The Establishment of Automatic Detection Procedures for Far and Near Outliers in Multibeam Echo Sounding Dataset

Othman Mohd Yusof
Department of Surveying Science and Geomatics
Faculty of Architecture, Planning and Surveying
Universiti Teknologi MARA
40450 Shah Alam, Selangor
Malaysia
email: omy2006@yahoo.co.uk

Mohd Razali Mahmud
Department of Geomatic Engineering
Faculty of Geoinformation Science and Engineering
Universiti Teknologi Malaysia
81310 UTM Skudai, Johor
Malaysia
email: razali@fksg.utm.my

Abstract

Multibeam echosounder (MBES) has tremendously improved the rate of data collection in terms of time spend and data density. MBES collects full data coverage of seabed within significantly short period of time. However more time is required for data cleaning process in post-processing mode (Cronin *et al*., 2003). The data cleaning process is an important event as the collected data comprised of not only real seabed profiles but associated with erroneous data called outliers. The integrity of acquired data could be validated if a ground truth validation could be performed. However a systematic checking on the grounds of multibeam data is impossible (Mori, 2003). This technical paper discusses on the development of automatic detection of MBES outliers which is divided into two categories namely far outliers and near outliers. From various programs developed using Microsoft Visual Basic, the study tries to investigate and establish the least procedures to detect MBES outliers.

Keywords: far outliers, near outliers, robust detection, statistical elimination, cross-validation

1.0 Introduction to Data Cleaning

Multibeam echosounder (MBES) system has the capability of producing multiple beams within each ping. Hundreds of beams continuously strike the seabed in a fan-shaped across the vessel tracks. The rate of ping increases with decreases of depth. The faster ping rate means more data being collected. This led to enormous dataset to be justified whether the data really represents the true seabed or outliers. Human visual interpretations and traditional line-by-line data inspections have limited capacity to deal with high data density. Time consuming and inconsistent of final data cleaning results contribute to the main weaknesses of these traditional procedures. Experienced operators are highly demanded in order to preserve consistency of the final results especially to distinguish between real seabed and outliers. To counter these downsides of the traditional procedures, fully automatic outlier's detection programs have been developed using Microsoft Visual Basic version 6.0. As no ground truth is available for data quality assessment, the validation of the data should be carried out using mathematical model. As the MBES outliers can be calculated, represented and modeled, it is strongly recommended that the process can be performed in fully automatic mode that can support the determination and elimination of MBES outliers.

MBES outliers can be grouped into two categories:

(i) far outliers, and
(ii) near outliers

The concept of program workflow is to clean the MBES dataset from far outliers in the first stage then followed by cleaning of near outliers in the next stage. Once the dataset is cleaned from far outliers, the remaining data is said to be in normal distribution.

2.0 MBES Far Outliers

Points are detected as far outliers if fall within threshold of 6σ to 9σ (Hekimoglu, 1999). A single displacement in a sample is the typical example of the far outlier where it can be referred as the point that deviate from its neigbourhood point trends or normal anomalies (Figure 1). This point does not represent the overall trend of seabed surface and is considered coming from different population from the sample being studied (Mohd Razali and Othman, 2005).

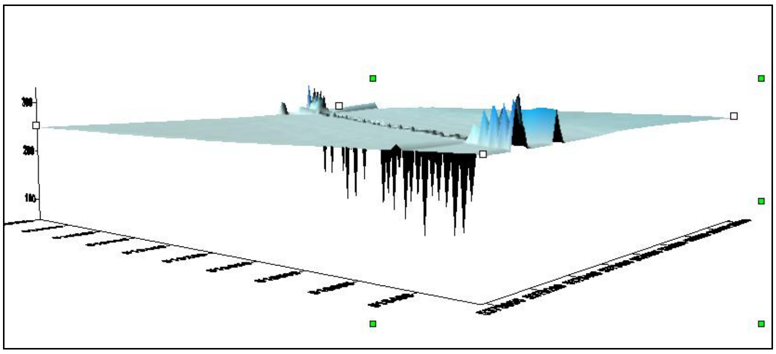

Figure 1 Far outliers in MBES dataset

In this study, far outlier detection programs used robust algorithm concept instead of using the threshold of 6σ to 9σ. Robust algorithm has certain advantages, especially the capability of the algorithm to find an approximation of the real value even with the existence of multiple outliers in the same sample (Capena *et. al*., 1999). The algorithm performance is not affected by the high number of outliers within that sample.

