

Hybrid Machine Learning and CUBE Method for Multibeam Data Cleaning

M. El-Diasty¹, R. Abdalla¹ & F. Alsaqa²

¹ Sultan Qaboos University, Muscat, Oman

² King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: Multibeam data contains different types of errors that are classified as systematic errors, random errors and gross errors (outliers). Accurate bathymetric base surface production requires efficient cleaning methods to detect and reject the outliers. The manual cleaning method is tedious and time-consuming method. The need for automation of data cleaning is essential to reduce the processing time for multibeam data processing tasks. The newly developed AI-based Machine Learning (ML) method is a promising supervised method for automatic multibeam outlier detection and rejection. In this paper, a Hybrid ML-CUBE method was introduced and evaluated using multibeam datasets collected by Kongsberg EM712 and Reson T50-P multibeam echosounders. It was also found that if the outliers were not successfully detected and rejected, the accuracy of the produced base surfaces are degraded by 0.61 m and 0.58 m for EM712 and T50-P tests, respectively, which exceed the International Hydrographic Organization (IHO) special order. The significance of the proposed Hybrid ML-CUBE method is that it is a rigorous and automatic outlier rejection method for multibeam data cleaning and a rigorous bathymetric surface generation method for random errors reduction.

1 INTRODUCTION

Multibeam data includes different error types. These errors can be classified as systematic errors, gross errors (outliers) and random errors [1]. The source of these errors can be humans (hydrographers), equipment (echosounders) or environment (sea state). The systematic errors can be corrected using the functional models utilizing misalignment errors (roll, pitch and yaw errors) and time delay calibrated by the patch test in addition to the sound velocity profile collected by the velocimeter. The random errors are corrected when the final base surface is produced. However, the outliers in multibeam data should be detected and rejected using cleaning methods to obtain accurate bathymetric surface [2]. Figure 1 shows an example of different types of multibeam errors. The main reasons for the outliers could be

cavitation/bubble sweep, loss of bottom on outer beams, loss of bottom lock and false returns from the water column [3].

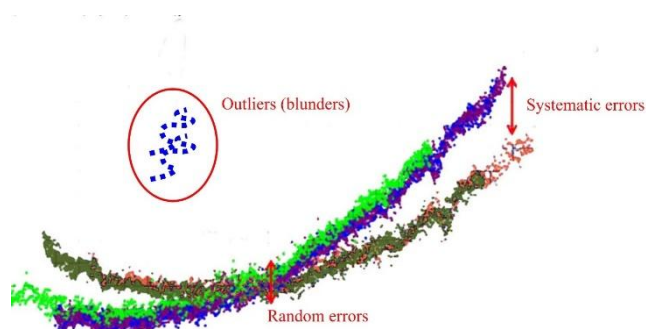


Figure 1. Types of errors exists in multibeam sounding where five survey lines displayed in different colors and outliers found in the blue survey line only

