

## COMPARISON AND ANALYSIS OF THINNING METHODS FOR MULTI-BEAM SOUNDING DATA

Wang Renjie<sup>1,\*</sup>, Li Chengpeng<sup>1</sup>

<sup>1</sup> Shandong University of Science and Technology, Qingdao 266590, China - wrjsdust@163.com

**KEY WORDS:** Multi-beam Sounding System, Underwater Topography, Thinning, Water Depth

### ABSTRACT:

Marine surveying and mapping is the basis of all marine development activities, and underwater topographic survey is one of the essential tasks of it. The multi-beam sounding system can give dozens or even hundreds of water depth values in the vertical plane perpendicular to the course at a time, and there is a lot of redundancy in these data. Efficient compression can make better use of water depth data, improve work efficiency, save system hardware resources, and facilitate rapid mapping and the construction of submarine topography model. Thinning requires an optimal balance between data accuracy and sampling density. In this paper, several commonly used thinning methods are selected and applied to the sounding data for experiments, and the application effects of different thinning methods are analyzed and compared. The results show that the mesh-based and system-based thinning methods are simple and efficient, and the results are more evenly distributed. It works well in areas with flat topography and low complexity. But in the area with large relief, the result of thinning may not take into account the topographical features, and the effect of topography representation is poor. The thinning method based on distance and elevation difference takes the elevation factor into account and has a better performance in preserving topography features. However, this method needs to search the points in a given range constantly, and it is inefficient to apply it to large amounts of data. The thinning method based on the Douglas-Peucker algorithm only considers the spatial relationship within each ping data, and the thinning result is not reasonable enough. This paper can provide reference for sounding data thinning.

### 1. INTRODUCTION

Marine surveying and mapping is the basis of all marine development activities, and underwater topographic survey is one of the essential tasks of marine surveying and mapping. With the development and support of modern science and technology, underwater topographic survey technology has been developed rapidly, and has become an important research field in marine surveying and mapping of the world's marine countries. With the development of marine surveying and mapping technology, The application of multi-beam bathymetric technique in hydrographic survey is more and more extensive, At present, it is one of that most advanced technical means of bottom topographic survey in Shanghai in the world, which can give dozens or even hundreds of water depths in plane perpendicular to the heading at a time, thus extending the surveying mode from the point and line surveying mode to the surface surveying mode (LI Haisen, 2013), and further developing to the three-dimensional sounding and automatic mapping mode. The multi-beam bathymetric technique can accurately and quickly measure the size, shape, and height of underwater targets within a certain width along the course, thus depicting the fine features of the seafloor topography and physiognomy reliably. Therefore, it has the characteristics of large measurement range, fast measurement speed, high accuracy and efficiency, digitization of records and real-time automatic mapping.

With the development of multi-beam sounding system hardware technology, the number of beams is increasing. The number of beams per sector has reached several hundred or more, and the update rate of the sector is also increasing, up to several

hundred pings per second. The data density of water depth points measured by the multi-beam sounding system can reach hundreds of points per square meter, so the data amount of TB level is usually formed. There are a lot of data redundancies in bathymetric data, which bring huge burden to the storage, organization, management, deep-level processing, and application of data, and are not conducive to the establishment of seafloor terrain model and chart production (Lurton X., 1994). Efficient compression of massive multi-beam bathymetric data can make better use of it, improve the efficiency of bathymetric processing, save system hardware resources, facilitate rapid mapping and the construction of seafloor terrain model, and create objective economic benefits. Therefore, the theory and method of multi-beam bathymetric data thinning has become an indispensable key technology in multi-beam bathymetric data processing.

