Article

A Reduction Method for Bathymetric Datasets that Preserves True Coastal Water Geodata

Marta Wlodarczyk-Sielicka [1,*] , **Andrzej Stateczny** [2] **and Jacek Lubczonek** [1]

[1] Institute of Geoinformatics, Department of Navigation, Maritime University of Szczecin, 70-500 Szczecin, Poland

[2] Department of Geodesy, Gdansk University of Technology, 80-233 Gdansk, Poland

[*] Correspondence: m.wlodarczyk@am.szczecin.pl; Tel.: +48-513-846-391

check for
updates

Abstract: Water areas occupy over 70 percent of the Earth's surface and are constantly subject to research and analysis. Often, hydrographic remote sensors are used for such research, which allow for the collection of information on the shape of the water area bottom and the objects located on it. Information about the quality and reliability of the depth data is important, especially during coastal modelling. In-shore areas are liable to continuous transformations and they must be monitored and analyzed. Presently, bathymetric geodata are usually collected via modern hydrographic systems and comprise very large data point sequences that must then be connected using long and laborious processing sequences including reduction. As existing bathymetric data reduction methods utilize interpolated values, there is a clear requirement to search for new solutions. Considering the accuracy of bathymetric maps, a new method is presented here that allows real geodata to be maintained, specifically position and depth. This study presents a description of a developed method for reducing geodata while maintaining true survey values.

Keywords: big data applications; data processing; data visualization; neural networks; reduction; coastal waters

1. Introduction

Currently, bathymetric data are one of basic data types used in systems modeling of phenomena occurring in coastal zones. In general, the quality of these data is much more important than the physical models of the phenomena, as they have a major influence on the outcome of the simulation [1]. An example is the use of bathymetry for numerical analyses related to water quality prediction in coastal waters [2] or for coastal hydrodynamics modelling [3]. Undoubtedly, the use of bathymetric data in electronic navigational charts is also an important factor. The accuracy of these data in this case determines the safety of maritime transport, especially in coastal areas where shallow water is common. In this work, the authors focused on processing reduced datasets from real bathymetric measurements, which can be used in practically any system. This approach is quite different from the standard forms of modeling bathymetry from high-density data, which is usually a GRID structure.

In order to model the shape of the bottom of water areas via a bathymetric chart or digital spatial model, a set of points with known parameters (i.e., location and depth) are required. Bathymetric data are, therefore, elaborated using appropriate processing methods and are then presented in final hydrographic products, which can be digital terrain model or charts. These products are then used during monitoring and analysis of the transformations of coastal water bottoms. A hydrographer is responsible for the proper collection, preparation, and presentation of bathymetric data, traditionally associated with labor-intensive processing that is generally time consuming [4–6].

Hydrographic remote sensors like echosounders are normally used to collect information about the depth of a given area of water where costal, especially, shallow waters are mostly selected. These devices utilize a hydro acoustic wave to measure the distance between the transducer and the sea bottom, or the location of objects on the latter. The use of multibeam echosounders allows the collection of large volumes of information in a relatively short period of time [7–12]. Even gathering data with a single-beam echosounder [13] collects a huge amount of data, which can be called spatial big data. The processing of LiDAR (Light Detection and Ranging) data is also a big data reduction problem [14–17]. Modern bathymetric data acquisition technologies enable the collection of huge volumes of points while ensuring full coverage across a tested water area [18]. Bathymetric big data are a collection of datasets so large that it is difficult to work with using traditional data processing algorithms. Indeed, this approach significantly exceeds the minimum accuracy requirements specified by the International Hydrographic Organization in either S-44 [19] or S-57 standards [20].

Existing reduction techniques for bathymetric data use interpolated values in the form of a regular rectangle GRID with an assumed mesh size [21]. The development and application of new measurement techniques and methods of data processing mean that it is now possible to present much more accurate positional values for a given point, including its depth. This is an important step forward in terms of determining real bottom characteristics in coastal areas instead of a commonly used interpolated seabed model. The efficient application of such approaches can be assessed on the basis of research results related to sea bottom shape reconstructions [22,23]. One interesting solution in this area might also include the use of artificial neural networks for modeling sea bottom shape, as these also continually implement a surface approximation process [24–26]. It is important to be able to present depth, as this maintains surface mapping accuracy as well as the position and value of depths. In Reference [27], the authors used bathymetry data and field data as inputs for GIS (Geographical Information System), GAM (Generalized Additive Models), and kriging methods to generate a series of maps that described bottom characteristics. An interesting approach using an artificial neural network for LiDAR data was presented in Reference [28], as well as in Reference [29] where a three-layer back propagation neural network is proposed to estimate bathymetry.

We emphasize the generalization of point object sets in this study, as these comprise bathymetric geodata. In earlier work, Li [30] divided data point operations into two basic groups, those comprising transformations of individual point objects and those made on groups. Bathymetric geodata can be included within a group of points whose transformations can be divided via the following processes:

- aggregation—combining more points into one;
- regionalization—drawing a border around a group of point objects and creating a new surface;
- selective omission—the selection of objects that are more important while omitting those of lesser importance;
- simplification—the removal of objects in order to correctly present the remainder, and;
- typography—the preservation of point object dominant source symbolization while removing points [30].

The problem addressed in this study can therefore be considered to be one of selective omission because of the choice of points with the smallest depths. This process is also a simplification because of the importance of point visualization on a bathymetric chart. As available methods utilize interpolated values, it is necessary to search for new solutions. The main assumption of the proposed method of reduction is the preservation of real data, without interpolation.

2. Materials and Methods

It is important to discuss the main assumptions that underlie the development of our method. The first assumption applied here was that bathymetric geodata should be subject to reduction, obtained using a multibeam echosounder. These datasets tend to be large and contain distributed characteristics. The second basic assumption utilized here was the position of geodata collected using a GPS RTK

(Global Positioning System Real Time Kinematic) system; these data were incorporated following correct preliminary processing such that their position and depth comprised true surveying values.

2.1. Proposed Solution: Initial Assumptions

The concept of a reduction method was developed in this study as the initial research stage. We assumed that this method comprised three basic stages: the pre-division of geodata into smaller subsets (pre-processing); data clustering using artificial neural networks; and the selection of a point or points at minimum depth for a given subset.

We designated a geodata border domain such that the range depended on the coordinates of the input area. In this approach, one of the following three conditions must be met: there is only one survey point within a given domain or there are none at all; the size of the side of a divided area must be smaller than the threshold value assumed by the operator; and the actual depth difference within a tested area must be smaller than that indicated by the operator. A domain will, therefore, either not be divided into smaller areas, or it will be cut into four equal squares. This process was then repeated until the assumptions discussed above were met; sets of data comprised the outputs from these iterations, their number varied depending on the characteristics of input area, and minimum and maximum values of location and depth range were associated with bottom shape. It has to be noted that the first stage of the proposed reduction method is intended for the initial division of bathymetric data and it prepares data for clustering; it is the pre-processing of big data. Big data problems were also taken from References [31–34].

The next step was data clustering. While working on this method, except for the artificial neural network (ANN), additional methods of clustering were taken into account: K-Nearest Neighbor (KNN) and traditional hierarchical algorithms. This stage is described in Reference [21]. From the analysis, it can be concluded that the best results were obtained by the KNN and ANN methods (Kohonen). This use of a Kohonen network for bathymetric data clustering has not yet been presented in the literature and is, therefore, innovative in this context. In addition, the choice of method also considered the fact that ANN is more sensitive to the depth value than the KNN method. Bathymetric data were then clustered so that all survey points were subdivided using a Kohonen network [35–37]. In the case of the Kohonen network's learning, there was no relationship between the input data and the output of the network. The competition between neurons provides the basis for updating values assigned to the weights. It can be adopted that x is the input vector, p is the number of data samples, as follows:

$$x = \left(x_1,\ x_2,\ x_3,\ \ldots,\ x_p\right)'$$

(1)

Variable w is the weight vector and connected with the node l, as follows:

$$w_l = \left(w_{l1},\ w_{l2},\ w_{l3},\ \ldots,\ w_{lp}\right)'$$

(2)

Various samples of the training dataset were presented to the network in random order. While neurons compete with each other, the one nearest to the input sources is a winner for the input dataset. The extent of adaptation depends on the distance of the neuron from the input data. The node l is shifted by some relation of the distance between it and the training data. This relation is dependent on the learning rate. For object i, the distance between the weight vector and the input signal is estimated. After the start of the competition, the node l with the nearest distance is the winner. Then, the weights of the winner are updated using the learning rule, as follows [38]:

$$w_q^{s+1} = w_q^s(1 - \alpha^s) + X_i \alpha^s = w_q^s + \alpha^s\left(X_i - w_q^s\right)$$

(3)

where α^s is the learning rate for the sth step of training. The weight vector for the lth node in the sth step of training can be characterized as w_l^s and the input vector for the ith training sample can be

described as X_i. After several epochs, a training sample is selected, and the index q of the winning neuron is defined:

$$q = \underset{l}{arg\ min}\|w_l^s - X_i\|$$

(4)