The manual and automatic cleaning methods can be used to detect and reject the multibeam data outliers. The manual cleaning depends on the visual inspection by the hydrographer to reject any obvious sudden spikes (outliers) in multibeam data line by line. The manual cleaning method is certainly tedious, critical and subjective method. Also, given the huge amount of data collected by the new generation of current multibeam acquisition systems, the manual cleaning is considered as a time-consuming task for hydrographic offices where the time required for data processing is almost one-to-one when compared against data acquisition time [4]. The need for automation of data cleaning is critical to reduce the processing time for multibeam data processing. The traditional and simple automatic methods for data cleaning are the methods that use the depth limit (based on prior knowledge of depth range), the nadir angles limit and the beam-to-beam angle limit [6]. However, these traditional methods can reject useful multibeam data soundings where these methods depend on the prior knowledge of depth limit which in most cases is not known or depend on the guessing of the nadir angles limit as well as the beam-to-beam slope angle limit. Therefore, advanced and more accurate automatic methods are developed such as AI-based supervised method and unsupervised methods [5] [7][8]. CARIS AI-based Sonar Noise Classifier is one of the promising supervised methods for automatic multibeam data cleaning and is evaluated in this project [7]. The Caris AI-based Sonar Noise Classifier uses the machine learning (ML) three-dimensional convolutional neural network technique ([3][9][10]). The Caris ML-based Sonar Noise Classifier method is based on the estimation of noise confidence level (0-100%) for all multibeam soundings and recommends the rejection of the soundings with noise confidence level value above 50% [3][10]. After the correction for systematic errors and outliers in multibeam data processing, the remaining outliers and random errors are reduced in the digital terrain model or base surface production with different methods such as shoalest depth, swath, uncertainty, and CUBE methods where CUBE filter is the most effective method for bathymetric surface production and random error reduction. Hybrid Uncertainty and Bathymetry Estimation (CUBE) filter initially utilizes the soundings and uncertainty of soundings for high-density multibeam data to create the CUBE surface. CUBE filter takes the depth of the CUBE Surface at a location and compares the sounding depths against the CUBE depth plus a threshold value. The threshold value is defined by the parameters set in the filter. If the difference between the depth of the sounding and the depth of the CUBE Surface exceeds the threshold, then the sounding with residual outliers will be rejected (Calder and Mayer, 2003; Calder, 2003). The primary aim of the CUBE algorithm is to use as much information as possible to determine the true depth at any point in the survey from the noise. CUBE assumes that you are processing good data which requires ML-based method to be initially applied to remove the majority of outliers where only residual outliers and random errors exist. It is worth noting that the bathymetric surface can be generated using uniform grid model or variable gride model [11].

Developing accurate multibeam data cleaning methods are critical for all hydrographic offices to achieve bathymetric maps that meet the International

Hydrographic Organization (IHO) standards [12]. The main objective of this research is to develop an optimal Hybrid ML-CUBE method for multibeam echosounder data cleaning where the outliers are rejected using ML-based method and random errors are reduced using CUBE method. The proposed method is evaluated and tested using two multibeam datasets collected from Kongsberg EM712 and Reson SeaBat T50-P multibeam echosounders in shallow water case studies areas.

2 HYBRID ML-CUBE METHOD

An optimal method for bathymetric surface creation using the Hybrid ML-CUBE method is investigated in this paper to automatically clean multibeam data from the outliers and rigorously reduce the random errors. The ML method is developed based on the three-dimensional convolutional neural network. The three-dimensional convolutional neural network is the most popular deep learning model that can learn and extract the features based on a unique network structure [9] [13] [14]. The ML method is employed for multibeam data outliers detection and rejection. Then, the CUBE method is employed for bathymetric surface generation where the Kalman filtering is used within the CUBE algorithm to process high-density bathymetry data for random errors reductions [15][16]. Therefore, the proposed Hybrid ML-CUBE method is considered as a rigorous and automatic method for outliers and random errors reduction in post-processing phase where systematic errors are corrected in the pre-processing calibration phase.

The methodology for implementing the proposed Hybrid ML-CUBE method is shown in Figure 2. In step 1, the high-density multibeam soundings are collected using two multibeam echosounders (Kongsberg EM712 and Reson SeaBat T50-P). In step 2, the attitude and navigation data are cleaned from errors, if required. In step 3, the multibeam soundings are georeferenced using the attitude and navigation data. In step 4, the noisy bathymetric surface is created using the shoalest depth true position method where the outliers and random errors are embedded in this surface. In step 5, the CARIS ML-based sonar noise classifier using three-dimensional convolutional neural network technique is employed to clean multibeam data from outliers. The CARIS ML-based sonar noise classifier technique includes: 1) the creation of voxels grids from the high-density multibeam soundings, 2) send the voxels grids to the Caris AI cloud computing server, 3) receive the classified voxels from the Caris AI cloud, 4) map the voxel back to the soundings and reject the soundings with noise confidence level values above 50% [3][10]. The ML-based sonar noise classifier detects and rejects the outliers embedded in the noisy high-density multibeam soundings. In step 6, the CUBE bathymetry is created using the cleaned multibeam high-density multibeam soundings estimated from second step. CUBE is an algorithm that rigorously uses the uncertainty of every cleaned sounding and simultaneously comparing its depth to its neighbors to create the best estimate of the bathymetric surface. The CUBE algorithm reduces the random errors embedded in the cleaned high-density multibeam soundings [15][16]. In the last step, the accuracy of Hybrid ML-

