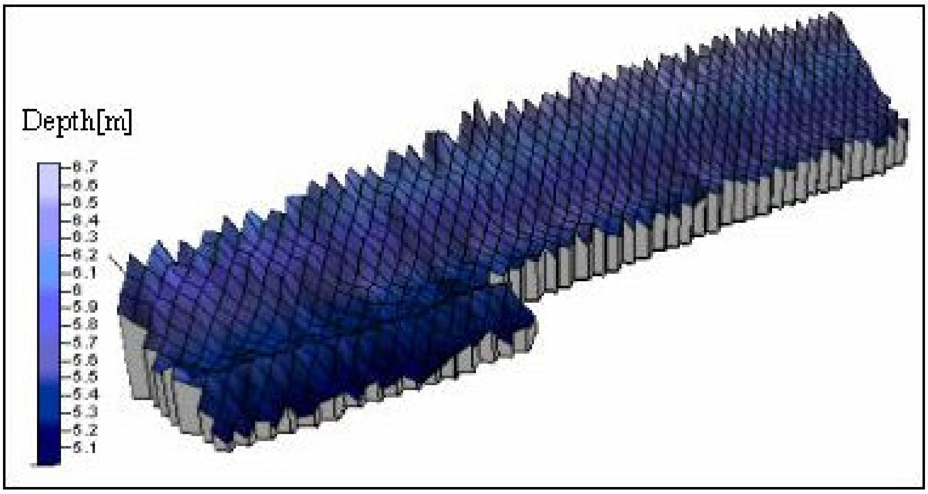

Figure 3: DTM'R' (60857 points; the fragment of Swinoujscie-Szczecin channel)

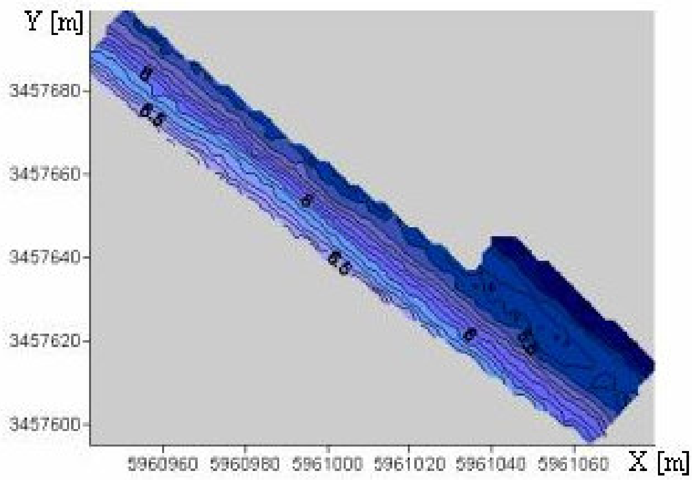

Figure 4: Contour model of the analyzed area

(60857 points; the fragment of Swinoujscie-Szczecin channel)

The further figures present variants *DTM'D-P'* and *DTM'V-W'* generated on the basis of reduced measurement data sets obtained as a result of the initial processing. The number of measurement points corresponding with *variant I* is 39148 and the *DTM'D-P'* generated on its base is presented in fig. 5. Figure 6 presents the systems of contour lines interpolated from all points of that set.

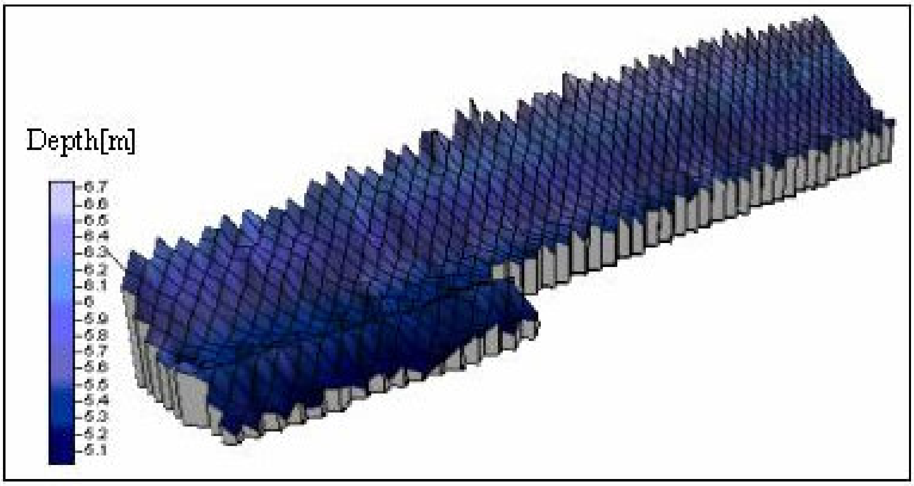

Figure 5: DTM'D- P' (39148 points; the fragment of Swinoujscie-Szczecin channel)