### 2. RELATED WORK

#### 2.1 Principle of Multi-beam Sounding System

The multi-beam sounding system is also called multi-beam swath bathymeter. In operation, the transmitting transducer transmits a beam with a narrow open-angle along the ship's course and a wide-open angle along the vertical course at a certain frequency. Corresponding to each transmitting beam, a plurality of receiving beams with narrow and wide-open angles in the vertical direction are obtained by the receiving transducer. Hundreds of narrow beams in the vertical direction can be obtained by superimposing the transmitting beams and the receiving beams one after the other. The position and water depth of the measuring point can be calculated by using the

\* Corresponding author

beam angle and travel time of each narrow beam, and a water depth strip with a certain width can be obtained with the progress of the measuring ship (ZHOU Tian, 2005).

The multi-beam sounding system uses the transducer installed at the bottom of the ship to continuously transmit and receive a plurality of beams with a certain open-angle. The transmitted beams form a sector perpendicular to the direction of the course, and the beams are projected to the seafloor to generate beam footprints. The multi-beam system records the round-trip time of each beam and determines the measured water depth of the beam according to the sound velocity profile. At the same time, the multi-beam system also records the amplitude information of the beam, that is, the echo intensity. After each measurement, a group of time-series observations of the echo intensity is obtained on the intersection line with the seafloor. The peripheral equipment simultaneously measures and records the position information, attitude data, sound velocity profile, and other data, and transmits the data to the central processing system for data processing. Then the workstation integrates the data and generates the original multi-beam data file. Finally, the data processing software analyzes and processes the data to generate the underwater topographic map.

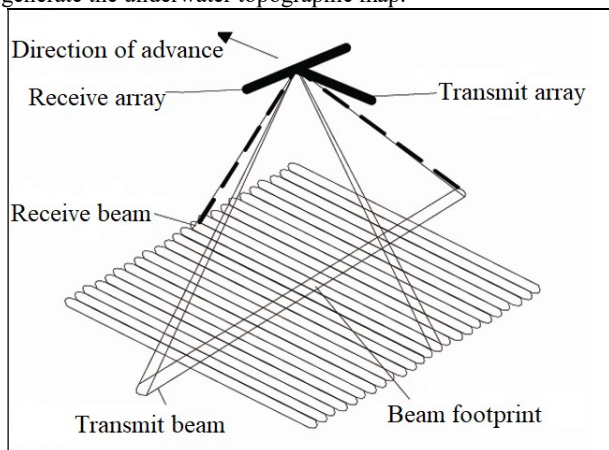


Figure 1. The multi-beam sounding system

## 2.2 Non-selective Data Thinning Method

**2.2.1 Thinning Method Based on Grid:** According to the depth data points and the measurement area, the spatial grid is established, and only one water depth point is retained in each grid (GAO Jianlin, 2001). For example, for the water depth point, the minimum depth point is more appropriate. The thinning threshold represents the size of the grid cell, which determines the final thinning degree. The larger the threshold, the higher the thinning degree, and the fewer data points will be retained. The algorithm is efficient and straightforward, but the algorithm is regular thinning, only considering the position relationship between water depth data points in each grid cell, ignoring the overall expression of terrain features. As a result, the thinning results can not take into account the topographic characteristics of the seafloor, some bathymetric feature points may be deleted, and key bathymetric data may be lost, which reduces the accuracy of bathymetric data.

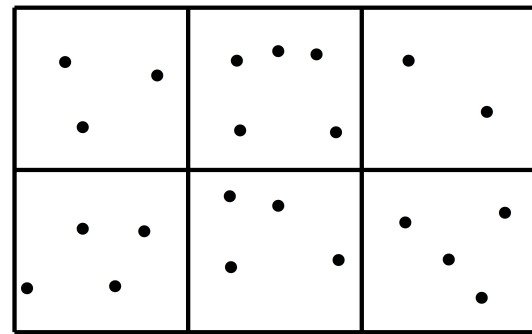


Figure 2. Grid-based thinning method

**2.2.2 Thinning Method Based on System:** For a large sample of water depth data, first determine the sampling interval  $N$ , every  $N$  sample points as a group, in  $N$  points to select a water depth sample point, and then in the next sample group to select the water depth point, until traversing the entire sample data (Calder B.R., Mayer L.A., 2003). Threshold is the sampling interval  $N$ . During the process of thinning, rules should be made for selecting water depth points for each group of water depth samples, such as selecting the minimum depth point, the median depth point of  $N$  points, or the first point and the last point of  $N$  points. This method is similar to the grid-based thinning method, which can extract water depth points quickly and simply, but the disadvantage is that it can not retain terrain features well. This method is often used to quickly view and display the massive point data, and to roughly express the undulation of the whole terrain (Peter Bottelier, 2000).