CUBE method is evaluated for the two EM712 and T50-P multibeam echosounders using Root Mean Squares (RMS) errors estimated from the error differences between the noisy and cleaned bathymetric surfaces. Then, the RMS errors are compared against the IHO standards for different hydrographic surveying orders at 95% confidence level.

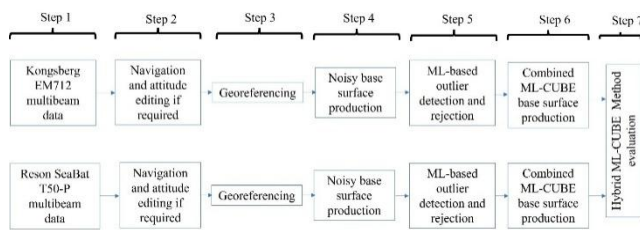


Figure 2. Hybrid ML-CUBE method technique and evaluation

3 DATASETS DESCRIPTIONS

The KAU Hydrography 1 vessel and KAU Hydrography 2 vessel, owned by faculty of maritime studies, King Abdulaziz University, were employed to collect multibeam data. The KAU Hydrography 1 vessel is equipped with Kongsberg EM712 multibeam echosounder and the KAU Hydrography 2 vessel is equipped with Reson SeaBat T50-P multibeam echosounder for multibeam data collection along with navigation solutions from POS-MV positioning and orientation system for the georeferencing of the collected multibeam datasets [17][18][19]. Figures 3 and 4 show the KAU Hydrography 1 vessel along with the Kongsberg EM712 multibeam echosounder and KAU Hydrography 2 vessel along with the Reson SeaBat T50-P multibeam echosounder, respectively. Figures 5 and 6 show the survey lines in two different shallow waters case study areas for test 1 using KAU Hydrography 1 vessel and for test 2 using KAU Hydrography 2 vessel data, respectively.



Figure 3. KAU Hydrography 1 vessel (3.a) along with Kongsberg EM712 multibeam echosounder (3.b) and POS-MV positioning and orientation system (3.c) [17][19]



Figure 4. KAU Hydrography 2 vessel (4.a) along with Reson SeaBat T50-P multibeam echosounder (4.b) and POS-MV positioning and orientation system (4.c) [18][19].

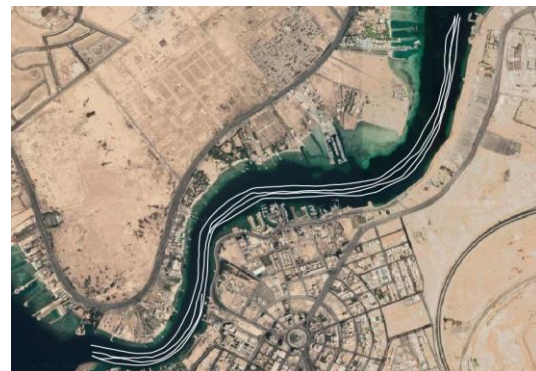


Figure 5. KAU Hydrography 1 vessel test 1 survey lines (Sharm Obhur)

