

# **THE IMPLEMENTATION OF POST-PROCESSING DATA THINNING FOR MULTIBEAM ECHO SOUNDING DATA**

**Mohd Razali Mahmud**

Institute for Geospatial Science and Technology  
Faculty of Geoinformation Science and Engineering  
Universiti Teknologi Malaysia  
81310 UTM Skudai, Johor  
Malaysia  
Tel/Fax: 607-5530827  
Fax: 607-5566163  
E-mail: razali@fkg.utm.my

**Othman Mohd Yusof**

Department of Surveying Science and Geomatics  
Faculty of Architecture, Planning and Surveying  
Universiti Teknologi MARA  
40450 Shah Alam, Selangor  
Malaysia  
E-mail: omy2006@yahoo.co.uk

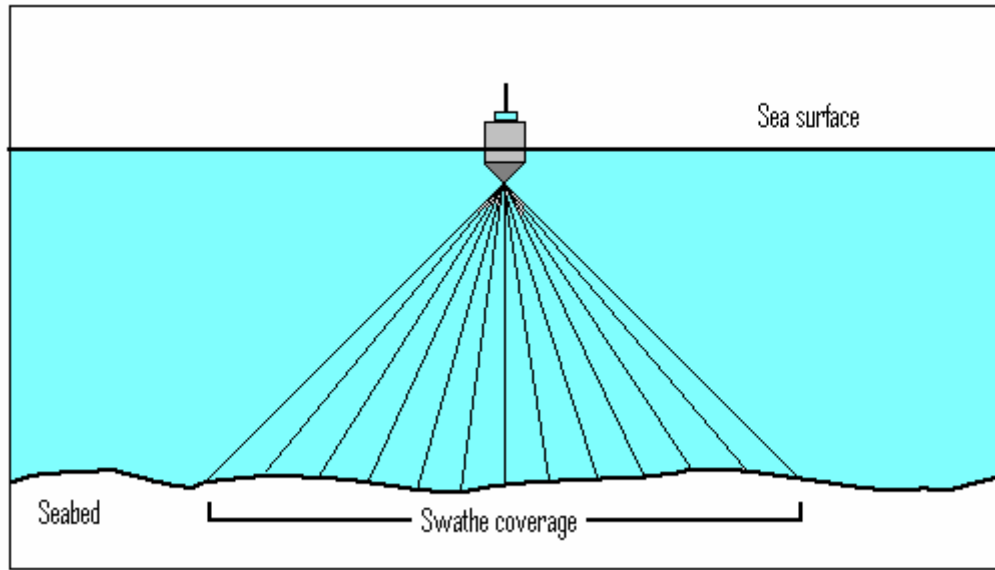
## **Abstract**

Multibeam echosounder (MBES) is considered the best innovation in depth measuring technique if full seabed coverage is the foremost concern. Multiple beams are generated through MBES transducer in a fan-shaped swathe across vessel track within fraction of second. From few to hundred of beams in every ping can be transmitted into different sections or angles from nadir beam direction then translated into multiple depth measurements. The rate of ping is subjected to survey area where the shallower the area is, the higher ping rates is expected compare to deeper area. Thus higher MBES data density is produced in shallow water area. In this shallow zone, over sampled or redundancy of MBES data are common. These over sampled data are too dense to be displayed and in consequence to represent the survey area on bathymetric plans. Another factor is these over sampled data make Digital Terrain Modeling (DTM) and contouring more intense especially on the computer processing software. Therefore these dataset should be reduced in term of it size. The process to reduce the size is called data thinning. Data thinning algorithms should be capable in handling high volume of MBES data. In the process to reduce the dataset, one must bear in mind that the process would not jeopardize the integrity and accuracy of final product. To adhere on these requirements, the resolution of the MBES used during the survey and the smallest expected detail to be mapped must be taken into considerations. This paper elaborates on the development of data thinning programs using Microsoft Visual Basic Version 6. Various algorithms namely Douglas Peucker, Single Swathe Reducer and Skip N Points are referred. Comparisons on the final results based on these three algorithms will be discussed in order to decide which algorithm is most favorable to be used in post-processing data thinning for cleaned MBES dataset. Finally the paper summarizes some of the distinguishing features of this data thinning approach.