As our focus was on the selection of network parameters for bathymetric data clustering, we selected a series of self-organizing settings for further research including a hexagonal topology and an initial neighborhood size of ten. We then applied a Euclidean distance analysis over 200 iterations via the Winner Take Most rule. All previous studies related to the use of ANN in the developed method have been published [21,39–42]. The stage related to the clustering is shown in Figure 1.
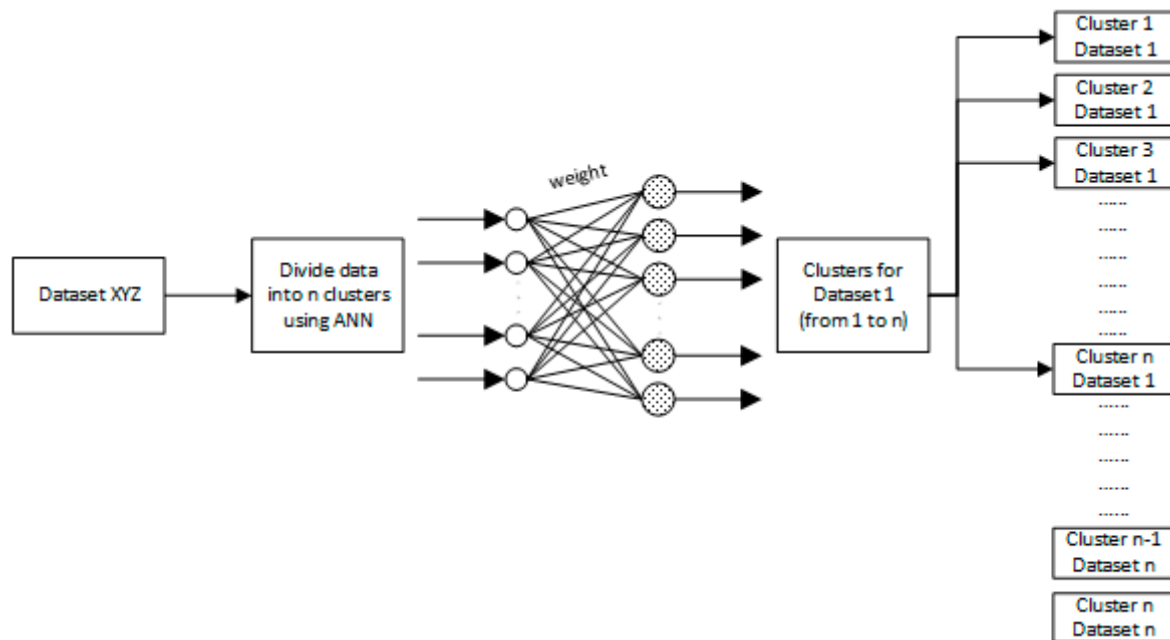

Figure 1. Clustering using an artificial neural network (ANN).

We then finally selected a series of survey points across previously designated clusters. In this context, the smallest depth (the shallowest) points will be the most important and inputs comprise collections of bathymetric geodata from individual clusters obtained in the previous phase of our method. A circle was then created around each XYZ point at a given size depending on characteristics. The radius of each circle was calculated in accordance with the authors' formula, which is presented, as follows:

$$R_i = C \times \frac{1}{Z}$$

(5)

In this expression, R_i denotes the radius at a given point, while C is a constant ($C > 0$), and Z is depth.

It is clear that points contained within a circle are of greater importance for analysis, as these indicate smaller depths (points with smaller values of Z) and will, therefore, be subject to reduction. This overall process was repeated until only districts containing the most significant objects remained and circles did not intersect. These reduced bathymetric geodata retained their real characteristics. In the next step, the initial assumptions of the proposed method were optimized on test data.

2.2. The Experiment

2.2.1. Test surfaces

Four distinct surface shape types were then determined in order to test our method that corresponded to various likely bottom surface forms. The test surfaces were built using the following mathematical functions:

- Test surface number one:

$$Z = \left(\left(\frac{3}{4}e^{-\left(\frac{(9X-2)^2+(9Y-2)^2}{4}\right)} + \frac{3}{4}e^{-\left(\frac{(9X+1)^2}{49}+\frac{(9Y+1)^2}{10}\right)} + \frac{1}{2}e^{-\left(\frac{(9X-7)^2+(9Y-7)^2}{4}\right)}\right.\right.$$
$$\left.\left. - \frac{1}{5}e^{-\left((9X-4)^2+(9Y-7)^2\right)}\right) + 0.5\right) * 0.1 \tag{6}$$

- Test surface number two:

$$Z = \left(X * e^{\left(-X^2-Y^2\right)}\right) + 1 \tag{7}$$

- Test surface number three:

$$Z = \left(4 - 2.1 * X^2 + \frac{X^4}{3}\right)X^2 + X * Y + \left(-4 + 4 * Y^2\right)Y^2 \tag{8}$$

- Test surface number four:

$$Z = \tfrac{1}{2.427}(\log(1 \quad + \left(\overline{X} + \overline{Y} + 1\right)^2 * \left(19 - 14\overline{X} + 3\overline{X}^2 - 14\overline{Y} + 6\overline{XY} + 3Y^2\right)) * (30$$
$$+ \left(2\overline{X} - 3\overline{Y}\right)^2 * \left(18 - 32\overline{X} + 12\overline{X}^2 + 48\overline{Y} - 36\overline{XY} + 27Y^2\right))) \tag{9}$$
$$-8.693)$$

where: $\overline{X} = 4X - 2$ and $\overline{Y} = 4Y - 2$.

These surfaces illustrate the possibilities of modeling real sea bottom shapes and encapsulate irregularities (Figure 2).
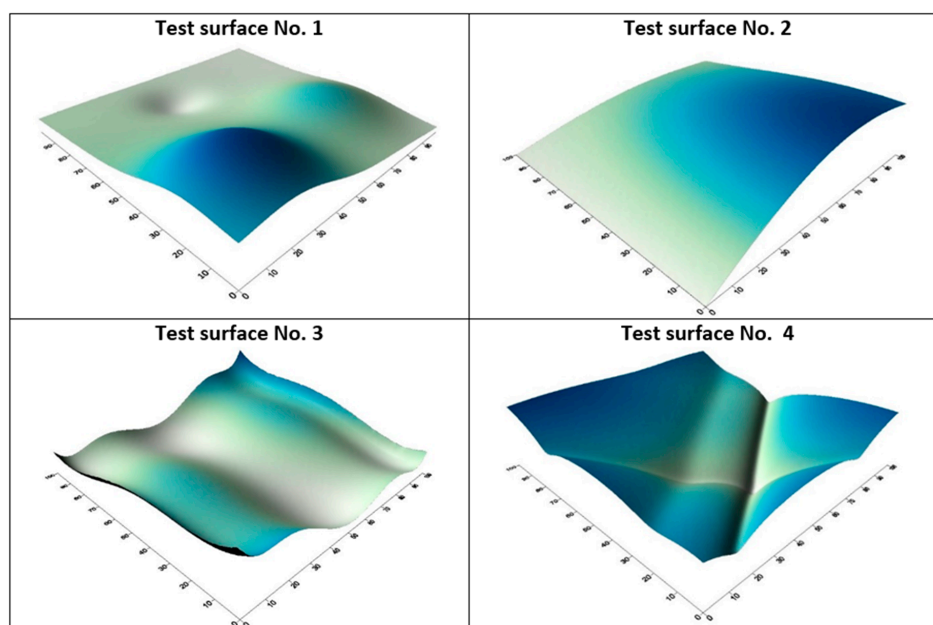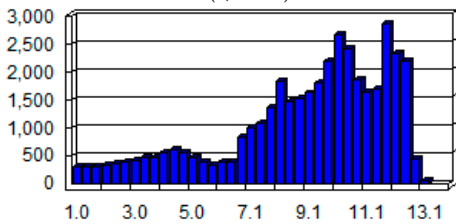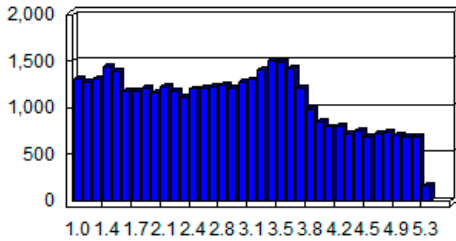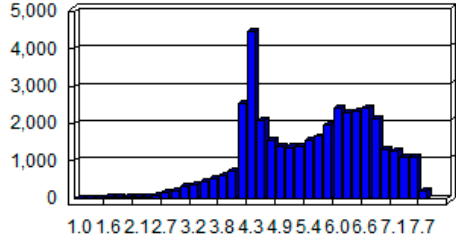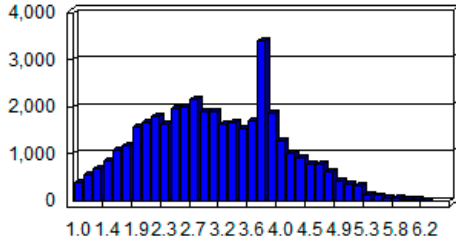

Figure 2. Alternative test surfaces considered in this analysis.

Test surface number one was characterized by irregular curvatures as it contained bottom faults, local elevations, and depressions, while test surface number two was more regular and is often seen in open water areas as it contains a gentle drop in depth associated with natural surface shapes. Test surface number three was characterized by larger changes in shape across its whole domain, while test surface number four had a shape that resembles waterway sea bottoms.

During the spatial distribution of test points on the surfaces, a simulation of scattered data was implemented, and all calculations were in the domain with sides from zero to one. Then for X and Y in this interval, Z was calculated according to the appropriate test function. In the next step, X, Y, and Z were scaled, i.e., $X \times 100$, $Y \times 100$, and, respectively, Z. The scaling was carried out in order to obtain sets of test points in the local metric system referring to the true measurements. Detailed characteristics of all test sets are presented in Table 1. The second and third columns of the table represent ranges of positions X and Y. In contrast, changes in the shape of the bottom were contained in appropriate intervals, which are specified in the fourth column. The last column also contains a histogram of the distribution of the depth value in a given set, where the *x*-axis shows the depth value and the *y*-axis the number of test points.

Table 1. Characteristics of test collections.

No.	X	Y	Z
1	$X \in (0, 99.99)$	$Y \in (0, 99.99)$	$Z \in (1, 13.18)$
2	$X \in (0, 99.99)$	$Y \in (0, 99.99)$	$Z \in (1, 5.28)$
3	$X \in (0, 100)$	$Y \in (0, 100)$	$Z \in (1, 7.71)$
4	$X \in (0, 99.99)$	$Y \in (0, 99.99)$	$Z \in (1, 6.24)$

All tests were carried out using authoring scripts written using MatLab software and they can be found in the Supplementary Materials S1, S2, S3, and S4 Scripts. Thus, in order to fulfill the assumptions of our method, geodata must be collected using a multibeam echosounder, must exhibit distributed characteristics, and comprise large datasets. A total of 40,000 test points were, therefore, generated across each test surface and the test datasets can be found in the Supplementary Materials S1, S2, S3, and S4 Files.