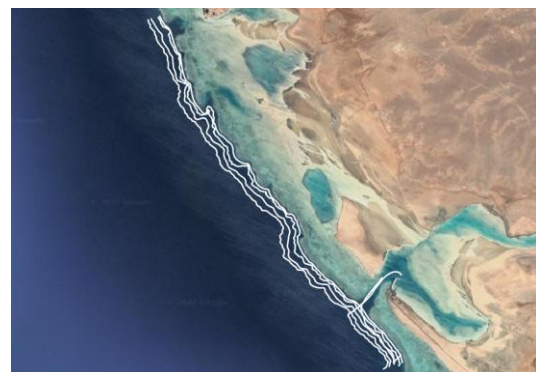


Figure 6. KAU Hydrography 2 vessel test 2 survey lines (Alrayes)

4 RESULTS AND DISCUSSIONS

The high-density multibeam soundings collected from test 1 (EM712 multibeam echosounder) and test 2 (T50-P multibeam echosounder) were about 1,134,026 soundings and 3,493,434 soundings, respectively. The multibeam soundings from test 1 and test 2 were georeferenced using the Caris HIPS and SIPS software and the base surfaces for test 1 and test 2 data were produced using the shoalest depth true position method without applying any cleaning method. Figures 7 and 8 show the produced base surfaces for test 1 and test 2 which include the outliers and we call these surface noisy bathymetries for test 1 and test 2.

The ML-based cleaning method was implemented using the CARIS Sonar Noise Classifier processor for test 1 and test 2 high-density multibeam soundings. The ML-based CARIS Sonar Noise Classifier processor works using the Caris AI Mira authentication service and the selected options for Sonar Noise Classifier processor to clean the soundings confidence noise level value was above 50%. Figure 9 and 10 show examples of the ML-based CARIS Sonar Noise Classifier results in swath editor and subset editor windows, respectively, for selected number of soundings which lists the values of the noise confidence level value (0-100%) for each sounding and the rejected soundings with the noise confidence level values above 50%. The Hybrid ML-CUBE method was implemented and Figures 11 and 12 show the cleaned bathymetries for test 1 and test 2, respectively, where the ML-based Sonar Noise Classifier method was implemented for outliers rejection and CUBE method was implemented for random errors reduction. To test the performance of Hybrid ML-CUBE method, the error differences between the noisy and Hybrid ML-CUBE based cleaned bathymetries for tests 1 and 2 were estimated and shown in Figures 13 and 14. Also, the bathymetric error differences from test 1 and test 2 were used to generate the histograms distributions and cumulative distributions plots as shown in Figures 15, 16, 17 and 18, respectively. The error differences from test 1 ranges from -10.61 m to 6.77 m with mean value around -0.12 m and error differences from test 2 ranges from -15.74 m to 8.31 m with mean value around -0.09.

Then, the bathymetric error differences from test 1 and test 2 were used to estimate the summery statistics as listed in Tables 1. The RMS errors at 95% confidence level listed in Table 1 are estimated using the absolute means and the RMS errors at 68% confidence level using the following formula [20]:

$$RMS_{95\%} = |mean| + 1.96 * RMS_{68\%} \quad (1)$$

where the mean and $RMS_{68\%}$ are estimated from the error differences the error differences between the noisy and Hybrid ML-CUBE based cleaned bathymetric surfaces for tests 1 and 2.

The estimated RMS errors at 95% confidence level caused by the outliers and random errors are 0.61 m and 0.58 m for test 1 and test 2, respectively, which exceed IHO special order requirement of maximum allowed total vertical uncertainty that equal 0.34 m but meets IHO order 1a requirement of maximum allowed total vertical uncertainty that equal 0.63 m both were estimated using the average depth of the area (30 m in the tests 1 and 2) at 95% confidence level.

Therefore, the rejection of the outliers and reduction of random errors using the proposed Hybrid ML-CUBE method are essential and critical to produce bathymetric maps with accuracy within the IHO special order level when multibeam data is processed. The results show that the Hybrid ML-CUBE method can significantly improve the overall accuracy of the produced final bathymetric maps. The advantage of the Hybrid ML-CUBE method that is based on the use of the AI-based sonar noise classifier subscription is that it is a very efficient and accurate automatic cleaning method in detecting and rejecting the outliers.