**Keywords:** MBES, data thinning, swathe, across-track

## 1.0 Introduction

One of the most significant achievements when using MBES is increasing data rates and densities. MBES has improved survey technique with the ability to provide a high data resolution and full bottom coverage to determine bathymetry of seabed. A brief noise impulse known as a "ping" is transmitted in a fan-shaped swathe in across-track survey vessel direction via MBES transducer. This acoustic pulse is wide in across-track direction and narrow in the perpendicular along-track direction. Within each ping, hundred of beams are transmitted and propagated through water column in different angles measured from nadir beam. The two-way travel time for each beam's propagation from it is transmitted, reflected by seabed, bounced back and then received by the transducer is measured and translated into sounding depth. As the ship moves, it sweeps along the sea bottom, resulting in multiple simultaneous depth measurements across the ship track. The swathe coverage is commonly assumed as four-times as wide as the water depth.



**Fig. 1.1:** Multiple beams are transmitted in a fan-shaped swathe across-track vessel direction

For each beam received, a two-way travel time ( $t$ ) and a beam angle ( $\phi$ ) are recorded. These parameters are then converted into depths ( $d$ ) and across-track position ( $y$ ) of the soundings, assuming that errors and vessel motion are neglected, using the following formula:

$$d = \frac{1}{2} v t \cos \phi \quad (1.1)$$

$$y = \frac{1}{2} v t \sin \phi \quad (1.2)$$

where

$d$  = depth

$y$  = across-track position

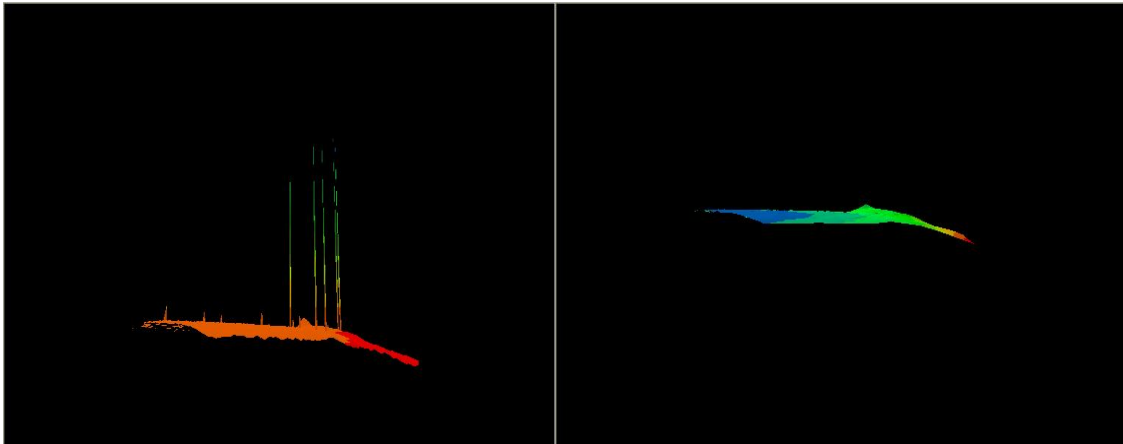
$t$  = two-way travel time

$\phi$  = beam angle

In swathe system, the above equations are used to determine series of depths and beam positions, taking into account the beam angle in each swathe. The footprint of MBES system increases, depended on the factors of the beam angle ( $\phi$ ), the beam width and the depth of the water column. The bigger the factors means the bigger the footprint will becomes. The smallest detectable object on the seabed totally depends on these parameters; the size of the footprint, the angle of incidence and the slope of the seabed. MBES resolution is further improved to the range of between 2.5 cm to 5 cm.