2.2.2. Method Optimization

We tested the new approach developed here using four previously presented collections and evaluated each across three bathymetric chart scales, 1:500, 1:1000, and 1:2000. The choice of large scales was closely related to research in coastal areas where the accuracy of the obtained results is very important. The scope of this experiment is summarized in Figure 3; our approach enabled us to distinguish three basic research stages: geodata reduction optimization, the evaluation of applied criteria, and an evaluation and comparative analysis of data reduction versus existing methods.
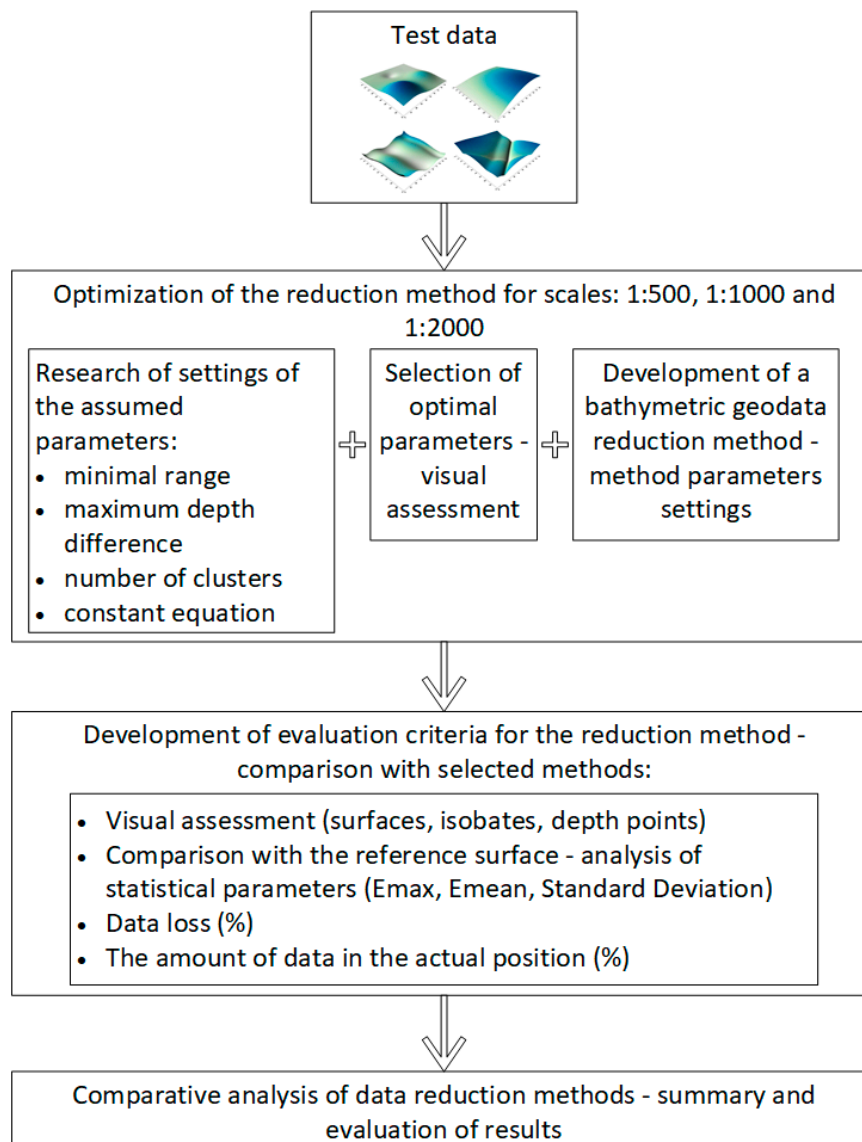

Test data

Optimization of the reduction method for scales: 1:500, 1:1000 and 1:2000

| Research of settings of the assumed parameters:
• minimal range
• maximum depth difference
• number of clusters
• constant equation | Selection of optimal parameters - visual assessment | Development of a bathymetric geodata reduction method - method parameters settings |

Development of evaluation criteria for the reduction method - comparison with selected methods:

• Visual assessment (surfaces, isobates, depth points)
• Comparison with the reference surface - analysis of statistical parameters (Emax, Emean, Standard Deviation)
• Data loss (%)
• The amount of data in the actual position (%)

Comparative analysis of data reduction methods - summary and evaluation of results

Figure 3. The research process used in this analysis.

The stage related to the optimization of the reduction method was the longest stage of the research. The four parameter settings used in our approach were optimized throughout this analysis such that:

- the threshold value in each case was related to the size of the area side under consideration—parameter value was entered on several levels from 1 to 100;
- the range between maximum and minimum depth in the study area—the minimum value was defined as 0.5 m and the maximum was equal to the maximum depth value in the tested set, changes were made every 0.5 m;
- the number of clusters—parameter value was changed from 4 to 289;
- the constant, *C*—as the minimum value was assumed 1, the maximum depending on the maximum depth value, changes were introduced every 0.5.

The list of the parameters examined for each of the sets is included in Table 2. The configuration of the four parameters was associated with carrying out different numbers of trials for each of the test datasets. In the case where in the specified intervals of one of the parameters the resulting number of points and their location did not change, other parameters of the method were introduced.

Table 2. List of the tested method's parameters.

Test Datasets	The Threshold Value	The Depth Range	The Number of Clusters	The Constant, *C*
1	6, 12, 18, 25, 50, 100 which gives 6 tested values	from 0.5 to 13.50 which gives 27 tested values	from 4 to 289 which gives 16 tested values	from 1 to 27 which gives 52 tested values
2	6, 12, 18, 25, 50, 100 which gives 6 tested values	from 0.5 to 5.50 which gives 11 tested values	from 4 to 289 which gives 16 tested values	from 1 to 13 which gives 24 tested values
3	6, 12, 18, 25, 50, 100 which gives 6 tested values	from 0.5 to 8.00 which gives 16 tested values	from 4 to 289 which gives 16 tested values	from 1 to 16 which gives 30 tested values
4	6, 12, 18, 25, 50, 100 which gives 6 tested values	from 0.5 to 6.50 which gives 13 tested values	from 4 to 289 which gives 16 tested values	from 1 to 14 which gives 26 tested values

Depending on the individual settings, a different number of output points were obtained—an example is shown in Table 3.

Table 3. An example of parameter method settings during tests—dataset number one.

Threshold Value	Depth Range (*minR*)	Number of Clusters (*K*)	Constant, *C*	Number of Points after Reduction
6	1	4	1	34,163
6	1	4	2	25,900
6	1	4	3	18,404
6	1	4	4	12,478
6	1	4	5	8341
6	1	4	5.5	6893
6	1	4	6	5756
6	1	4	6.5	4910
6	1	4	7	4225
6	1	4	7.5	3724
6	1	4	8	3364
6	1	4	8.5	3072
6	1	4	9	2852

Table 3. *Cont.*

Threshold Value	Depth Range (*minR*)	Number of Clusters (*K*)	Constant, *C*	Number of Points after Reduction
6	1	4	9.5	2678
6	1	4	10	2574
6	1	4	10.5	2495
6	1	4	11	2414
6	1	4	11.5	2367
6	1	4	12	2325
6	1	4	12.5	2296
6	1	4	13	2279

The table contains a small part of the research: a constant threshold equal to 6, *minR* equal to 1, *K* equal to 4, and *C* taking values from 1 to 13. The last column determines the number of XYZ points obtained after reduction. A total of nearly 800 tests were carried out for test set no. 1.

The number of points and the visual assessment in a given dataset were accepted as preliminary criteria. The point data that meet the criteria for visual assessments should be evenly distributed. Figure 4 shows examples of results that were taken into account in the visual assessment.
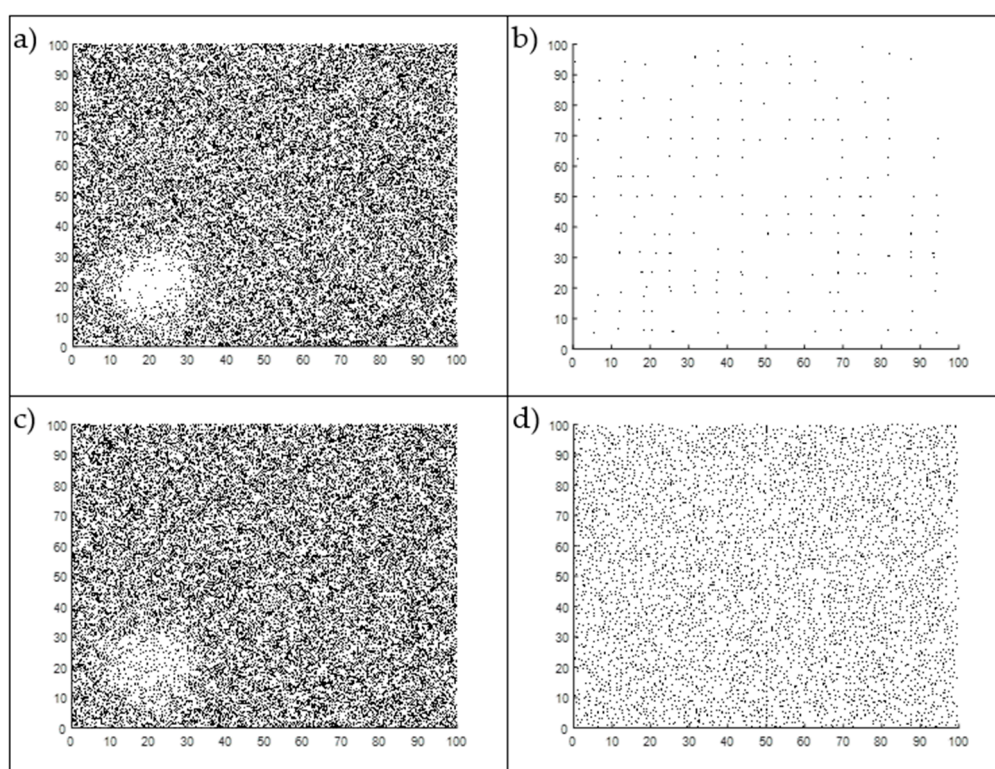
Figure 4. Examples of reduced dataset number one for the following parameters: (**a**) threshold: 18; range between max and min depth: 0.5; number of clusters: 4 and constant *C*: 1; (**b**) threshold: 18; range between max and min depth: 0.5; number of clusters: 4 and constant *C*: 25; (**c**) threshold: 18; range between max and min depth: 0.5; number of clusters: 100 and constant *C*: 1; (**d**) threshold: 18; range between max and min depth: 0.5; number of clusters: 100 and constant *C*: 25.