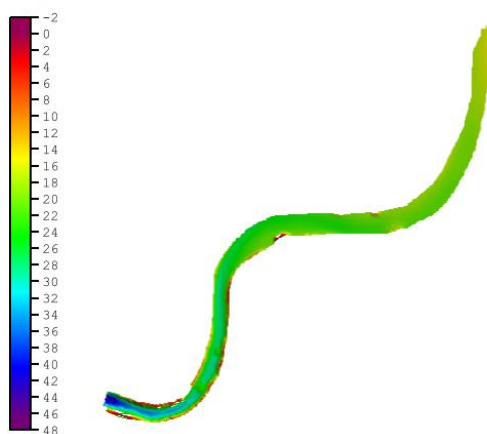


Figure 7. Noisy bathymetry for test 1 (EM712 multibeam echosounder) referenced to the WGS84 ellipsoid (units in meters)

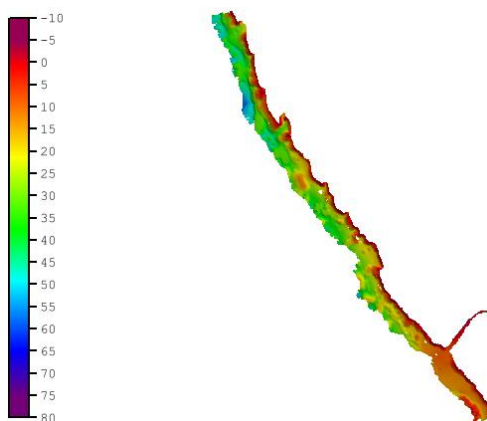


Figure 8. Noisy bathymetry for test 2 (T50-P multibeam echosounder) referenced to the WGS84 ellipsoid (units in meters)

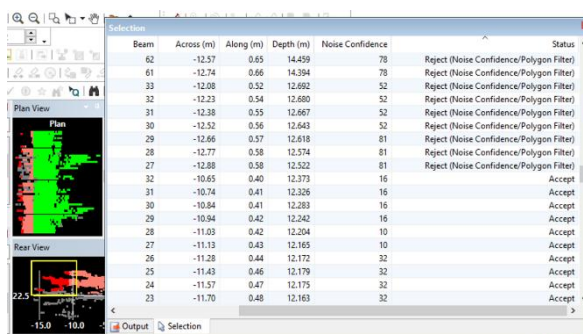


Figure 9. ML-based CARIS Sonar Noise Classifier Results for selected soundings located inside the yellow box using CARIS swath editor

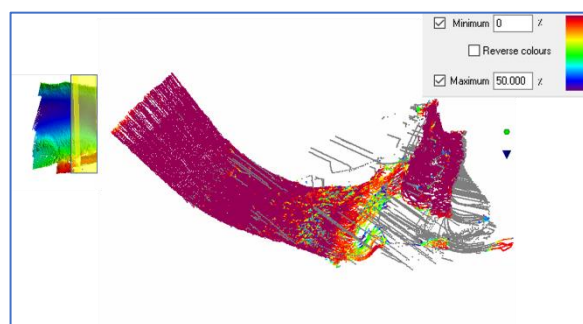


Figure 10. ML-based CARIS Sonar Noise Classifier noise confidence values for soundings using subset editor where the grey color shows the rejected soundings and the red-to-blue color soundings shows the accepted soundings

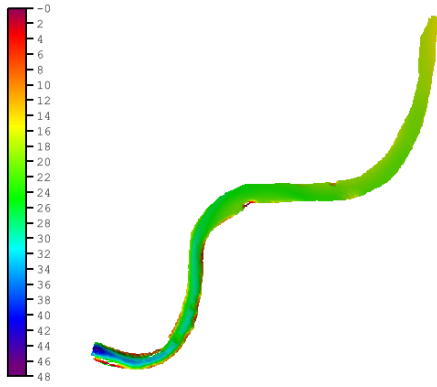


Figure 11. Hybrid ML-CUBE based cleaned bathymetry for test 1 (EM712 multibeam echosounder) referenced to the WGS84 ellipsoid (units in meters)