Non-selective thinning method focuses on the rapid thinning of data. For general data (not including coordinate dimensions and special feature information, etc.), this processing method can meet most of the requirements, and the speed of thinning processing is a very important factor for a simple data system. But this method is not good enough for special data (coordinate information, feature information), Although the non-selective thinning method can reduce the amount of data, it lacks the process of selecting water depth feature information when processing water depth data, and can not retain the feature information of water depth points. Therefore, the non-selective thinning algorithm is not suitable for preserving terrain feature information. Many scholars have proposed some selective thinning algorithms, which can retain key feature points and delete general data points as far as possible. The critical feature points mainly refer to the data points which can represent the terrain features.

## 2.3 Selective Data Thinning Method

**2.3.1 Thinning Method Based on Douglas-Peucker Algorithm:** In this method, the Douglas-Peucker algorithm (D.H. Douglas, T.K. Peucker, 1973), which simplifies the curve vector data, is introduced into the multi-beam sounding data thinning based on Ping, and the sounding section of each ping is regarded as a spatial curve for data thinning (XIA Wei, 2009). The classical Douglas-Peucker algorithm is to select the two ends of the line segment, and then calculate the distance between the other points of the line segment and the straight line connecting the two ends. If the maximum of the vertical distances from these points to the line is still less than a predetermined threshold, then all of these points are omitted; If

the maximum distance is greater than the threshold value, this point is reserved, and the line segment is divided into two sections by this point, and the above-mentioned operations are respectively performed on the two sections until all parts are processed and the broken lines formed by each dividing point are connected in turn, which can be regarded as the approximation of the curve.

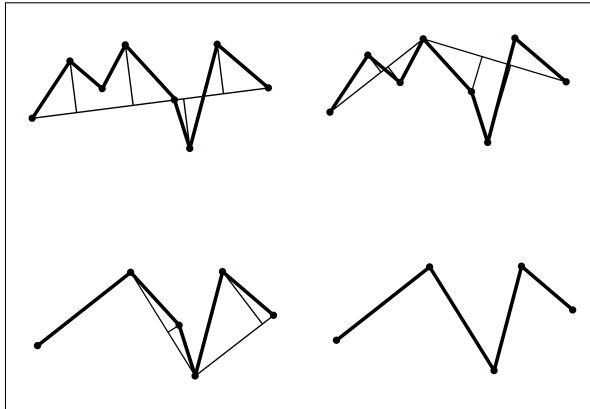


Figure 3. Douglas-Peucker algorithm

**2.3.2 Thinning Method Based on Radius and Tolerance:** Pamela's data density reduction algorithm (DDR) mainly relies on two parameters: radius and tolerance (Pamela S., 2000). The radius is a horizontal distance value (in the measurement units of the data) which determines the neighborhood around a given point. It should be selected based on the variability of the topography being considered. The points within a defined neighborhood are then compared to the center point in the thinning process. The tolerance is a vertical distance value (in the measurement units of the data) which defines redundancy. If the z-value of a point within the neighborhood differs from the z-value of the center point by less than the stated tolerance, then it is considered redundant, and one of the two points will be removed. Because the algorithm constantly searches for data points within a certain radius, if the amount of data is too large, it will greatly reduce the speed of data processing. Because of the calculation of elevation difference, this method has a good effect on preserving topographic features.