For the parameter settings in Figure 4a, 33,942 points were obtained. It can be seen that the constant *C* was too small. It should be noted that in the case of test set no. 1, the depth ranged from 1 m to 13.18 m. In Figure 4b, when the example was extended to 25, only 267 points were obtained. In Figure 4c, with a minimum *C* value and a number of clusters equal to 100, 34,123 points were obtained;

for Figure 4a, similar results were obtained. During the Figure 4d trial, the collection was reduced to 6409 points—this collection meets the criteria of visual assessment.

Then, after rejecting the wrong results, the optimal parameter settings were selected, and the reduction method was developed in detail, which will apply to the bathymetric geodata. This enabled us to determine the optimal settings of our built-in reduction method following the detailed analysis of selected parameters including the simultaneous visualization of reduced geodata on bathymetric reporting site plans at different scales. The steps included in our reduction method with determined parameters are summarized in Figure 5.
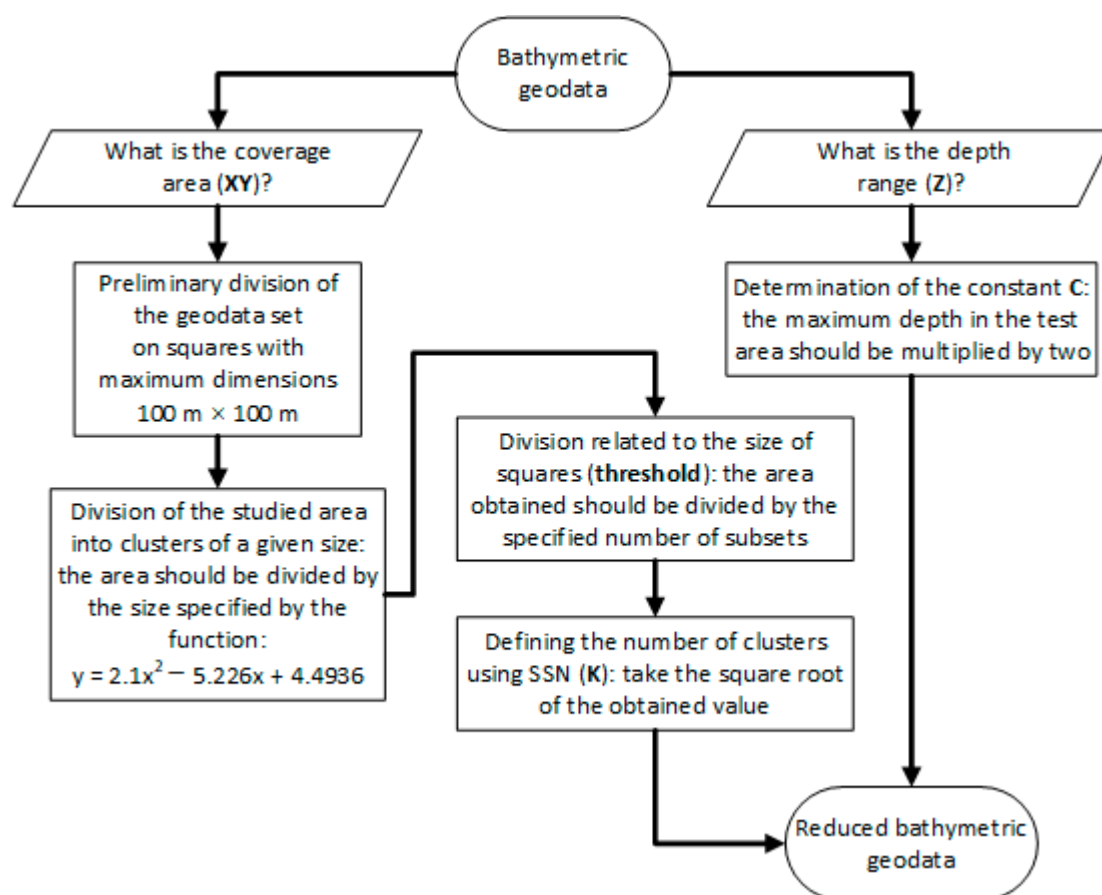

Figure 5. The bathymetric geodata reduction method used in this study.

We initially divided each coverage area into squares with the maximum dimensions of 100 m × 100 m because of the limitations inherent in processing large volumes of data. We then measured the extent of each area along *x*- and *y*-axes and assumed the smaller value to be the side of each square. Thus, if this value was greater than 100, we then divided it into smaller areas such that each received square would be subject to reduction. This means that the area covered by our data was divided depending on the desired scale of a given bathymetric chart, itself dependent on the cluster size obtained via a function with empirically selected coefficients, as follows:

$$f(x) = 2.1x^2 - 5.226x + 4.4936 \tag{10}$$

for which the graph is shown in Figure 6.
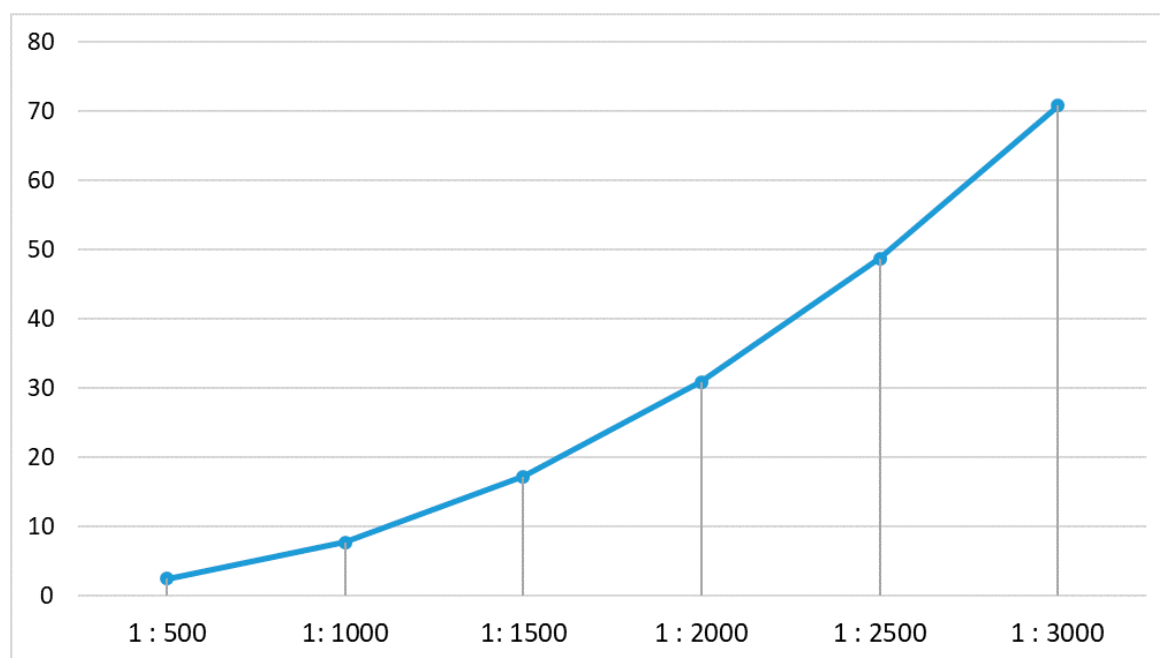
Figure 6. Graph of the cluster size function.

The horizontal axis represents the scale, the vertical axis shows the cluster size (m^2) through which the area covered by data will be divided.

This approach enabled us to select the best research results that met our visual assessment criteria and determined the functional relationship between map scale and cluster size. To this end, the value obtained using Equation (10) was then divided by a certain number of squares (Supplementary Materials Table S1); the number of subsets depended on the scope of input data and the scale of each result chart, obtained in pre-processing on the basis of each threshold value parameter for geodata areas. We then extracted the square root in each case from previously obtained values, including a value for parameter *K* that corresponds to the number of clusters obtained using an ANN. The last parameter in each case was the constant *C*, the optimal value of which corresponded to twice the maximum depth in each study area. We assumed that this constant (*C*) was an integer that could be obtained after rounding, while all applied parameter values in our bathymetric geodata reduction method depended on data scope and depth. The method was implemented in the Matlab environment.