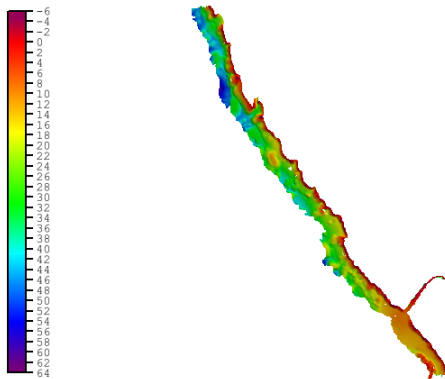


Figure 12. Hybrid ML-CUBE based cleaned bathymetry for test 2 (T50-P multibeam echosounder) referenced to the WGS84 ellipsoid (units in meters)

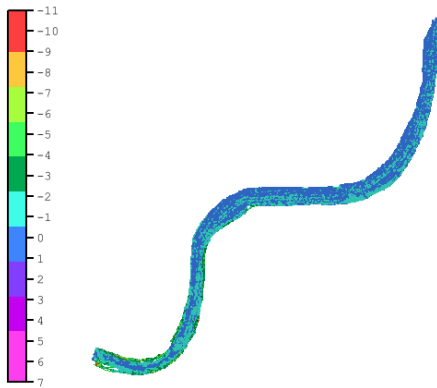


Figure 13. Error differences between the noisy and Hybrid ML-CUBE based cleaned bathymetries for test 1 (EM712 multibeam echosounder)

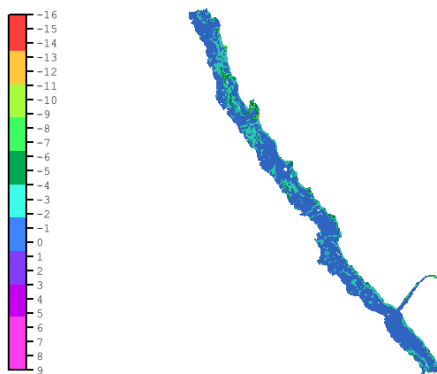


Figure 14. Error differences between the noisy and Hybrid ML-CUBE based cleaned bathymetries for test 2 (T50-P multibeam echosounder)

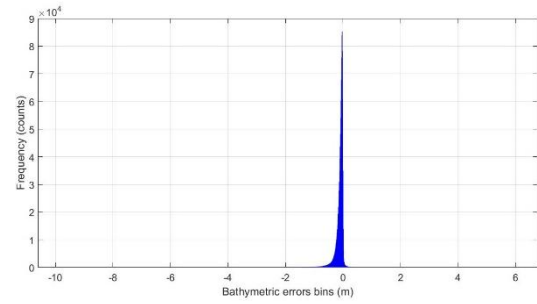


Figure 15. Histogram distribution for test 1 (EM712 multibeam echosounder) bathymetric error differences

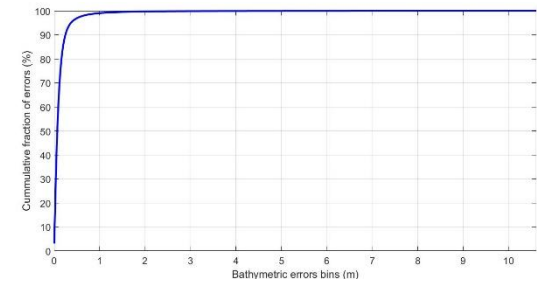


Figure 16. Cumulative distribution for test 1 (EM712 multibeam echosounder) bathymetric error differences