## 2.4 Spatial Data Index

The multi-beam bathymetric data is large and distributed in three-dimensional, so it is difficult to query and analyze it directly. In order to achieve efficient access to and processing of these data, and to improve the efficiency of thinning, it is necessary to find an appropriate data organization algorithm to re-establish the index of water depth data.

Kd-tree (k-dimensional search tree) is a main memory data structure which extends binary search tree to multi-dimensional data. A kd tree is a binary tree whose internal nodes have an associated attribute  $a$  and a value  $V$ , which divides the data point into two parts: the part with a value less than  $V$  and the part with a value greater than or equal to  $V$ . Because the attributes of all dimensions loop through the layers, the attributes on different layers of the tree are different.

In a typical kd tree, data points are stored in nodes, just as in a binary search tree. However, two changes were made when the idea was first introduced:

- (1) The internal node has only one attribute, a partition value of the attribute, and pointers to the left and right subtrees.
- (2) The leaf node is a block, and there are as many records in the block space as possible.

Compared with the regular grid indexing method, the kd-tree indexing method avoids the imbalance caused by the uneven distribution of spatial data, and saves the time of recursive processing. In terms of computational complexity and robustness, it has more advantages and is more suitable for static tree search. The kd tree can easily and efficiently query the local point sets in the neighborhood of points, and can be applied to high-dimensional data.

## 3. EXPERIMENTS AND RESULTS

### 3.1 Introduction to Experimental Data

The experimental data is a multi-beam sounding data with 568,496 points in a swath, and the bathymetric values are between -22.570 and -11.577. Each ping has approximately 280 points, and each ping data is located almost on the same straight line. The data of sounding are stored in order of measuring direction and sailing direction, that is, each ping is stored along the measuring direction, and then the data of the next ping is stored along the sailing direction. Therefore, the order of data storage is clear and easy to process. Through the preliminary analysis of the three-dimensional topographic map (Figure 6), we can see that the experimental data are mostly flat terrain, low complexity. A small part of the terrain undulations, high complexity.

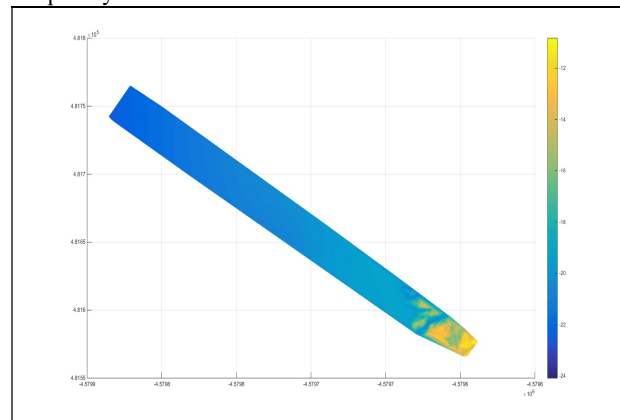


Figure 4. swath plot(X-Y view)

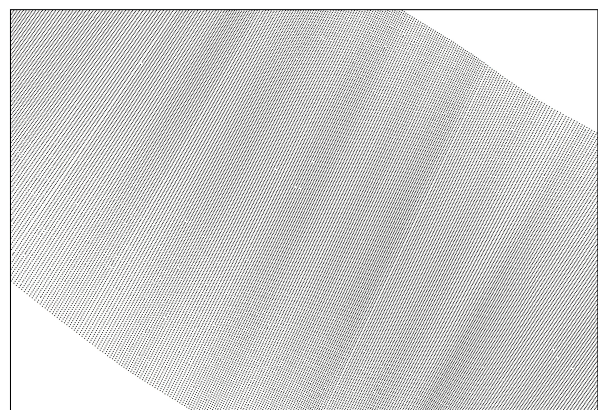


Figure 5. Local scatter plot(X-Y view)