2.3. Evaluation Criteria

We then established a series of criteria for assessing our reduction method and compared its operation with a selection of existing approaches. Using the reduction method, each test dataset was reduced for the three scales adopted in the tests, i.e., 1:500, 1:1000, and 1:2000—in total 18 sets of XYZ were created. In order to compare the approach developed in this study with alternatives that had previously been implemented, we reduced the points included in our test sets using our method as well as three-dimensional double buffering and a third technique used to visualize points that is implemented in the Caris software [43]. These two alternatives are hereafter referred to as method one and two, respectively. It is noteworthy that the reduction process implemented in method number one is similar to "rolling a ball" on a given surface at intervals defined by a vertex. Thus, once an input area was defined, it was then necessary to determine the scale of the resultant desired bathymetric plan; in this context, vertices corresponded to circle centers that each had a radius that equaled 1/100th of a predetermined scale. A surface was then obtained following buffering which was then re-buffered, this time in the reverse direction. This means that the final surface yielded via this process would be generalized and could, therefore, be used to build bathymetric charts at a given scale [43]. In contrast,

method two reduced the points that overlapped on a resultant bathymetric plan at each assumed scale. The essence of this reduction is based on the removal of deeper (or shallower, depending on indications from an operator) values when their labels overlap; this method emphasizes just visualization, and therefore, does not consider sea bottom shape [43]. We reduced each of our test datasets at different scales for three resultant bathymetric charts. All received files were saved in the *.txt format and had three attributes: X position, Y position, and Z depth. With the help of Surfer 9 software, we modeled surfaces from all received output datasets. We compared all received surfaces with reference surfaces (test surfaces).

We applied a series of evaluation criteria when assessing the operation of our bathymetric geodata reduction method, as follows:

1. Visual assessment of the distribution of obtained points.
2. Visual assessment of obtained surfaces.
3. Visual assessment of isobaths obtained from evaluated surfaces.

Visual evaluation is subjective, but with such a number of examples, you can certainly classify them. Criteria related to the visual assessment of the distribution of reduced points was made for all test collections.

4. Comparison of obtained surfaces with our model via the analysis of statistical parameters, including:

    a. the maximum difference in Z-values among surfaces,
    b. mean difference in Z-value among surfaces, and
    c. standard deviation.

We used Surfer 9 software for the calculations. We calculated the differences in the Z-values among the test sets before the reduction and the values obtained for the created surfaces. In the next step, we estimated their absolute values and basic statistics, which would be analyzed.

5. Calculation of percentage data loss after reduction.

We calculated the percentage of the number of points lost after the reduction.

6. Percentage of the amount of data the XYZ coordinates preserved.

For this purpose, we used ArcGIS software with basic spatial analysis, which is selection by location.
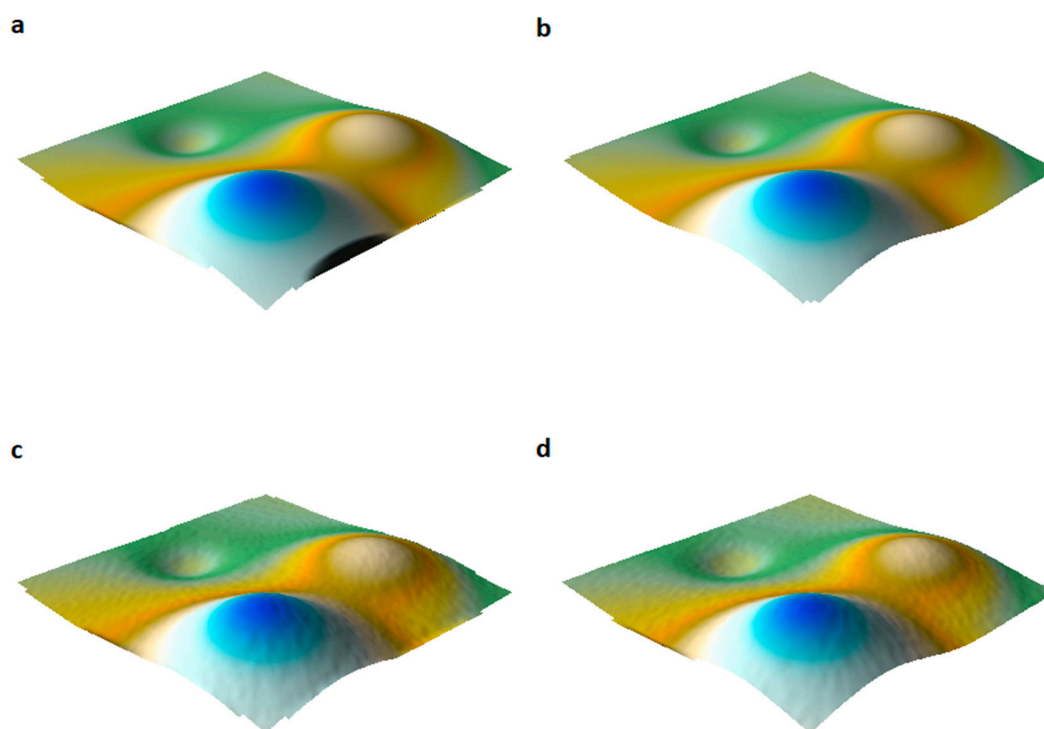
3. Results

3.1. Test Datasets

We evaluated each of our test datasets, but due to the large amount of data, we present only a description of test surface No. 1 for the purposes of this paper. The surface from which data were generated in this case comprises an irregular shape containing shallow and deep areas (test surface No. 1 in Figure 1). Initial bathymetric plans of seawater at selected scales showing depth points were developed on the basis of each received set; our visual assessment of depth points in each case was focused primarily on their spatial distribution such that they were ideally regularly distributed and their descriptions did not overlap. Subsequent to analysis of all visualizations for the test surface No. 1, it was clear that the points obtained by our reduction did meet the visual assessment criteria as they only overlapped in a few places. Therefore, we concluded that in this case, a hydrographer's interference cannot be avoided during data development using our new approach. It was also the case that points remained clearly visible when method number one was applied, as these were interpolated values that had a regular distribution. The last case we examined also met our visual assessment

criteria, although there was a tendency for values to cluster on isolines. One representative example of the results obtained at a scale of 1:1000 is presented in Figure 7.
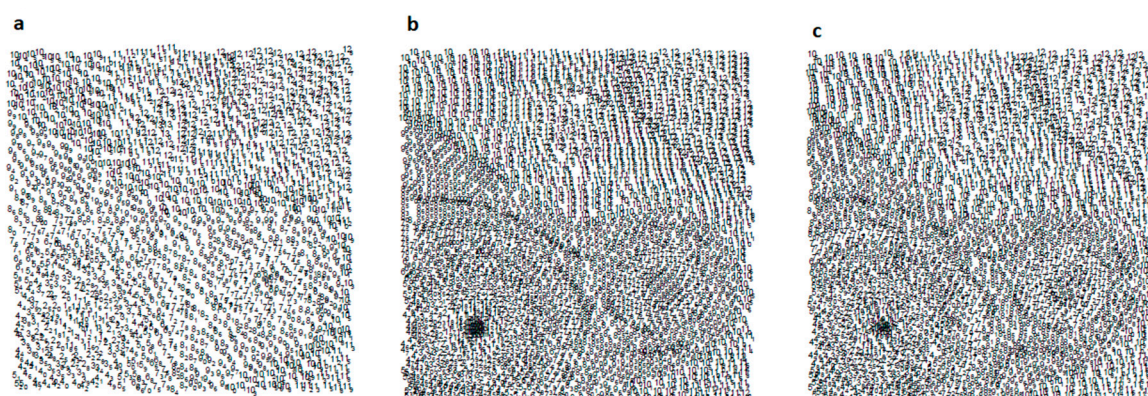
Figure 7. Depth points generated from test surface No. 1 at a scale of 1:1000. (**a**) Resolution based on our new method; (**b**) method number one; (**c**) method number two.

We also performed a visual assessment of the surface shapes generated from reduced point sets. Results showed that the surface we obtained at a scale of 1:500 from the collection reduced by our new method had almost no differences from that of the model in this case, with the only differences present at borders area. A perfect surface shape was obtained by maintaining real positions and depth values, as well as a sufficient number of points in the case of test surface No. 1. Surface roughness phenomena were seen when other methods were applied; these were associated with interpolated point values in method number one, while in method number two, these were due to an emphasis on the visualization of depth points and not on bottom shape mapping. We also illustrate the surfaces obtained at a 1:500 scale in Figure 8.

Figure 8. Surfaces obtained at a 1:500 scale using test surface No. 1. (**a**) Reference; (**b**) our method; (**c**) method number one; (**d**) method number two.

The results revealed roughness on all the surfaces obtained at scales of 1:1000 and 1:2000, likely closely related to the number of points in each set. Indeed, the surface obtained from points following reduction via our method at a scale of 1:1000 did not differ significantly from the model surface—the least of the methods studied. All tested methods generated very similar visualization results at a 1:2000 scale; however, overall, the surfaces obtained using points from our method very clearly illustrated surface shapes of shallower water but were less efficient at depth. The desirability of such an approach results from a need to enable correct bathymetry mapping in areas characterized by navigational danger.

The next step in this process comprised a visual assessment of the isobaths obtained from previously modeled surfaces. The results showed that those obtained via points generated from our reduction method almost completely coincided with surface shapes; indeed, there also seemed to be no clear difference between scales, although minimal variations were noticeable following a more detailed analysis and were indicative of very good test area mapping. The data also showed that the other methods we considered were unable to achieve such good results; the least accurately generated isobaths were obtained by our use of reduced points with method number two at a 1:500 scale. As in the case of the surfaces, this result was related to model specificity; isobaths were smoother (as is normal) at other scales; however, although their shapes were not entirely consistent with the standard. We also present a series of resultant isobaths at a 1:500 scale in Figure 9.
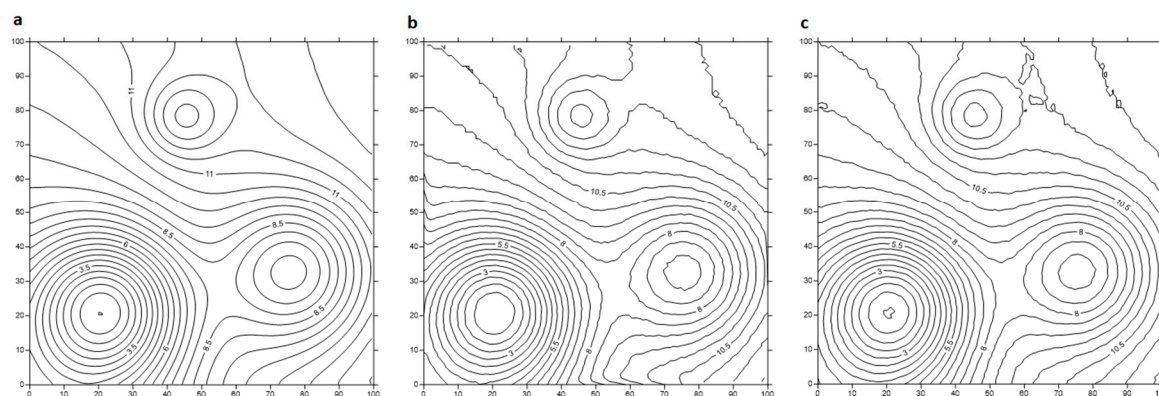
Figure 9. Isobaths at a 1:500 scale using test dataset number one. (**a**) Our method; (**b**) method number one; (**c**) method number two.