Under the MBES robust filtering algorithms, the following programs were developed to facilitate the detection and elimination of MBES far outliers:

(i) Quick View
(ii) Depth gating
(iii) Outer beam limit
(iv) Across-track line anomalies
(v) Along-track line anomalies

From the list above, various robust filtering algorithms are selectable by a user during processing. The user is not restricted to any sequence of the above algorithms. It is not a rigid procedure. The user can choose to use either all algorithms or skip any algorithm as required. The programs for the detection and removing outliers are explained accordingly in the following discussions.

2.1 Robust Outliers Detection Using Quick View Algorithm

Far outliers can be identified through visualization if the points are situated at far off above from the rest of the data profile. Figure 2 clearly explains this phenomenon, displayed by Quick View program.

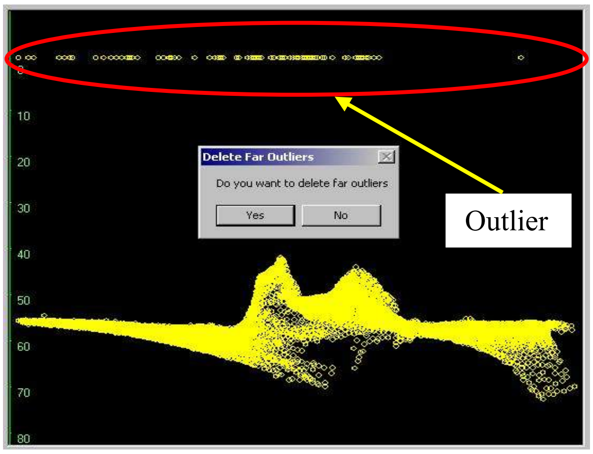

Figure 2 Typical MBES far outliers located at less than zero value depth

2.2 Robust Outliers Detection Using Depth Gating Algorithm

Depth gating program uses 1-D technique. Knowledge of overall depth of surveyed area is essential especially the minimum and maximum depth values of the surveyed area. With these threshold values, any point exceeded the maximum depth limit or less than minimum depth limit is trapped and considered as an outlier, thus will be deleted from the MBES dataset. The testing will be carried out on every single point against these threshold values specified by the user. It is expected that most of the far outliers are detected by this technique. The red dots displayed in Figure 3 are the detected outliers by this robust technique.