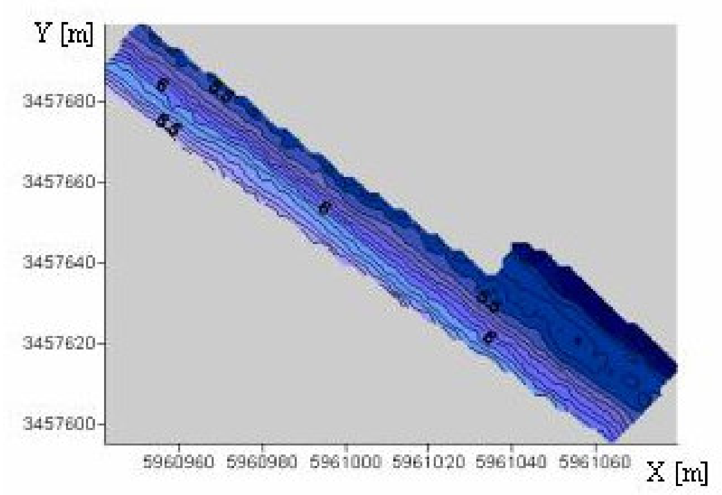

Figure 6: Contour model of the analyzed area

(39148 points; the fragment of Swinoujscie-Szczecin channel)

Figure 7 presents *DTM'V-W'* corresponding to *variant II* generated on the basis of the set of points obtained from optimization algorithm based on the V-W method. The system of contour lines generated on its basis is presented in fig. 8.

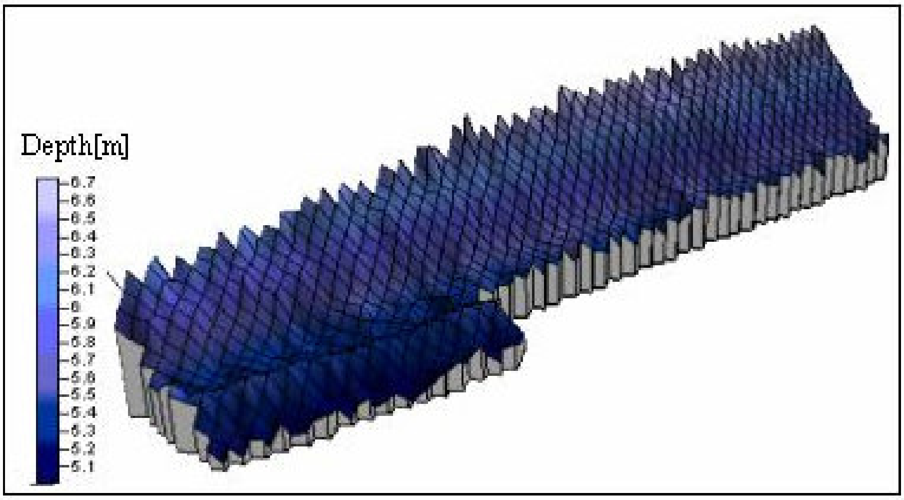

Figure 7: DTM'V-W' (51062 points; the fragment of Swinoujscie-Szczecin channel)

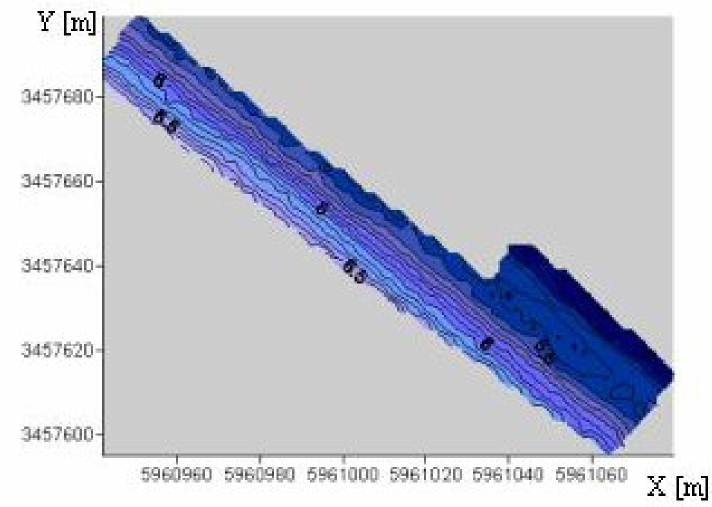

Figure 8: Contour model of the analyzed area

(51062 points; the fragment of Swinoujscie-Szczecin channel)

Comparing the courses of contour lines presented in fig. 6 and fig. 8 we can notice that the system does not deviate significantly from the system of bathymetric contour lines presented in figure 4. On the basis of graphic presentations alone we could conclude that we do not observe distortions in the courses of bathymetric contour lines presented in the figures with the decreasing number of points in the set.

5. Conclusions

In the presented figures highly similar courses of contour lines and digital terrain models for *variant I* and *variant II* can be seen even though each of them was interpolated on the basis of a different number of points in the reduced set. In *variant I* the difference between the actual number of set points and the number of set points after application of the initial processing algorithm is 21709 points while in *variant II* 9795 points. Those difference result from application of different set reduction methods. It can be concluded that in the analyzed example, in case of very similar tolerance values (for V-W method 0.008m for triangle side and for D-P method tolerance section equal to 0.008m) and the same search width, the V-W method reduces fewer points than the D-P method. Such a difference in the result of algorithm application does not cause significant deviations from the real system of contour lines. The results presented in this paper do not exhaust all the issues of *large data sets* reduction and it is difficult to formulate general conclusions on the basis of that experience. The issues presented above will be the subjects of further detailed theoretical and empirical analyses.

References:

[1] Błaszczak W.: The measurement data preliminary processing in study of large information set. Technical Sciences. UWM Olsztyn, 2005.

[2] Błaszczak W., Kamiński W.: 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, 2005.

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

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