We also evaluated different reduction methods on the basis of their maximum and mean error as well as standard deviations from all tested variants (Supplementary Materials Table S2). These surfaces were then compared with the reference and all errors were calculated; in this context, maximum error can be helpful in capturing large deformations in received surface shapes. Data show that highest values of maximum error were seen in data obtained using method number one; values ranged between 67.90 cm and 72.09 cm. In contrast, the smallest errors were generated using method two and ranged between 7.03 cm and 18.42 cm in this case. Maximum errors from this approach were not much bigger, ranging between 8.38 cm and 28.32 cm; the values from each method increased in concert with scale, a natural phenomenon as the number of points decreased (Figure 10).
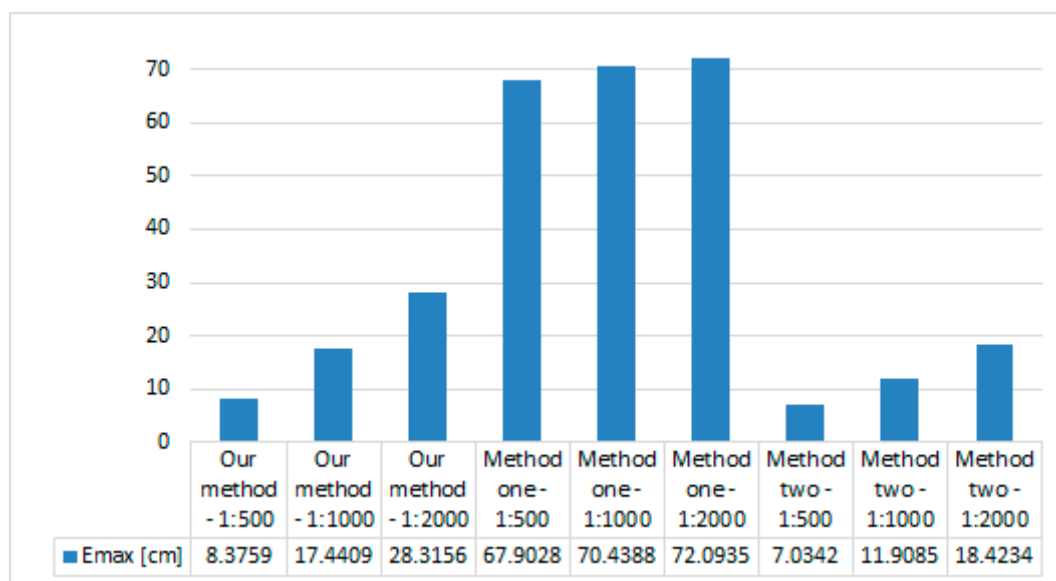
Figure 10. Maximum errors obtained for test dataset number one.

Indeed, the largest mean error at the level of 8.47 cm was obtained for points reduced using method one at a 1:2000 scale, while similar mean errors ranging between 1.32 cm and 1.82 cm were also generated using method number two. Smallest error values were obtained using our novel method; however, for 1:500 and 1:1000 scales, these errors were less than 0.50 cm and were just 1.24 cm at a 1:2000 scale (Figure 11).
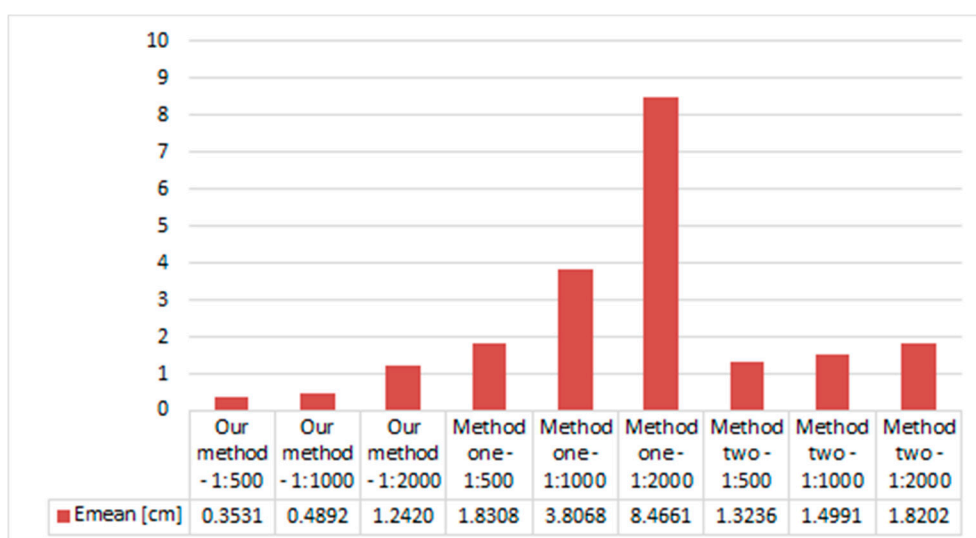

Figure 11. Mean errors obtained for test dataset number one.

These results indicate that accurate mapping of bottom shape can be achieved via our reduction method. Indeed, standard deviation results derived from test surface number one were very similar in their values and distribution to received average errors; we were, therefore, able to conclude that the best results overall can be obtained by applying our method for bathymetric data reduction, while the worst were derived from method number one. Errors in data turn out to be related to the interpolation of depth to form a square grid, as illustrated by similar research on other test datasets. The processing time during data transformation were similar and fluctuated from 15 s to 20 s for test areas (processor 2.3 GHz and RAM 16 GB).

The final evaluation criteria we applied referred to the percentage level of reduction of measurement points in each case as well as the percentage presentation of true values (Table 4).

Table 4. Comparison of data reduction levels and the number of points in each actual position in reduced datasets.

	Data Reduction Level	Number of Points in Real Positions	Data Reduction Level	Number of Points in Real Positions	Data Reduction Level	Number of Points in Real Positions	Data Reduction Level	Number of Points in Real Positions
	Test surface number one		Test surface number two		Test surface number three		Test surface number four	
Our method: 1:500	90%	100%	90%	100%	90%	100%	90%	100%
Our method: 1:1000	97%	100%	97%	100%	97%	100%	97%	100%
Our method: 1:2000	99%	100%	99%	100%	99%	100%	99%	100%
Method one: 1:500	79%	0%	70%	0%	75%	0%	73%	0%
Method one: 1:1000	95%	0%	92%	0%	94%	0%	93%	0%
Method one: 1:2000	99%	0%	98%	0%	98%	0%	98%	0%
Method two: 1:500	85%	81%	80%	81%	83%	80%	82%	82%
Method two: 1:1000	95%	81%	93%	81%	94%	80%	94%	81%
Method two: 1:2000	99%	80%	98%	81%	98%	78%	98%	81%

The analytical results of this study (Table 4) showed that our new method reduced the largest number of depth points. The data show that at a scale of 1:500, a 90% reduction can always be achieved regardless of the tested surface, while 97% and 99% were obtained at other scales. We also showed that more points were received following a reduction in the case of method one and method two; true measuring points (position and depth) were held at 100% taking into account the position and value of the depth of the reduced points when our method was applied. In contrast, the accuracy of method number one decreased following simultaneous interpolation, and therefore, depth points did not coincide with actual results. In method number two, however, just between 78% and 82% of the data overlapped despite a true position assumption in the software. It was also the case that the remainder were moved to enable a better presentation on our water area bathymetric plan; the source points for test surface one as well as those for individual scales are shown in Figure 12.
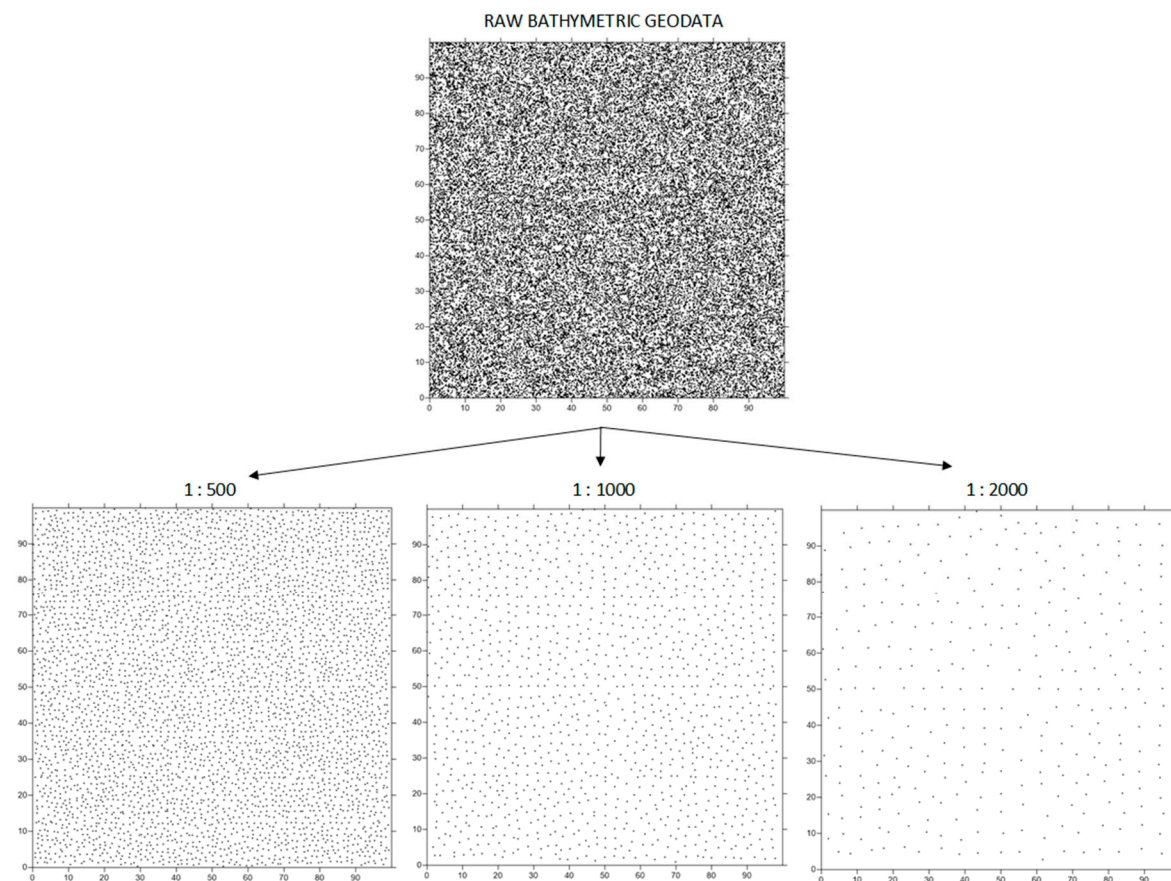
Figure 12. The points reduced in this study with our new method based on test surface one.