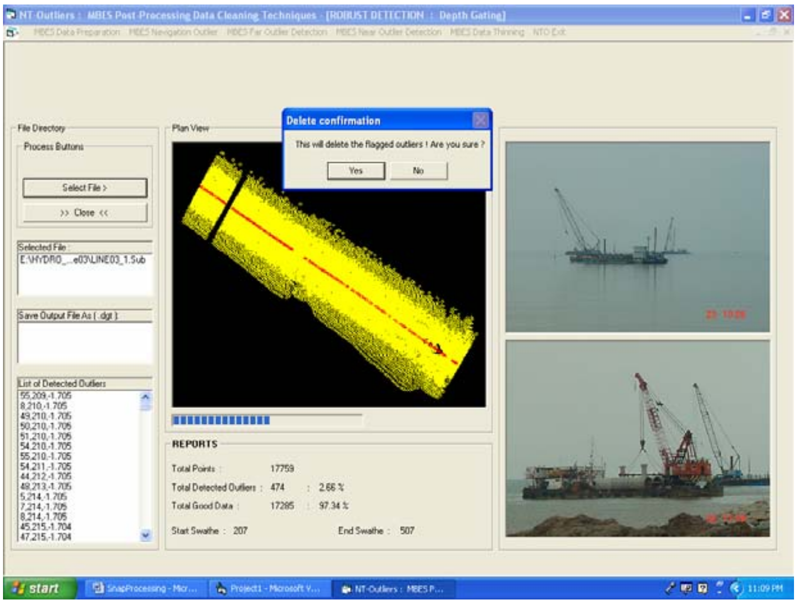

Figure 3 Far outliers detection by Depth Gating algorithm

2.3 Robust Outliers Detection Using Outer Beam Limit Algorithm

This robust technique introduces outer beam limit value on both sides of the swathe. Phenomenon that most outer beam sectors experienced lower signal-to-noise ratio that created noisier bottom had been proved by de Moustier (2003). Thus more occurrences of far outliers are detected in these outer sectors of the swathe, especially for the beams generated at more than 65 degrees, measured from the nadir beam. The red dots on the graphic display box indicated the detected outliers and will be rejected (Figure 4).

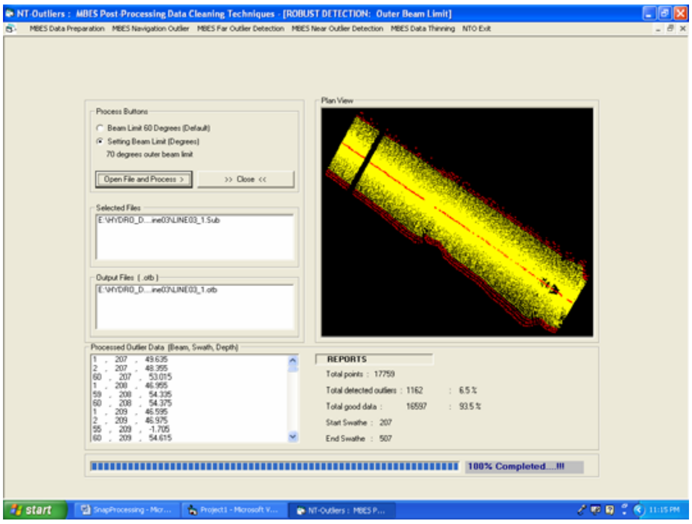

Figure 4 Far outliers detection by Outer Beam Limit algorithm

2.4 Robust Outliers Detection Using Reject Beam ID Algorithm

During data collection at sea, hydrographic surveyors are enable to identify problematic beam ID. This beam ID consistently produced false depth and can be visualised by monitoring the swathe profile display box (waterfall display box) on the navigation computer screen. Direct robust outliers detection program has been developed to search and delete beam IDs that specified by a user. The user can specify any number of beams by typing beam ID numbers and the program will delete all the declared beam ID points located in every swathe within the file. The red dots on the graphic display box indicated the detected outliers and will be rejected (Figure 5).

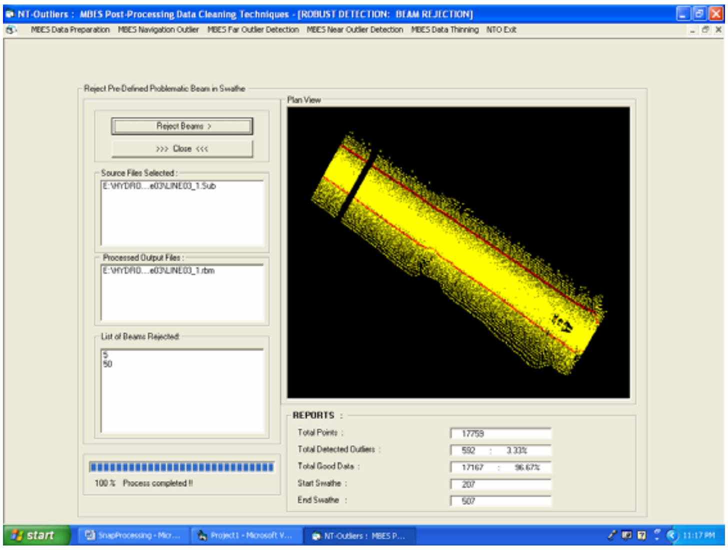

Figure 5 Beam Rejection sub-menu

2.5 Robust Outliers Detection Using Across-Track Line Algorithm

Seabed anomaly is always represented by continuously smooth surface. Gardner *et al.* (1998) suggested that the across-track slope anomalies if exceeded 25° will be considered as an outlier. The red dots represented the detected outliers as shown in Figure 6.