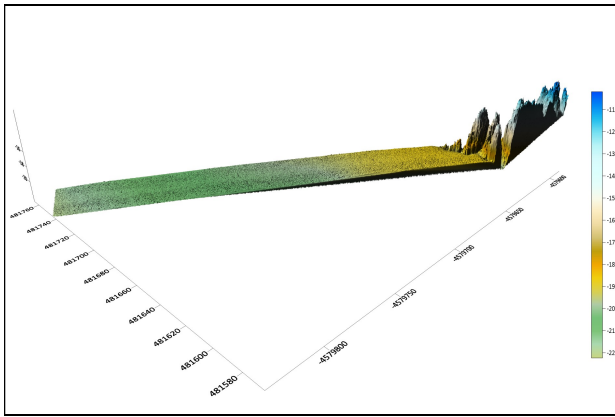


Figure 6. 3-D plot

### 3.2 Precision Evaluation

(1) Four thinning methods are used to process the multi-beam sounding data, and for group of experimental results is selected.

(2) Taking the point deleted after thinning as the test point, the elevation of the test point is interpolated by the point retained after thinning, and the interpolation method is Kriging interpolation.

(3) The mean error (MAE), mean square error (RMSE), and fitting superiority ( $R^2$ ) were calculated by the difference between the actual water depth and the interpolated water depth of the sample points (TAN Qulin, XU Xiao, 2014).

$$MAE = \frac{1}{n} \sum_{i=1}^n (|Z_i - z_i|) \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_i - z_i)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Z_i - z_i)^2}{\sum_{i=1}^n (Z_i - \bar{z})^2} \quad (3)$$

where  $n$  = number of points  
 $Z_i$  = interpolation of point  $i$   
 $z_i$  = Measured value of point  $i$   
 $\bar{z}$  = bathymetric mean

### 3.3 Experimental Results

| Method | Reserved points | Thinning rate | MAE(m) | RMSE   | $R^2$  |
|--------|-----------------|---------------|--------|--------|--------|
| Grid   | 28771           | 5.06%         | 0.0851 | 0.2052 | 0.9925 |
| System | 29921           | 5.26%         | 0.0410 | 0.1155 | 0.9976 |
| DP     | 27730           | 4.88%         | 0.0539 | 0.1198 | 0.9972 |
| DDR    | 27893           | 4.91%         | 0.0597 | 0.1349 | 0.9965 |

Table 1. Experimental results of the first group

| Method | Reserved points | Thinning rate | MAE(m) | RMSE   | $R^2$  |
|--------|-----------------|---------------|--------|--------|--------|
| Grid   | 67409           | 11.86%        | 0.0520 | 0.1259 | 0.9972 |
| System | 63167           | 11.11%        | 0.0385 | 0.1109 | 0.9978 |
| D-P    | 66459           | 11.69%        | 0.0351 | 0.0983 | 0.9981 |
| DDR    | 64103           | 11.28%        | 0.0379 | 0.1020 | 0.9979 |

Table 2. Experimental results of the second group

| Method | Reserved points | Thinning rate | MAE(m) | RMSE   | $R^2$  |
|--------|-----------------|---------------|--------|--------|--------|
| Grid   | 94060           | 16.55%        | 0.0472 | 0.1190 | 0.9975 |
| System | 94750           | 16.67%        | 0.0380 | 0.1097 | 0.9978 |
| DP     | 94504           | 16.62%        | 0.0326 | 0.0951 | 0.9982 |
| DDR    | 92637           | 16.30%        | 0.0342 | 0.0978 | 0.9980 |

Table 3. Experimental results of the third group

| Method \ Thinning Rate | 5%      | 11%     | 16%     |
|------------------------|---------|---------|---------|
|                        | Grid    | 3.43    | 3.69    |
| System                 | 1.02    | 1.03    | 1.05    |
| DP                     | 585.18  | 542.61  | 462.78  |
| DDR                    | 1606.55 | 1425.83 | 1205.42 |

Table 4. Experimental time (second)