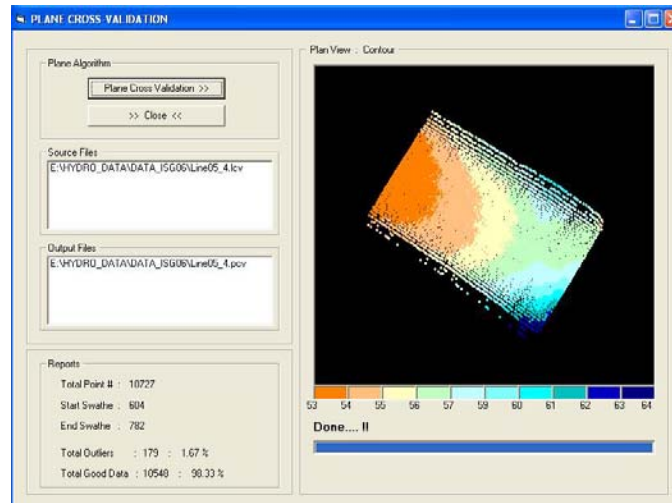
Another ping will be generated once the previous transmitted beams are received at the transducer. Therefore the rate of pings is variables subject to depth of survey area. In shallow area, a faster ping rate is expected compared to deeper area because less two-way travel time is required for shallow depth, thus more data will be collected. Sounding data collected at sea can easily be more that  $10^6$  depth measurements per hour leading to over  $10^7$  per survey (Capena *et al.*, 1999). In shallow water area, the case is more critical where high density of data is produced in a more spatially dense and more uniformly distributed over the survey area.

## 2.0 Data Thinning



**Figure 2.1:** Comparison between dataset with outliers (left box) and cleaned dataset (right box) over the same area

Data thinning will only deal with cleaned MBES dataset from outliers to give meaningful results for final contouring stage. Figure 2.1 clearly shows the significant differences between raw MBES dataset and cleaned MBES dataset. The raw dataset should be filtered by various outlier filters prior to this event. Figure 2.2 shows the cleaned MBES dataset after being screen for outliers. The same dataset was reduced by different data thinning algorithms and a comparative analysis can be performed.



**Figure 2.2:** Cleaned MBES dataset

The ideal concept of data thinning in dataset size is reduced intelligently. Those points which do not affect or marginally effect on the details of DTM have been discarded (Grounds, 2005). Where the detail is needed, points are retained and points which do not affect the model are discarded. The main objective is to decrease point density and at the same time maintain the accuracy of the Digital Terrain Modeling (DTM) and contouring process.

In this study, only a triangulated irregular network (TIN)-based DTM is considered in programming. A TIN consists of a set of non-equilateral triangles with the data points at the corner points, which have minimum length based on vertices. On the other hand, Gridded DTM is commonly used in a more traditional way of decreasing the amount of data points in a multibeam dataset (Bottelier *et al.*, 2000). This method would create new data points on a regular horizontal grid interval that represents a certain area of the original data set. The effect is downgraded its resolution in DTMs. If the depth is not at the intersection of grid lines, then the system will estimate the value based on the sample depths for that point (Brennan, 2006). Users are allowed to choose either average, shallowest, deepest, median or statistical mode depth value of the sample (Coastal Oceanographics, Inc., 2003) or it can be the nearest point to the grid intersection point considering it carries more weight. The discrepancy between volumes computation of a gridded DTM compared to that of an original TIN-based DTM is large. As concluded by Bottelier *et al.* (2000), the larger the grid interval, the less detail remains in the resulting DTM and in some test where rough terrain occurs, these volume differences can be as much as 10%.

In view of the drawbacks to data thinning by the grid method, an alternative data thinning technique, called line algorithm was utilised. In this line algorithm data thinning, the line generalization concept was adopted. Each multibeam echosounder swathe is actually a profile consisting of sounding points. Z coordinates (depths) of the points in each profile are a function of the horizontal distances to the first point in the profile. The generalization concept was applied in swathe-to-swathe basis meaning that the generalization is carried out on an individual single swathe profile at one time.

Three methods of line generalization were introduced in the data thinning procedures, namely as:

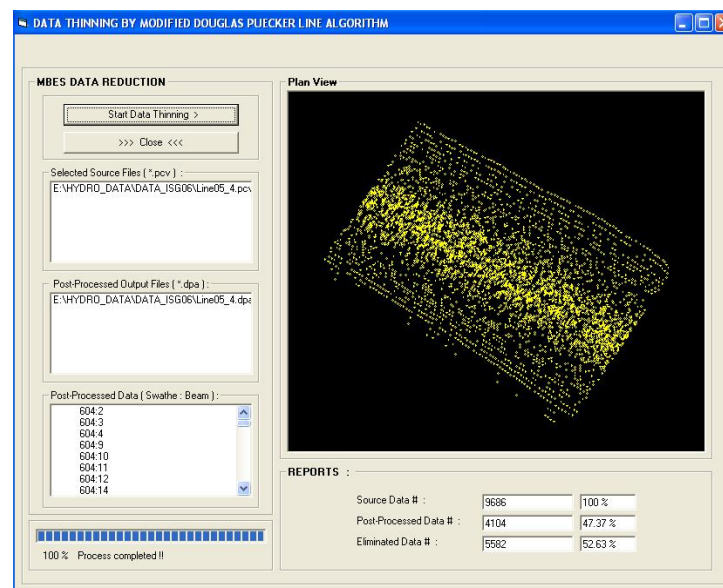
- a. Douglas Peucker Line algorithm
- b. Single Swathe Reducer algorithm
- c. Skip N Points algorithm

## 2.1 Douglas Peucker Line Algorithm

The line algorithm starts by taking the two outermost points of each profile called anchor points, so that it forms a straight line defined by a line equation.

$$y = mx + c \quad (2.1)$$

Comparison on the height differences between the measured depths and the straight line is conducted for intermediate points in between those two anchor points. The maximum absolute value of these height differences is determined and tested against the pre-defined target accuracy normally a MBES resolution value is accepted. In case of this maximum value exceeds the target limits, the point becomes a new end point. Next, two new lines will be formed using the end points (anchor points) and the maximum value point. This iterative procedure is repeated until all height differences are below the target limit. The final result is the XYZ coordinates of the end points of the line segments.



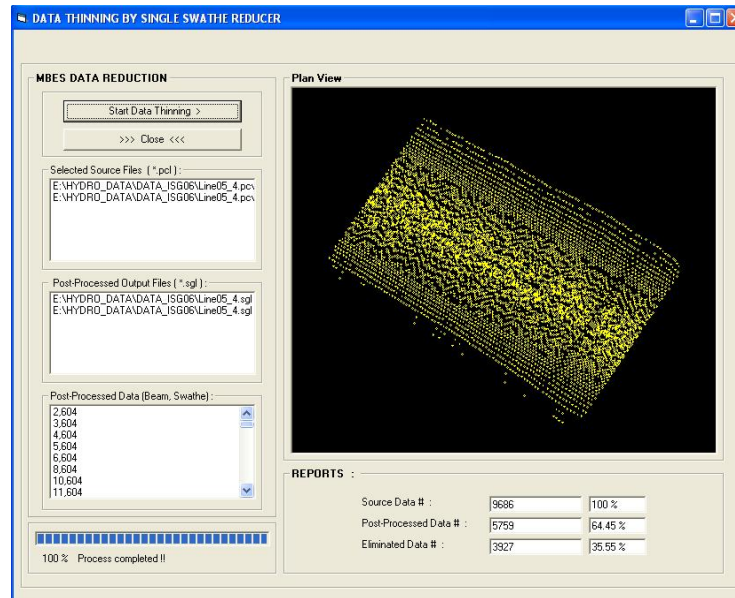
**Figure 2.3:** Results based on Douglas Peucker Line algorithm

## 2.2 Single Swathe Reducer Algorithm

In this Single Swathe Reducer algorithm, two pre-set values are tested against all MBES swathe profile. The first pre-set value is difference in height between the first anchor point and the second anchor point. The test limit for the height difference is normally referred to a resolution of MBES. The first anchor point is the first point in the swathe profile and the second anchor point is the second consecutive point in the swathe profile. If the height difference is within the limit then the second anchor point is shifted to the third consecutive point in the swathe. The test

is to be performed again. If the height difference exceeding the limit given then the second consecutive point becomes a new node as the first anchor point. The second anchor point will be shifted to the third consecutive point and the testing is carried out for its height difference. This chain reaction process is continued until to the end point of the swathe.