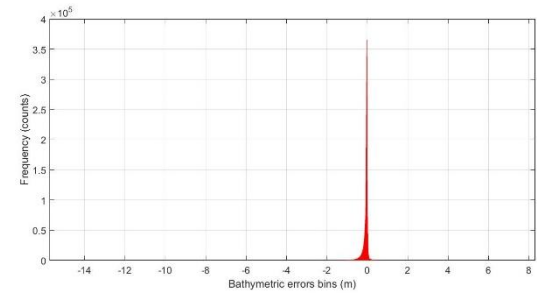


Figure 17. Histogram distribution for test 2 (T50-P multibeam echosounder) bathymetric error differences

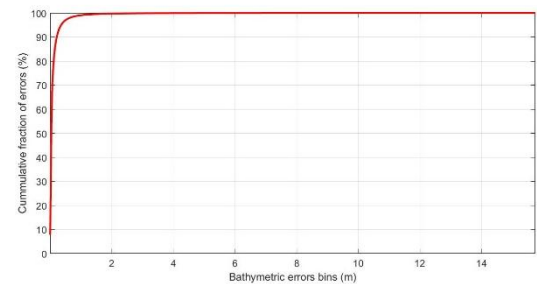


Figure 18. Cumulative distribution for test 2 (T50-P multibeam echosounder) bathymetric error differences

Table 1. Summary of statistical results of the error difference between noisy and the Hybrid ML-CUBE cleaned bathymetries produced from test 1 and test 2

Parameter\solution	Test 1 (EM712 multibeam echosounder)	Test 2 (T50-P multibeam echosounder)
Maximum	6.77 m	8.307 m
Minimum	-10.61 m	-15.743 m
Soundings total count	1,134,026	3,493,434
Mean	-0.12 m	-0.09 m
RMS (68% confidence level)	0.25 m	0.25 m
RMS (95% confidence level)	$0.12 + 1.96 * 0.25 = 0.61$ m	$0.09 + 1.96 * 0.25 = 0.58$ m

5 CONCLUSION AND RECOMMENDATIONS

The objective of this paper is to evaluate the proposed Hybrid ML-CUBE method for outliers rejections and random error reduction using high-density multibeam soundings collected by Kongsberg EM712 and Reson SeaBat T50-P multibeam echosounders. The noisy bathymetries were produced using the shoalest depth method from the noisy Kongsberg EM712 and Reson SeaBat T50-P multibeam datasets. Afterwards, the Hybrid ML-CUBE method was employed to reject outliers and reduce random errors in the Kongsberg EM712 and Reson SeaBat T50-P multibeam datasets and produce the cleaned bathymetries. The difference surfaces between the noisy and cleaned bathymetries showed that outliers in Kongsberg EM712 and Reson SeaBat T50-P datasets can successfully be detected and rejected using the ML-based CARIS Sonar Noise Classifier automatic cleaning method and the random errors can successfully be reduced using the CUBE method. It was found that if the outliers were not successfully rejected and random errors were not successfully reduced by Hybrid ML-CUBE method, the accuracy of the produced bathymetries are degraded by 0.61 m and 0.58 m in the tests conducted by EM712 and T50-P, respectively, which exceed the IHO special order but meet order 1a requirements. Therefore, it is recommended to use Hybrid ML-CUBE method for the bathymetric mapping with multibeam echosounders. The significance of the Hybrid ML-CUBE method is that it is an automatic outlier rejection method for multibeam data cleaning and a rigorous bathymetric surface generation method for random errors reduction.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of deanship research fund, Deanship of Research, Sultan Qaboos University, Oman, grant number RF/ENG/CAED/22/02. The data was provided to the authors by the Research and Consulting Institute (RACI), King Abdulaziz University. The author, therefore, gratefully acknowledges RACI's support.