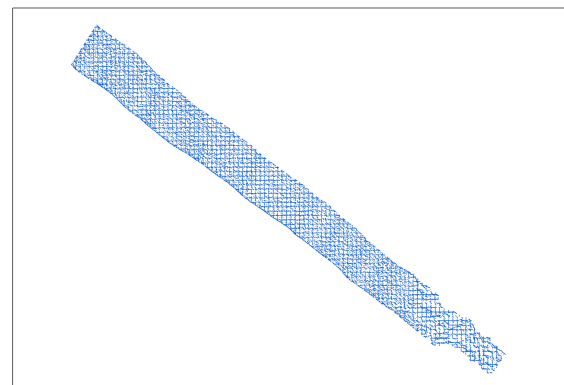


Figure 7. Grid-based thinning method of 5% thinning rate

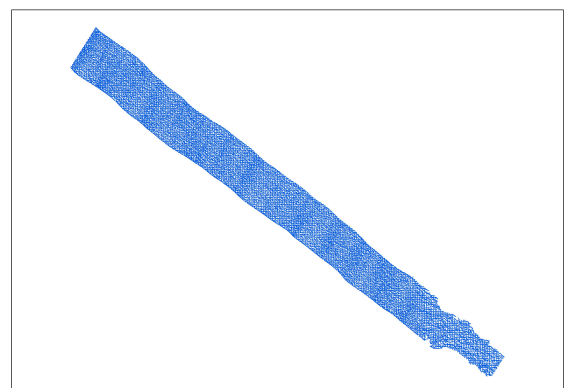


Figure 8. Grid-based thinning method of 11% thinning rate

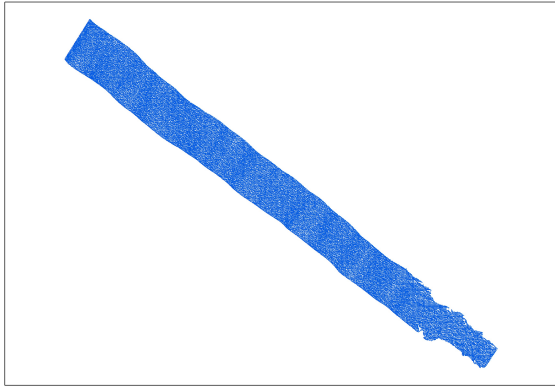


Figure 9. Grid-based thinning method of 16% thinning rate



Figure 13. D-P thinning method of 5% thinning rate

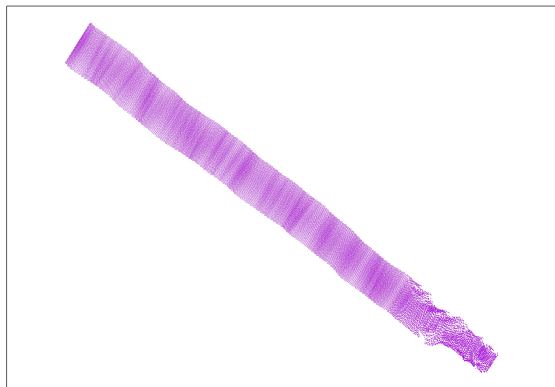


Figure 10. System-based thinning method of 5% thinning rate



Figure 14. D-P thinning method of 11% thinning rate

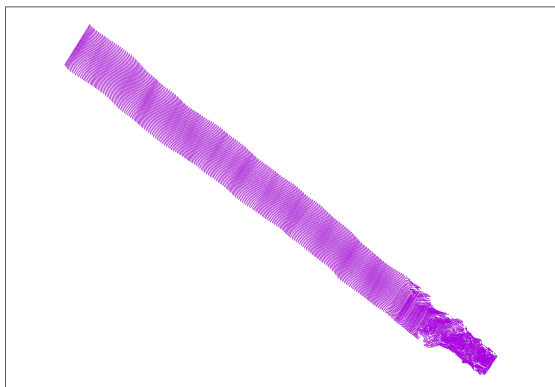


Figure 11. System-based thinning method of 5% thinning rate



Figure 15. D-P thinning method of 16% thinning rate

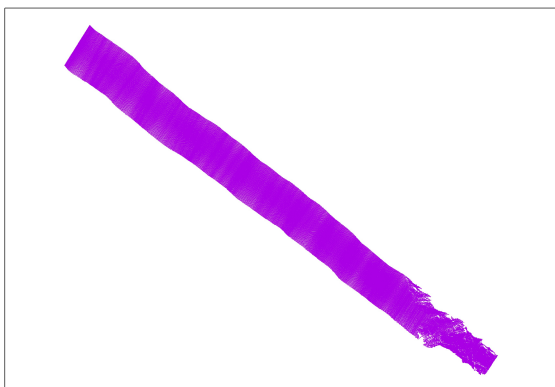


Figure 12. System-based thinning method of 5% thinning rate

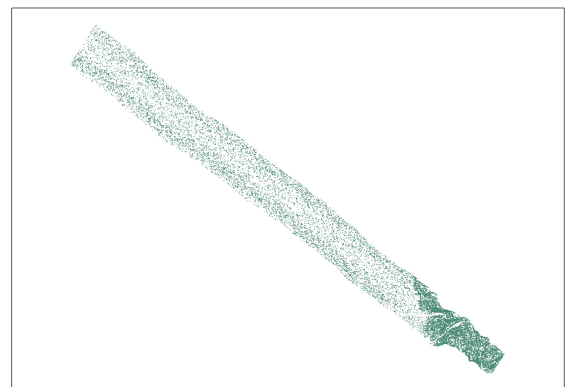


Figure 16. DDR thinning method of 5% thinning rate

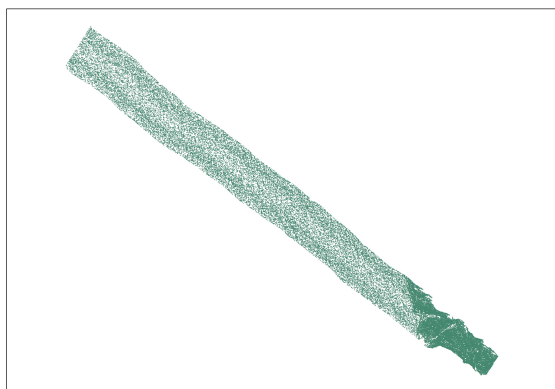


Figure 17. DDR thinning method of 11% thinning rate