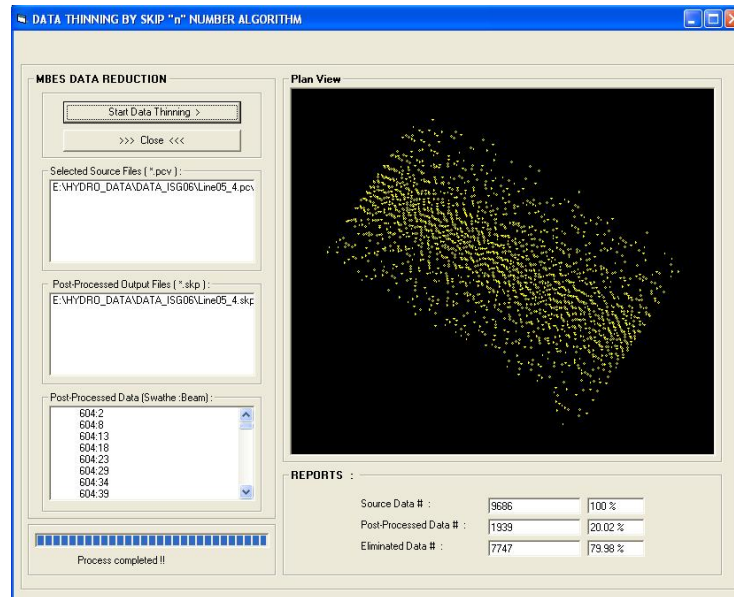
The second pre-set value is distance between the first and the second anchor points, limited to user specified range. If the height difference between the first and the second anchor points is still within the tolerance but the range limit is exceeded then the point will becomes a new node. This will ensure that there will always be adequate points for accurate DTM in the next process.



**Figure 2.4:** Results based on Single Swathe Reducer algorithm

### 2.3 Skip N Points

In the Skip N Points algorithm, the procedure is very straightforward by simply discarding every N points throughout swathe profile, for example five points are discarded and the next sixth point is retained. These routines are very simple that do not accounts for the topological relationship with the neighbouring points in the sample (McMaster, 2004). The routines are repeated until the end point in the swathe. Although it looks that the routines do not retained points where the detail is needed, by nature due to redundancy and over sampling in MBES dataset, the effect is very minimal especially for smooth and flat seabed profiles.



**Figure 2.5:** Results based on Skip N Points algorithm

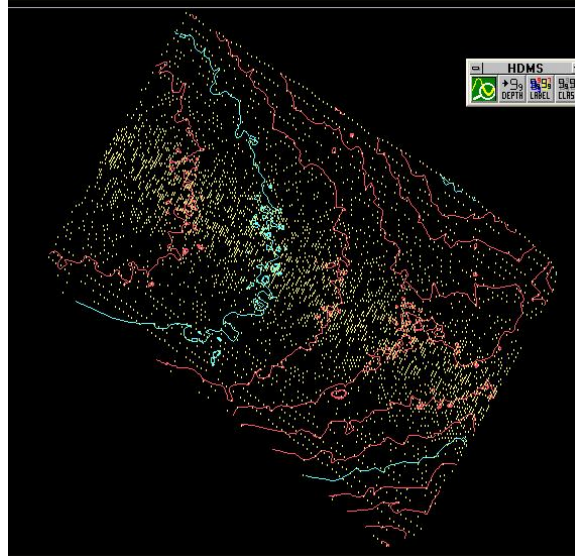
### 3.0 Results

The data thinning process can only be considered as meaningful if the MBES dataset is cleaned from any systematic, blunders and random errors. The final aim is to reduce the MBES dataset size but maintain the accuracy of final contour lines. By using these three line algorithms, differences in results are expected and can be analysed as tabulated below in the Table 3.1. Using the same dataset will permit the analysis to carry the same weight.

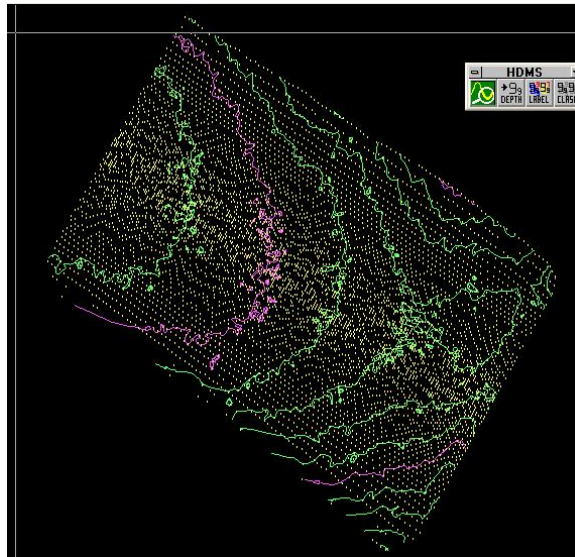
**Table 3.1:** Results from different algorithms