3.2. Real Data

In the last step of our research, we verified the new reduction method on real datasets. The assessed data were gathered by the usage of a GeoSwatch Plus multibeam echo sounder on board a Hydrograf XXI laboratory. As the main evaluation criteria, it was assumed that bathymetric data after reduction should still be in the real position and have real value of depth. The areas used in the research differed from each other in terms of the shape of the bottom and the characteristics of the occurring shore. The reduction was carried out for the three examined scales: 1:500, 1:1000, and 1:2000. According to the assumptions of the method, appropriate parameters were adopted for each scale.

The first test set comes from the Babina area. The input data included 5,864,171 bathymetric points. The minimal depth was 0.3 m and maximum depth was 9.34 m. After the reduction for the first area, three sets of reduced bathymetric geodata were obtained, which contained the following number of points:

- for scale 1:500—25,843 points XYZ with minimum depth 0.3 m;
- for scale 1: 1000—8583 points XYZ with minimum depth 0.3 m;
- for scale 1: 1000—2385 points XYZ with minimum depth 0.3 m.

For the result points, digital terrain model using triangulation with linear interpolation method was developed, which are visualized in Figure 13.
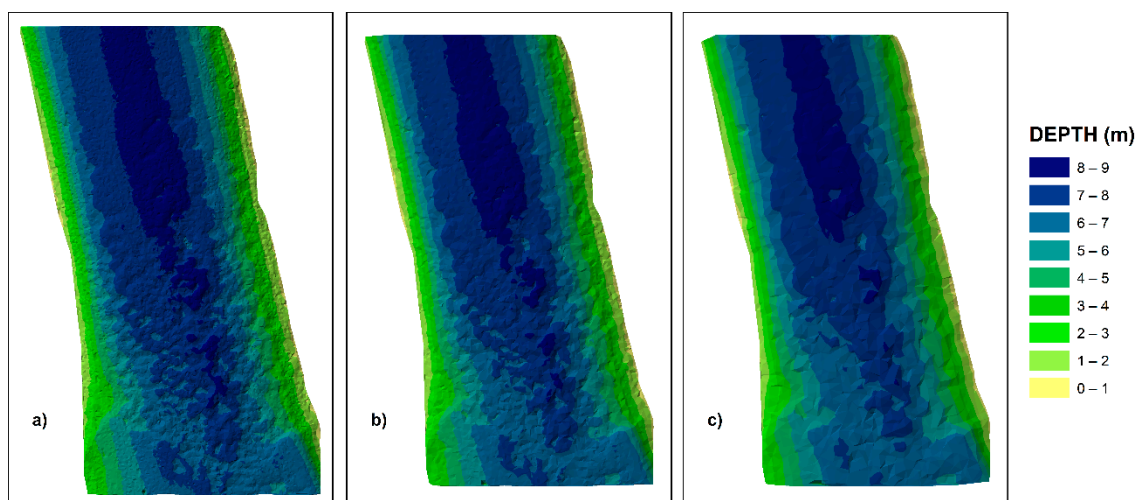
Figure 13. Surfaces from the Babina area for scale: (**a**) 1: 500, (**b**) 1:1000, and (**c**) 1:2000.

The second dataset was collected in the Debicki canal. The tested collection had 27,362,303 points and the minimal depth was 0.04 m and the maximum depth was 14.42 m. In this case, the following output datasets were obtained:

- for scale 1:500—54,346 points XYZ with minimum depth 0.3 m;
- for scale 1:1000—18,184 points XYZ with minimum depth 0.3 m;
- for scale 1:1000—7466 points XYZ with minimum depth 0.3 m.

The collections obtained during the reduction process were used to create digital terrain models, which are illustrated in Figure 14. As in the previous case, triangulation with linear interpolation method was used.
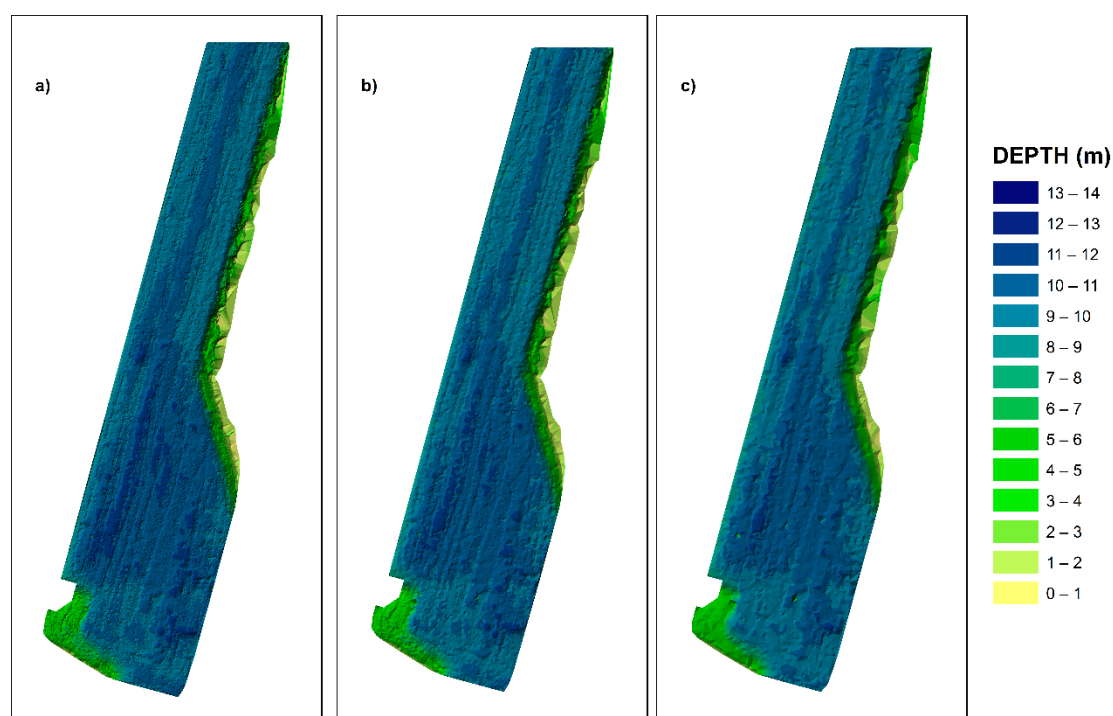
Figure 14. Surfaces from the Debicki canal for scale: (**a**) 1:500, (**b**) 1:1000, and (**c**) 1:2000.

The use of the method presented made it possible to retain 100% of the characteristics of the bottom surfaces. The geodata were commonly characterized by a lower rate of surface projection failure as real

depth was preserved without interpolation. Reduction also made it possible to minimize the dataset from which results could be accurately obtained and opened a range of future processing possibilities.

4. Conclusions

Our research thesis was related to the reduction of bathymetric geodata, an issue that has a direct relationship to research in coastal zones. Our main goal was, therefore, to develop a method to reduce geodata while maintaining real measurement values. The method proposed here to reduce geodata enabled a selection of depth points to maintain the necessary accuracy of surface mapping. We tested our new method with a range of generated datasets that exhibited generally different characteristics related to the shape of the sea bottom across special areas. In order to properly evaluate the performance of our method, we defined a series of criteria and compared our results with those obtained from other methods. A comprehensive analysis of our results enabled us to present an objective evaluation of our new method. The reduction method was tested on real data. We chose coastal areas with a natural and constructed coastline and with different depth values.

The novel reduction method developed in this study allows real depth values to be maintained at their exact measurement locations. At the same time, the parameters of this method depend on the scope of input data and the depth of the water area being tested. This means that the set of input geodata is significantly reduced depending on the scale of the bathymetric map as a result of faster analysis and subsequent processing possibilities. The data we obtained was further characterized by a smaller surface mapping error, associated with maintaining real depth values without interpolation. Our new method also processes data via a novel artificial neural network approach; the bathymetric geodata obtained as a result of this reduction process can be implemented during the creation of a map or digital seabed model. We showed that a map in which XYZ points were reduced using our novel method does possess the necessary surface mapping accuracy. The results of this analysis showed that our new method for the reduction of bathymetric geodata can be utilized for the development of hydrographic products that require real data. Our research can be the basis for methods related to the creation of large numerical geographical surface models, as well as in the processing, management, and visualization of large geodata sets that comprise measurements obtained by modern remote sensing instruments.