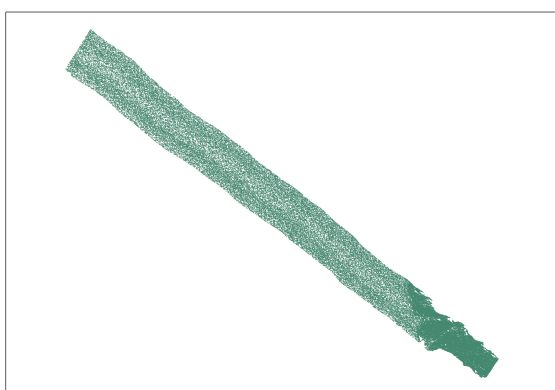


Figure 18. DDR thinning method of 16% thinning rate

#### 4. CONCLUSION

##### 4.1 Analysis of Experimental Results

The experimental data contains more than half a million points, Comparing the results of the four groups of thinning rates of 5%, 11%, 16% and 21%, it can be seen that the data accuracy after thinning is still very high at the lowest thinning rate of 5% (Table 1-3). This shows that only a small number of bathymetric points are needed to express the overall seafloor topographic characteristics and it is necessary to thin the multi-beam bathymetric data (CAO Hongbo, 2010).

Comparing the four groups of experiments (Table 1-4), we can see that. The selective thinning method is more accurate and effective than the non-selective thinning method, but the treatment efficiency is very low. The advantage of the non-selective thinning method is that the processing speed is very fast, but the precision is a little low.