	Algorithms	Total data	Remained data	Eliminated data	Remarks
1.	Douglas Peucker	10,643 (100%)	3,477 (32.7%)	7,166 (67.3%)	Resolution=15cm
2.	Single Swathe Reducer	10,643 (100%)	5,731 (53.8%)	4,921 (46.2%)	Resolution=15cm Range = 10m
3.	Skip N Points	10,463 (100%)	2,130 (20.0%)	8,513 (80.0%)	Skip every 5 pts

From the above table it shows that the most effective algorithm to be used for data thinning is Skip N Points method as it reduces 80% of the data. In this study, skipped point number is set to five. User can specified other values and different result would be achieved. While it is good to reduce high amount of data, the most important thing is not only the percentage of the point reduction but also the integrity of the final contour. The reliable method to identify which algorithm suit the best is by a comparative study on the final contours produced by the three different ways of data thinning.

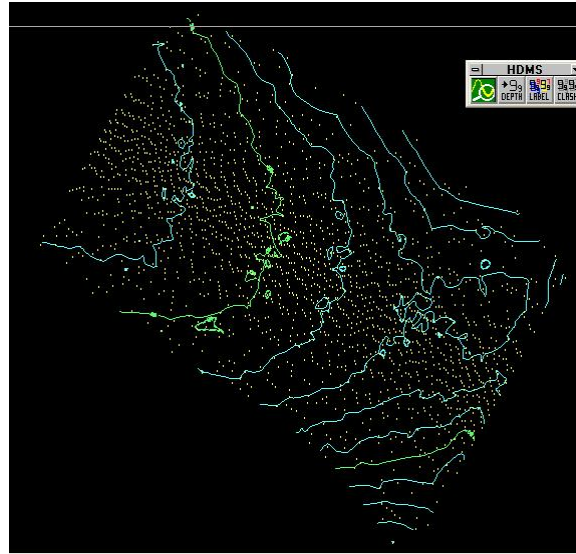


**Figure 3.1:** Result based on reduced dataset by Douglas Peucker algorithm



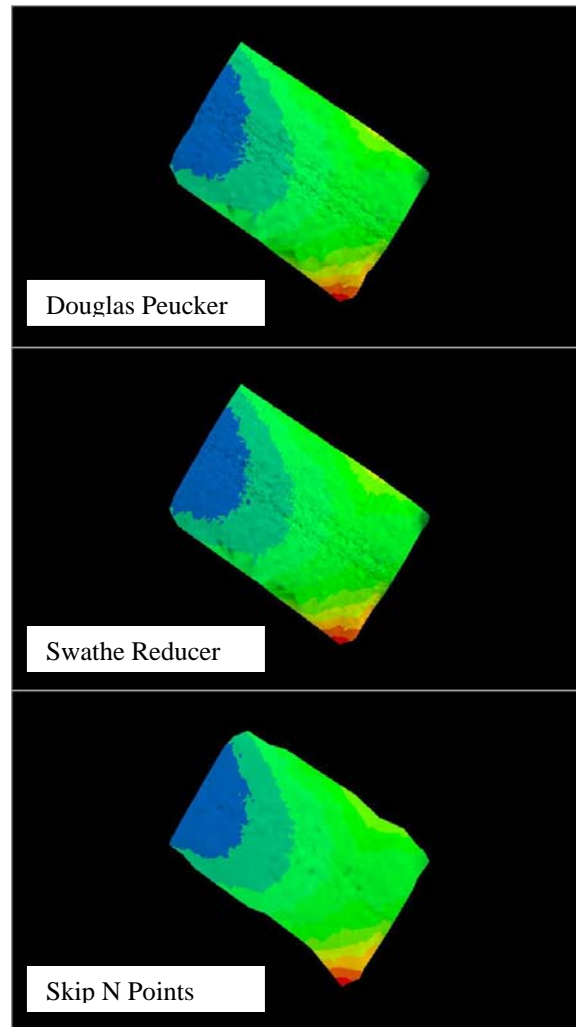
**Figure 3.2:** Result based on reduced dataset by Single Swathe Reducer algorithm





