Computers & Geosciences 186 (2024) 105572

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

An optimized outlier detection function for multibeam echo-sounder data

Tobias Ziolkowski [a,*], Agnes Koschmider [b], Colin W. Devey [a]

[a] GEOMAR - Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
[b] University of Bayreuth, Bayreuth, Germany

ARTICLE INFO

Dataset link: https://doi.pangaea.de/10.1594/PANGAEA.918716

Keywords:
Multibeam
Data science
Bathymetric outlier removal
Seafloor mapping

ABSTRACT

Multibeam sonar data are a valuable tool for seafloor mapping and geological studies. However, the presence of outliers in multibeam data can distort the results of analyses and reduce the accuracy of seafloor maps. In this paper, we define a weighting function based on the performance of various outlier detection techniques (ODTs) for detecting outliers in multibeam data, which calculates an outlier probability score for each sounding. Our results show that each ODT has its own strengths and weaknesses, and that a combination of outlier detection techniques is promising to improve reproducibility, explainability and the accuracy of the detection process. To address the challenge of detecting outliers in multibeam data, we propose a weighted outlier detection function that outperforms individual outlier detection techniques in terms of precision, recall and F1 scores by considering their strengths and combining them in a way that accounts for variations in the data. The function detects various types of outliers with high precision and recall values, resulting in valuable improvements in outlier detection performance for multibeam data. Overall, our proposed workflow has the potential to significantly improve the way multibeam data cleaning is performed, with the weighted outlier detection function being applied first, detecting most of the outlier automatically, followed by a domain-expert review of a small group of soundings whose automatic outlier labeling is not unequivocal.

1. Introduction

Acoustic measurements of seafloor depth are essential for oceanographic research, providing insights into the geomorphology and geology of the seafloor. The advancement of multibeam echo-sounder (MBES) technology (Farr, 1980) has enabled the acquisition of large quantities of high-resolution depth data. The multibeam system works by emitting a fan-shaped acoustic beam (or "ping") in a cross-track direction and measuring the time it takes for the echo to return from the seafloor. By measuring the round-trip time of the signal and assuming a speed of sound propagation in water, the system can calculate the depth of the ocean at various cross-track distances along the beam. The cross-track distance refers to the perpendicular distance between a target (such as the seafloor) and the track of the survey vessel.

The survey vessel generally cruises in a raster pattern, gradually insonifying a large area of seafloor. However, the complex marine environment, including stratified water bodies, organisms in the water column, and ship's noise, can result in spurious depth measurements (either due to false echo detection or variations in the actual speed of sound in water) termed outliers, see Fig. 1 (left). Outliers, with calculated depths significantly and erroneously different from other data points in the data set, can degrade the accuracy of the final map and potentially lead to false conclusions (Enderlein and Hawkins, 1987). Traditionally, outliers have been detected and flagged manually, and various software tools are available for such "editing" work, resulting in cleaned data, see Fig. 1 (right). This process, however, is inherently subjective and prone to human error.

While various automatic outlier detection techniques have been developed, their comprehensive application in the context of multibeam data, a critical aspect of geostatistical modeling, has not yet been fully exploited. Machine learning-based outlier detection techniques, such as those explored in Yang et al. (2022), Lirakis and Bongiovanni (2000), Lu et al. (2010), have shown good precision and recall scores in detecting outliers in multibeam data. These techniques, however, may vary in performance depending on the specific attributes of the data and the nature of the outliers. Some algorithms are particularly adept at handling typical outlier scenarios, while others show promise in more complex, multi-dimensional scenarios. This variation can lead to different algorithms misinterpreting the relationship between the soundings and the seafloor. The evolving role of these machine learning techniques in geostatistical modeling highlights their significant impact on improving the accuracy and precision of data modeling. This diversity in performance emphasizes the need for a customized approach to outlier detection within geostatistical modeling, reinforcing the concept that no single method can be universally applied to all outlier scenarios.

* Corresponding author.
E-mail address: tziolkowski@geomar.de (T. Ziolkowski).

https://doi.org/10.1016/j.cageo.2024.105572
Received 21 September 2023; Received in revised form 7 March 2024; Accepted 7 March 2024
Available online 11 March 2024

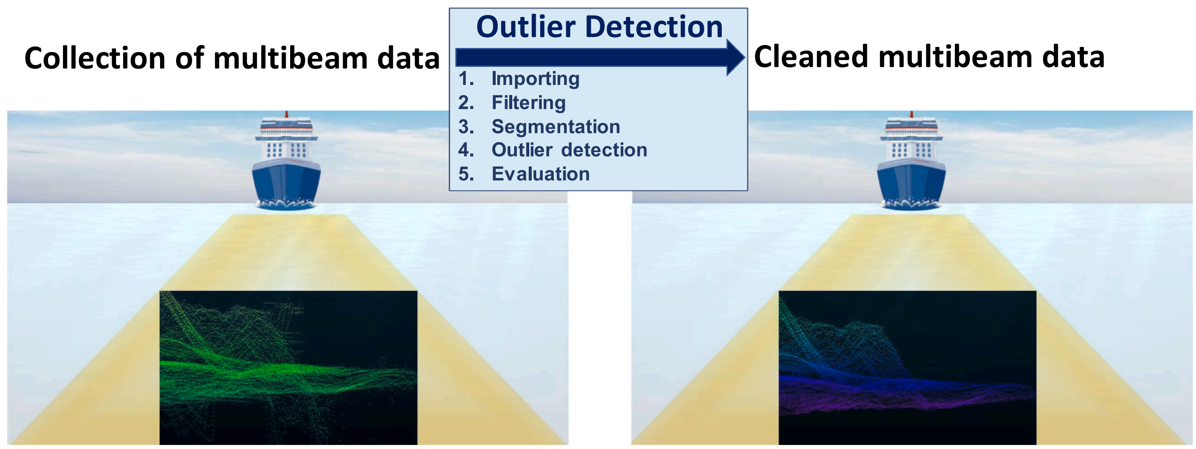


Fig. 1. Comparison of raw and cleaned multibeam data: Raw data (left) vs cleaned data (right), revealing a previously obscured seamount. Raw multibeam data with quality issues is processed using filtering, segmentation, and outlier detection techniques (ODT) resulting in cleaned multibeam data.