The result of thinning is accidental. In the experimental group with 5% thinning rate, the system-based method has the highest precision, while in the other groups, the precision of this method is not so outstanding.

Considering the accuracy and efficiency, the D-P method is the best of the four methods. But its thinning result retains all the points on both sides of the swath, it seems very unreasonable (Figure 13-15).

##### 4.2 Summary of Thinning Methods

The Grid-based thinning method and system-based thinning method have the highest efficiency, followed by Douglas-

Peucker method, and the method based on radius and tolerance has the lowest efficiency (Table 4).

The Grid-based and system-based thinning method can estimate the degree of thinning by threshold, and the grid-based method can compress all data to the same degree, so the result of thinning is relatively uniform, but it can not well preserve topography features and remove redundant information in areas with large topographic fluctuations. Because the system-based thinning method is equal-interval extraction, the results after thinning look very regular, but the reasonable degree is lower than the grid-based thinning method (Figure 7-12).

The Douglas-Peucker-based thinning method preserves the start and end points of each ping, and then all points on either side of the entire strip are retained, so the result seems unreasonable (Figure 13-15). The D-P algorithm obviously considers the topographic features, for example, the former part of the thinning results are more sparse than the latter part, because the former part of the topographic change is more gentle, and the latter part change is more intense. However, it is not reasonable to treat the three-dimensional water depth information as a spatial curve in a small local area. and only the data of each ping are processed separately without considering the adjacent ping and the overall topographic features, some characteristic information will be lost. The higher the threshold value is, the higher the degree of thinning is.

The method based on radius and tolerance(DDR) is computationally expensive and has the lowest efficiency among the four methods (Table 4). Repeated search and judgment based on radius and height difference reduces the speed, but improves the accuracy (Table 1-3). Compared with the results based on D-P algorithm, this method takes into account the overall topographic characteristics and is more reasonable (Figure 16-18). But there are two thresholds, so it is difficult to estimate the thinning rate directly.

#### REFERENCES

- Calder B.R., Mayer L.A., Automatic processing of high-rate, high-density multibeam echosounder data. *Geochemistry Geophysics Geosystems*, 2003, 4(6).
- CAO Hongbo, ZHANG Lihua, ZHU Muhua, Comparison and Analysis of Thinning Methods for Mass Data Acquired by Multibeam Sounders. *Hydrographic Surveying and Charting*, 2010, 30(5):81-83.
- D.H. Douglas, T.K. Peucker, Algorithms for the reduction of the number of points required to represent a line or its caricature. *Classics in Cartography*. 1973.
- GAO Jianlin, Two-dimensional and three-dimensional surface image processing algorithms and their applications. New Jersey Institute of Technology, May, 2001.
- LI Haisen, ZHOU Tian, XU Chao, New developments on the technology of multi-beam bathymetric sonar. *Technical Acoustics*, 2013, 32(2):73-80.
- Lurton X., Dugelay S., Augustin J.M., Analysis of multibeam echo-sounder signals from the deep seafloor. *Oceans. IEEE*, 1994.

Pamela S., A data density reduction algorithm for Post-processed airborne lidar bathymetric survey data. Florida: University of Florida, 2000.

Peter Bottelier, Fast Reduction of High Density Multibeam Echosounder Data for Real Time Application. The Hydrographic Review, 2000.

TAN Qulin, XU Xiao, WANG Haoyu, HU Jiping. Accuracy Assessment of DEM Interpolation Algorithms in Different Landform Regions. Journal of Basic Science and Engineering. 2014(1).

XIA Wei, HUANG Motao, LIU Yanchun, Application of Douglas-Peucker method in multibeam echo sounding system data thinning. Science of Surveying and Mapping, 2009, 34(3):159-160.

ZHOU Tian, Research on super wide coverage and high resolution seafloor bathymetry and physiognomy detection. Harbin:Harbin Engineering University, 2005. 2-9.