**Figure 3.3:** Result based on reduced dataset by Skip N Points algorithm

From the above results (Figure 3.1, Figure 3.2 and Figure 3.3), it can be concluded that the three data thinning algorithms had yield almost the same contour line pattern. The only major difference is the contour lines produced by Douglas Peucker and Single Swathe Reducer look more uneven and irregular even though during the contouring process they were smoothed by the Spline method. This is due to high dataset density still exists after being reduced by both algorithms especially in Single Swathe Reducer algorithm. The produced contour lines would follow every detail of the entire sounding points. As a result the lines were seen to be crooked although true. These applications will be good for mapping of the area where suspected man-made features such as pipelines on the seabed or coral outcrops within the survey areas. It can show more details with high resolution dataset available. For a flat seabed, the potential of using the Skip N Points is highly recommended as the result shows that there is no significant difference in generated contour lines with other results produced by other algorithms. The contour lines produced by this algorithm look smoother. It was generated by a much smaller reduced dataset. It is common that the contour should be smoothed although it does not show every point value. Bear in mind that contours do not necessary follow exactly every single depth point (Brennan, 2006), especially when smoothing algorithm is applied. The results are still accepted as true contours.



**Figure 3.4:** Colour-coded contour resulting from different data thinning algorithms

From Figure 3.4, Skip N Points algorithm produced more smooth surface compare to the other two algorithms, based on the resultant colour-coded contour between these three algorithms.

#### 4.0 Conclusion

The line algorithm yields data reduction in the across-track direction by examining only one swathe at a time. No reduction in the along-track direction between different swathes is carried out. In the above dataset example, the area is relatively flat seabed. It is expected that more data points will be retained if the seabed is rough and uneven. Based on the results, it can be concluded that all three data thinning algorithms will produce almost similar results. The advantages of having data thinning are:

- Reduced plotting time; as the number of depth points is reduced through the data thinning, the plotting speed is increased.
- Reduced storage; data thinning can reduce a dataset by 80% without jeopardize the accuracy on the contour lines. This process is significantly savings in computer memory.
- The advantages of using TIN-based DTM over the grid method; in this case the line algorithm preserves high resolution with respect to the smallest object to be detected. A

volume computation using the TIN-based DTM of the reduced dataset proves more accurate results compared to that using a gridded dataset (Bottelier *et al.*, 2000). In contrast, a gridded DTM flattens small sized structures at the bottom in case a limited number of grid points were used.

- Fewer points can represent smooth bottom topography.
- It is possible to represent the bottom within pre-defined target accuracy.

From a comparative study, for fairly flat seabed area in the above example, the best data thinning algorithm to be used which provides the highest rate of data reduction and preserves accurate contour lines is Skip N Points algorithm. Furthermore the method is a very straight forward in programming algorithms due to non-topological relationship with neighbouring beams is involved in computation.

## **REFERENCES**

- Bottelier, P., Haagmans, R. and Kinneging, N. (2000). *Fast Reduction of High Density Multi-beam Echo sounder Data for Near Realtime Applications*. *The Hydrographic Journal*, No 98, October 2000.
- Brennan, C. (2006). *RESON Training - Universiti Teknologi Malaysia*. Training material-CD January 2006.
- Capena, G., Bergem, O. and Pace, N.G. (1999). *Trismap: A Fast way to Deal With New Multibeam Sonar Data*. *Sea Technology*, Vol. 40, No. 10, October 1999, pp. 49-52.
- Coastal Oceanographics, Inc. (2003). *MBMax Sounding Filters*.  
[http://www.coastalo.com/newsletter/8\\_01/mike.htm](http://www.coastalo.com/newsletter/8_01/mike.htm); Site visited, September 06, 2003
- Grounds, M. (2004). *Intelligent Reduction of Sweep Survey/LIDAR Large Data Sets*.  
<http://www.riversrus.com>; Site visited, February 12, 2004
- McMaster, R. (2004). *UNIT 48 - Line Generalization*.  
<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u48.html#SEC48.2.2>; Site visited, February 10, 2004