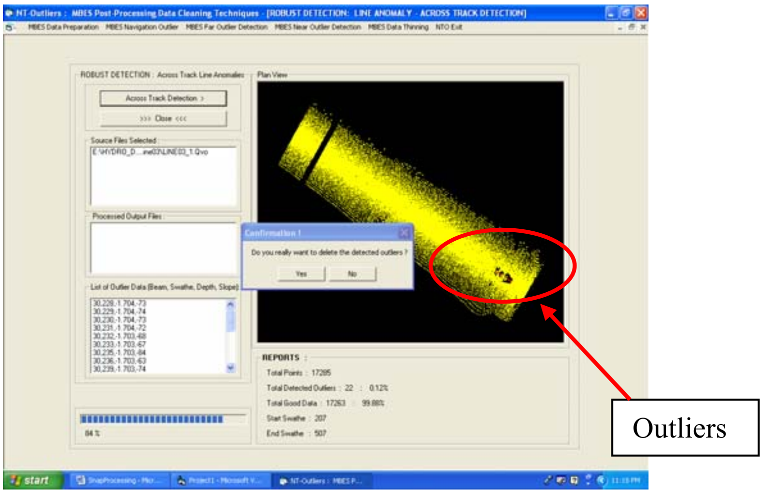

Figure 6 Far outliers detection by Across-Track Line algorithm

2.6 Robust Outliers Detection Using Along-Track Line Algorithm

The along-track line anomaly investigates the slope angle anomalies along the vessel track direction. The limit is set to 55° angle, used to identify and remove a single outlier among each set of three soundings (from three consecutive swathes) when the resultant angle has exceeded the criteria angle limit specified by the developed program. The 55° value is accepted based on the study carried out by Gardner *et. al.* (1998).

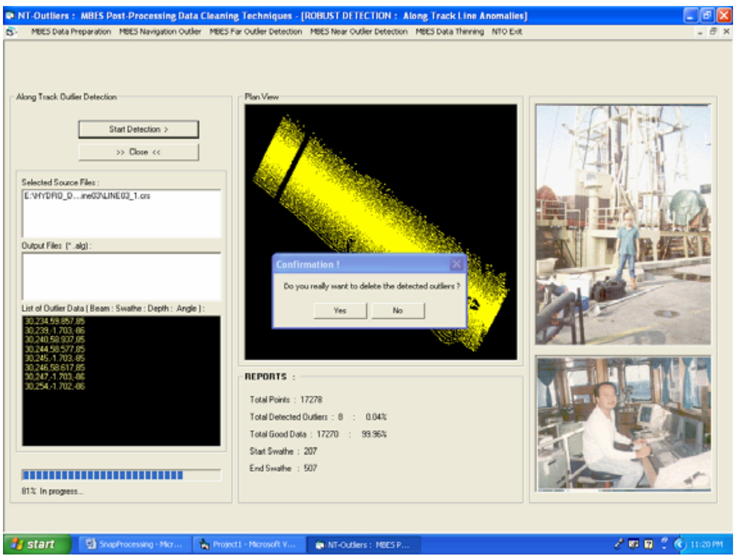

Figure 7 Far outliers detected by Along-Track Line algorithm

3 MBES Near Outliers

Once the dataset has been cleaned from far outliers, the MBES dataset is assumed to be randomly distributed. With this situation, the application of statistical analysis is highly recommended. The approach for near outliers detection in NT-Outliers program is by interpolation technique, the process of estimating the value of a quantity at a location of point to be investigated without considering the point's value. The process is called cross-validation. Most common statistical assumptions in statistical elimination tools are that the data to be estimated have some spatial correlation with data values in the neighbourhood and the dataset should be in a random or near-random character. The density and distribution of control data should be adequate, thus to ensure that accurate estimation could be produced (GEBCO SCDB, 1997).