Supplementary Materials: S1 File: Test dataset number one, S2 File: Test dataset number two, S3 File: Test dataset number three, S4 File: Test dataset number four, S1 Script: Script used to generate test surface number one, S2 Script: Script used to generate test surface number two, S3 Script: Script used to generate test surface number three, S4 Script: Script used to generate test surface number four, S1 Table: The number of subsets depending on the data range and the bathymetric chart scale, S2 Table: The results—statistical parameters.

References

1. Brown, M.E.; Kraus, N.C. *Tips for Developing Bathymetry Grids for Coastal Modeling System Applications*; Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory: Vicksburg, MS, USA, 2007.
2. Mishra, P.; Panda, U.S.; Pradhan, U.C.; Kumar, S.; Naik, S.; Begum, M.; Ishwarya, J. Coastal Water Quality Monitoring and Modelling Off Chennai City. *Procedia Eng.* **2015**, *116*, 955–962. [CrossRef]
3. Stansby, P.K. Coastal hydrodynamics—Present and future. *J. Hydraul. Res.* **2014**, *51*, 341–350. [CrossRef]
4. Bottelier, P.; Haagmans, R.; Kinneging, N. Fast Reduction of High Density Multibeam Echosounder Data for Near Real-Time Applications. *Hydrogr. J.* **2000**, *98*, 23–28.

5. Burroughes, J.; George, K.; Abbot, V. Interpolation of hydrographic survey data. *Hydrogr. J.* **2001**, *99*, 21–23.

6. Hansen, R.E.; Callow, H.J.; Sabo, T.O.; Synnes, S.A.V. Challenges in Seafloor Imaging and Mapping with Synthetic Aperture Sonar. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3677–3687. [CrossRef]

7. Jong, C.D.; Lachapelle, G.; Skone, S.; Elema, I.A. *Hydrography*, 2nd ed.; DUP Blue Print: Delft, The Netherlands, 2010.

8. Moszynski, M.; Chybicki, A.; Kulawiak, M.; Lubniewski, Z. A novel method for archiving multibeam sonar data with emphasis on efficient record size reduction and storage. *Pol. Marit. Res.* **2013**, *20*, 77–86. [CrossRef]

9. Rezvani, M.; Hadi, S.; Abbas, A.; Alireza, A. Robust Automatic Reduction of Multibeam Bathymetric Data Based on M-estimators. *Mar. Geod.* **2015**, *38*, 327–344. [CrossRef]

10. Calder, B.R.; Mayer, L.A. Automatic processing of high-rate, high-density multibeam echosounder data. *Geochem. Geophys. Geosyst.* **2003**, *4*, 1048. [CrossRef]

11. Yang, F.; Li, J.; Han, L.; Liu, Z. The filtering and compressing of outer beams to multibeam bathymetric data. *Mar. Geophys. Res.* **2013**, *34*, 17–24. [CrossRef]

12. Kazimierski, W.; Wlodarczyk-Sielicka, M. Technology of Spatial Data Geometrical Simplification in Maritime Mobile Information System for Coastal Waters. *Pol. Marit. Res.* **2016**, *23*, 3–12. [CrossRef]

13. Specht, C.; Switalski, E.; Specht, M. Application of an Autonomous/Unmanned Survey Vessel (ASV/USV) in bathymetric measurements. *Pol. Marit. Res.* **2017**, *24*, 36–44. [CrossRef]

14. Kulawik, M.; Lubniewski, Z. Processing of LiDAR and multibeam sonar point cloud data for 3D surface and object shape reconstruction. In Proceedings of the 2016 Baltic Geodetic Congress (BGC Geomatics), Gdańsk, Poland, 2–4 June 2016. [CrossRef]

15. Blaszczak-Bak, W. New Optimum Dataset method in LiDAR processing. *Acta Geodyn. Geomater.* **2016**, *13*, 379–386. [CrossRef]

16. Blaszczak-Bak, W.; Sobieraj-Zlobinska, A.; Kowalik, M. The OptD-multi method in LiDAR processing. *Meas. Sci. Technol.* **2017**, *28*, 075009. [CrossRef]

17. Blaszczak-Bak, W.; Koppanyi, Z.; Toth, C. Reduction Method for Mobile Laser Scanning Data. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 285. [CrossRef]

18. Holland, M.; Hoggarth, A.; Nicholson, J. Hydrographic processing considerations in the "Big Data" age: An overview of technology trends in ocean and coastal surveys. *Earth Environ. Sci.* **2016**, *34*, 012016. [CrossRef]

19. International Hydrographic Organization (IHO). *Transfer Standard for Digital Hydrographic Data*, 3rd ed.; Special Publication No. 57; International Hydrographic Organization: Monte Carlo, Monaco, 2002.

20. International Hydrographic Organization (IHO). *Standards for Hydrographic Surveys*, 5th ed.; Special Publication No. 44; International Hydrographic Organization: Monte Carlo, Monaco, 2008.

21. Wlodarczyk-Sielicka, M.; Stateczny, A. Clustering bathymetric data for electronic navigational charts. *J. Navig.* **2016**, *69*, 1143–1153. [CrossRef]

22. Lenk, U.; Kruse, I. Multibeam data processing. *Hydrogr. J.* **2001**, *102*, 9–14.

23. Maleika, W.; Palczynski, M.; Frejlichowski, D. Interpolation Methods and the Accuracy of Bathymetric Seabed Models Based on Multibeam Echosounder Data. *Lect. Notes Artif. Intell.* **2012**, *7198*, 466–475.

24. Cao, J.; Cui, H.; Shi, H.; Jiao, L. Big Data: A Parallel Particle Swarm Optimization-Back-Propagation Neural Network Algorithm Based on MapReduce. *PLoS ONE* **2016**, *11*, e0157551. [CrossRef]

25. Liu, S.; Wang, L.; Liu, H.; Su, H.; Li, X.; Zheng, W. Deriving Bathymetry from Optical Images with a Localized Neural Network Algorithm. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5334–5342. [CrossRef]

26. Lubczonek, J. Hybrid neural model of the sea bottom surface, Artificial Intelligence and Soft Computing-ICAISC. *Lect. Notes Comput. Sci.* **2004**, *3070*, 1154–1160.

27. Sibaja-Cordero, J.A.; Troncoso, J.S.; Benavides-Varela, C.; Cortés, J. Distribution of shallow water soft and hard bottom seabeds in the Isla del Coco National Park, Pacific Costa Rica. *Rev. Biol. Trop.* **2012**, *60*, 53–66.

28. Kogut, T.; Niemeyer, J.; Bujakiewicz, A. Neural networks for the generation of sea bed models using airborne lidar bathymetry data. *Geod. Cartogr.* **2016**, *65*, 41–53. [CrossRef]

29. Huang, S.Y.; Liu, C.L.; Ren, H. Costal Bathymetry Estimation from Multispectral Image with Back Propagation Neural Network. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 1123–1125. [CrossRef]

30. Li, Z. *Algorithmic Foundation of Multi-Scale Spatial Representation*; CRC Press: Boca Raton, FL, USA, 2007.

31. Habib ur Rehman, M.; Chang, V.; Batool, A.; Wah, T.Y. Big data reduction framework for value creation in sustainable enterprises. *Int. J. Inf. Manag.* **2016**, *36*, 917–928. [CrossRef]

32. Habib ur Rehman, M.; Jayaraman, P.; Malik, S.; Khan, A.; Medhat Gaber, M. RedEdge: A Novel Architecture for Big Data Processing in Mobile Edge Computing Environments. *J. Sens. Actuator Netw.* **2017**, *6*, 17. [CrossRef]

33. Zhong, R.Y.; Newman, S.T.; Huang, G.Q.; Lan, S. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Comput. Ind. Eng.* **2016**, *101*, 572–591. [CrossRef]

34. Aykut, N.O.; Akpinar, B.; Aydin, O. Hydrographic data modeling methods for determining precise seafloor topography. *Comput. Geosci.* **2013**, *17*, 661–669. [CrossRef]

35. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [CrossRef]

36. Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, *78*, 1464–1480. [CrossRef]

37. Tang, X. Fuzzy clustering based self-organizing neural network for real time evaluation of wind music. *Cogn. Syst. Res.* **2018**, *52*, 359–364. [CrossRef]

38. Osowski, S. *Artificial Neural Networks for Information Processing*; Warsaw University of Technology Publishing House: Warszawa, Poland, 2000. (In Polish)

39. Wlodarczyk-Sielicka, M.; Lubczonek, J.; Stateczny, A. Comparison of Selected Clustering Algorithms of Raw Data Obtained by Interferometric Methods Using Artificial Neural Networks. In Proceedings of the 2016 17th International Radar Symposium (IRS), Krakow, Poland, 10–12 May 2016. [CrossRef]

40. Wlodarczyk-Sielicka, M. Importance of neighborhood parameters during clustering of bathymetric data using neural network. In *International Conference on Information and Software Technologies*; Dregvaite, G., Damasevicius, R., Eds.; Springer: Cham, Switzerland, 2016; pp. 441–452. [CrossRef]

41. Wlodarczyk-Sielicka, M.; Stateczny, A. Selection of SOM Parameters for the Needs of Clusterisation of Data Obtained by Interferometric Methods. In Proceedings of the 2015 16th International Radar Symposium (IRS), Dresden, Germany, 24–26 June 2015; pp. 1129–1134. [CrossRef]

42. Wlodarczyk-Sielicka, M.; Lubczonek, J. The Use of an Artificial Neural Network to Process Hydrographic Big Data during Surface Modeling. *Computers* **2019**, *8*, 26. [CrossRef]

43. *Caris, Bathy DataBASE Manager/Editor Reference Guide*, 2011.