REFERENCES

- [1] Ghilani, C. (2017). *Adjustment Computations: Spatial Data Analysis*, Sixth Edition. John Wiley & Sons, Inc.
- [2] Le Deunf, J.; Debese, N.; Schmitt, T.; Billot, R. (2020), "A Review of Data Cleaning Approaches in a Hydrographic Framework with a Focus on Bathymetric Multibeam Echosounder Datasets" *Geosciences*, 10, no. 7: 254.
- [3] Hoggarth, A. (2019), "Using Artificial Intelligence to Clean Multibeam Echo Sounder Data", *GEBCO Symposium*, University of New Hampshire, November 2019.
- [4] Debese, N.; Bisquay, H. (1999) "Automatic detection of punctual errors in multibeam data using a robust estimator." *International Hydrographic Review*, 76, pp. 49–63.
- [5] Debese, N.; Moitié, R. (2012), "Multibeam echosounder data cleaning through a hierarchic adaptive and robust local surfacing", *Computers & Geosciences*, vol. 46, pp.330-339.
- [6] Bjørke, J.; Nilsen, S. (2009), "Fast trend extraction and identification of spikes in bathymetric data." *Computers & Geosciences*, 35, 6, pp. 1061-1071.
- [7] Foster, B. (2019), "Applications of Machine Learning in Hydrographic Data Processing", *US HYDRO 2019*.
- [8] Hou, H.; Huff, L. C.; Mayer, L. (2001), "Automatic Detection of Outliers in Multibeam Echo Sounding data", *US HYDRO 2001*.
- [9] Stephens, D.; Smith, V.; Redfern, T.; Talbot, A.; Lessnoff, A.; Dempsey, K. (2020), "Using three dimensional convolutional neural networks for denoising echosounder point cloud data.", *Applied Computing and Geosciences*, 5, 100016.
- [10] Redmayne, M. (2019), "Using Artificial Intelligence to Clean Multibeam Echo Sounder Data", *Oceans in Action Workshop 4 November 2019*.
- [11] Calder, B.R.; Rice, G. (2011) "Design and implementation of an extensible variable resolution bathymetric estimator." In *Proceedings of the U.S. Hydrographic Conference*, Tampa, FL, USA, 25–28 April.
- [12] IHO (2020), "IHO standards for hydrographic surveys." https://iho.int/uploads/user/pubs/Drafts/S-44_Edition_6.0.0-Final.pdf. 6th Edition March 2020.
- [13] Dong, Z.; Wang, M.; Wang, Y.; Liu, Y.; Feng, Y.; Xu, W. (2022) "Multi-oriented object detection in high-resolution remote sensing imagery based on convolutional neural networks with adaptive object orientation features.", *Remote Sensing*, 14, 950.
- [14] He, K.; Zhang, X.; Ren, S.; Sun, J. (2014) "Spatial pyramid pooling in deep convolutional networks for visual recognition.", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37, pp. 1904–1916.
- [15] Calder, B.R.; Mayer, L. (2003), "Automatic processing of high-rate, high-density multibeam echosounder data", *Geochemistry Geophysics Geosystems*, 4, 6.
- [16] Calder, B. R. (2003) "Automatic Statistical Processing of Multibeam Echosounder Data.", *International Hydrographic Review*, 4, 1.
- [17] Kongsberg (2025), "https://www.kongsberg.com/content/tasets/2c197a1f0a33461b9642276e31b67fb7/datasheet_em_712.pdf", last accessed on February 10, 2025.
- [18] Teledyne (2025), "Teledyne Reson SeaBat T50-P multibeam echosounder datasheet.", https://www.str-subsea.com/uploads/Teledyne-Reson-Seabat-T-50-Multibeam-Echo-Sounder-Datasheet_190104_111650.pdf, last accessed on February 10, 2025.
- [19] Woolven, S.; Scherzinger, B.; Field, M. (1997) "POS/MV-system performance with inertial/RTK GPS integration.", In *Proceedings of the Oceans' 97. MTS/IEEE Conference*, Halifax, NS, Canada, 6–9 October 1997; Volume 2, pp. 1104–1108.
- [20] El-Diasty, M. (2016), "Development of Real-Time PPP-Based GPS/INS Integration System Using IGS Real-Time Service for Hydrographic Surveys." *Journal of Surveying Engineering*, Volume 142, Issue 2, 05015005: pp. 1-8.