3.1 Statistical Elimination Tools by Cross-Validation Technique

To identify near outlier, one needs a model for the MBES dataset. A point that violated the model is detected as outlier (AUTOBOX, 2004). The model used in statistical elimination tools is based upon Inverse Distance Weighting (IDW) model in order to estimate depth value. In this research, the local concept using IDW was introduced, that will only consider a small niegbourhood in its sampling. Cross-validation using IDW will be performed in two stages:

(i) line algorithm (1-D) and
(ii) plane algorithm (2-D).

The task of line algorithm is to detect suspected outliers and the plane algorithm is for outlier confirmation and cleaning.

3.2 1-D Cross-Validation Using Line Algorithm

The 1-D cross-validation will trace suspected near outliers data using statistical analysis IDW, and to process one swathe at a time. 1-D cross-validation by line anomalies is a technique that performs a reliability test for each sounding in each swathe, working from the first sounding to the last sounding across the swathe. This technique only detected and flagged potential near outliers using more stringent criterion; one standard deviation (±1σ) as an acceptance region from the statistical sample mean value. The red dots displays in the display box indicated the detected potential outliers using ±1σ (Figure 8).

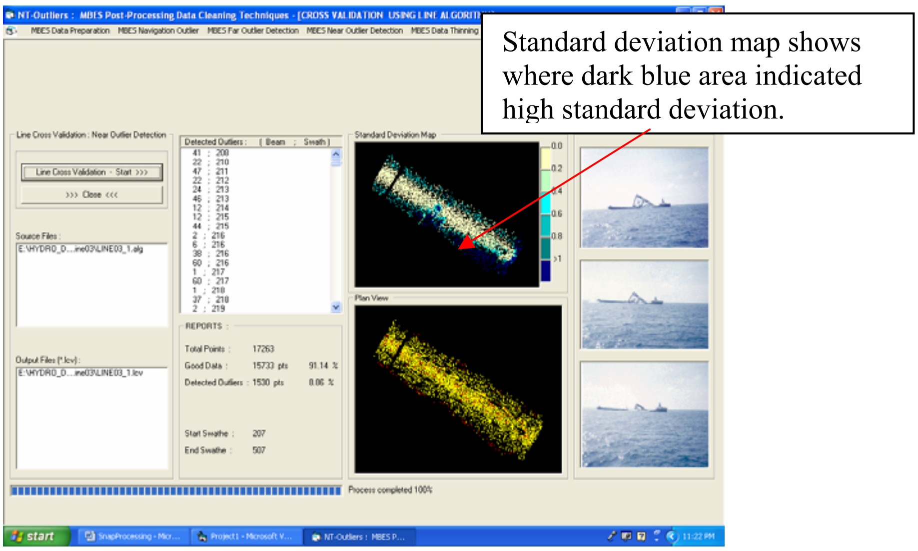

Figure 8 Cross-validation using line algorithm to detect near outliers

3.3 2-D Cross-Validation Using Plane Algorithm

The developed program has considered 2-D estimation based on IDW method, fundamentally a function that averages the surrounding points. This can be computationally very extensive particularly when dealing with large datasets. However the program is designed in such a way that it only investigates on those flagged points detected in the previous 1-D algorithm. Not the entire data will be tested for their ±2σ limit. The plane algorithm (2-D) will involve the surrounding neighbourhood (fore and rear swathes) of the suspected outlier points in its estimation. Any point that falls beyond this limit will be detected as outlier and deleted. Figure 9 shows colour coded contour surface that has been cleaned from outliers.

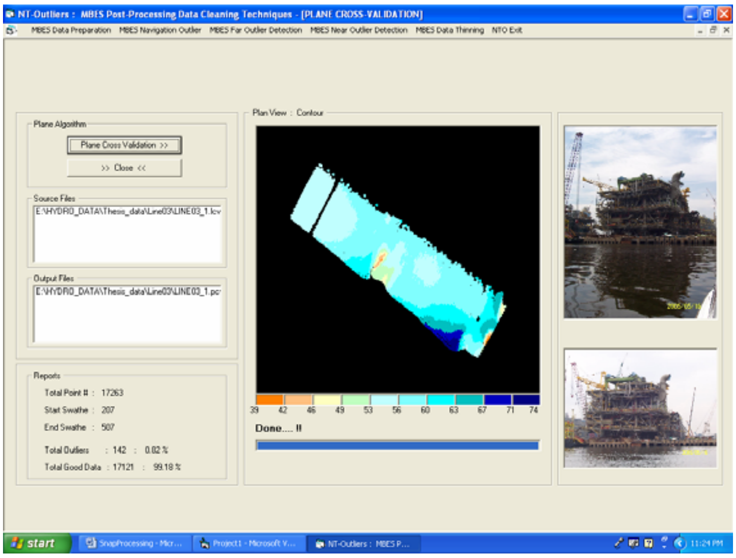

Figure 9 Plane cross-validation eliminated near outliers and displayed colour coded contour surface

4 Conclusion

Results generated through robust algorithms in detecting MBES far outliers were very impressive although the algorithms were straightforward and quite simple. The robust algorithms can be used as alternative ways of elimination of far outliers to replace time consuming line-by-line visual investigation technique by traditional method. Human visual analysis has a limited capability which it fails to process data quickly and efficiently. Further more, visual interpretation cannot identify MBES near outliers as the outliers tend to mix very close to the real seabed profile. Therefore the detection process needs to implement a mathematical model in its statistical analysis called IDW. The results indicated that cross-validation technique (statistical elimination tools) provides sufficient capability to detecting near outliers. The two main issues now is to determine the minimum algorithms needed and what type of algorithms serve the least requirement for the outlier detection process. From the study, the following sequence of the least detection procedures has been established:

(i) Depth gating – to detect far outliers
(ii) Across-track line anomalies – to detect far outliers
(iii) Along-track line anomalies – to detect far outliers
(iv) 1-D cross-validation using line algorithm – to detect near outliers
(v) 2-D cross-validation using plane algorithm – to detect near outliers

The above procedures proved that most of the far and near outliers have been tracked and eliminated. Although various techniques of outlier's detection have been applied in the programs, the results showed that the above procedures can be considered as the least steps that should be carried out for MBES outlier's detection. These five criteria have served perfectly to the expectation standard and fulfill the research objectives.

REFERENCES

AUTOBOX (2004). *Outlier Detection and Plus/minus 3-sigma charts*. http://www.autobox.com/outlier.html Site visited 30 April 2004.

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.

Cronin, D., Broadus, M., Byrne, S., Simmons, W. and Gee, L. (2003). Hydrographic Work Flow – From Planning to Products. The Hydrographic Society of America, U.S. Hydro 2003 Conference, Mississippi: March 24-27, 2003.

de Moustier, C. (2003). Field Evaluation of Sounding Accuracy in Deep Water Multibeam Swathe Bathymetry. IEEE Webmaster.
http://www.ieee.org/organizations/society/oes/html/fall02/Field_eval.html;
site visited: September, 2003.

Gardner, J. V., Mayer, L. A. and Hughes Clarke, J. E. (1998). Cruise report, RV inland surveyor Cruise is-98; The Bathymetry of Lake Tahoe, California-Nevada. Open-File Report 98-509, USGS.

GEBCO SCDB (1997). *On the preparation of a gridded dats set from the GEBCO Digital Atlas contours* – Draft Version 9 Appendix A

Hekimoglu, S (1999). Robustifying Conventional Outlier Detection Procedures. Journal Of Surveying Engineering, May 1999, pp. 69-86

Hughes Clarke, J. E. (1999). Provisional Swathe Sonar Survey Specifications. National Topographic and Hydrographic Authority, Land Information New Zealand, TH Technical Report #2, August 1999.

Mohd Razali Mahmud and Othman Mohd Yusof (2005). Automatic Filtering of Far Outliers in Multibeam Echo Sounding Dataset Using Robust Detection Algorithms. *International Symposium and Exhibition on Geoinformation 2005*. Penang, Malaysia. September 2005

Mori, A. (2003). Seamless Multibeam Data Analysis and Management. Hydro INTERNATIONAL, Vol. 7, No. 3, April 2003, pp. 43-45.