In this study, we apply data analytics techniques to acoustic measurements of seafloor depth to detect outliers, comparing the performance of statistical tests, distance- and density-based methods, and machine learning algorithms with manual editing. We find that no single approach can be applied to an entire multibeam data set while ensuring reproducibility, explainability, and accurate results. Instead, we propose a new weighting function for case-specific outlier detection, which considers the strengths of each individual technique and combines them in a way that accounts for variations in the data. This approach is shown to outperform individual automatic techniques in terms of precision, recall, and F1 scores.

The paper is structured as follows: Section 2 sets the stage by discussing data quality in multibeam systems and the current practices in manual data cleaning. It also includes a review of related work, providing a backdrop for our novel approach. This is followed in Section 3 by a detailed exposition of the algorithms and implementation strategies we employed. Section 4 presents our results and Section 5 interprets and discusses these findings in the context of their broader implications and limitations. Finally, Section 6 concludes the paper with a summary of our key contributions.

2. Methodology

2.1. Data quality in multibeam data

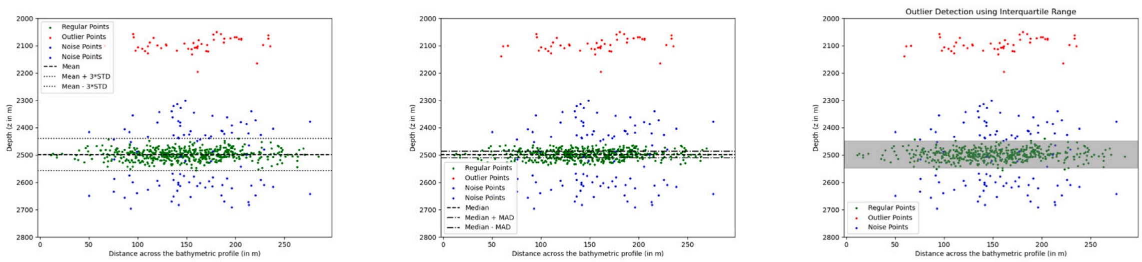
The terms “outlier” and “noise” in multibeam data are often used interchangeably but in fact have different meanings. Aggarwal and Yu (2001) distinguish between two types of false data: those that are considered noise are points lying outside a set of defined clusters, and those that are defined as outliers are points lying outside of the set of clusters but also separated from the noise. This is shown in Fig. 2, which is a plot of calculated depth (y-axis) against cross-profile distance (x-axis). The seafloor is flat and lies at 2500 m depth. The limiting precision of an acoustic depth measurement is determined by the physics of sound in water and is usually quoted by sonar manufacturers as +/- 1% of water depth. The green dots in Fig. 2 are randomly distributed with this precision around the depth 2500 m and are accurate measurements of the true seafloor depth. In this example, any deviation from this distribution is false and needs to be detected and flagged. The blue dots in Fig. 2 represent noise. Outliers are shown as red dots. Two types of outlier are distinguished: “Local outliers”, also known as point anomalies, are data points which differ significantly from their neighboring points within a small region of the data set (Souiden et al., 2017). They can be detected using methods that measure the density of soundings in a given region and identify points with significantly lower density than their neighbors. Local outliers are often caused by measurement errors or data entry mistakes. “Global outliers”, on the other hand, are data points that deviate significantly from typical values in the entire data set and can be detected using methods that measure the deviation of a sounding from the mean or median of the entire data set (Knorr and Ng, 1998). Global outliers are often caused by systematic errors or other factors that affect the entire data set, such as a malfunctioning sensor.

2.2. Manual data cleaning

Traditionally, multibeam data have been preprocessed (“cleaned”, “edited”) by scientists or students using manual, visual methods. Several software tools are available for this, both open-source resources such as MBSystem (Monterey Bay Aquarium Research Institute (MBARI), 2023) and commercial packages such as Qimera (QPS, 2023) or Caris (Caris, 2023). Most of these packages include a simple filtering algorithm to exclude impossible values (negative depths or depths greater than 11.000 m which do not exist on Earth) and then a graphical representation of the depth data collected from either a sequence of successive pings or a selected region of the seafloor. Using a mouse and various selection tools, one or more soundings can be selected and flagged as “excluded”. This label generally implies that the user does not believe them to be real measurements of the seafloor depth. This process is prone to several sources of bias or error and is inherently subjective. The user’s ability to identify outliers is strongly dependent on their level of experience and especially the number of seafloor maps they have studied and their understanding of seafloor processes — a region of the seafloor showing huge differences in water depth between adjacent measurements can be an outlier to be excluded for one user and a geologically significant structure for another. The people processing the data get tired, make mistakes or, if working at sea, can be affected by motion sickness, all of which contribute to poor and/or inconsistent data editing.

2.3. Related work

Statistical and Filtering-Based Methods: Statistical measures and filters have been at the forefront of outlier detection in multibeam data for quite some time. Techniques that deploy statistical analysis, often combined with filtering, have shown promising results in outlier detection. For instance, the work by Šiljeg et al. (2022) introduced CloudCompare, leveraging the Point Cloud Library (PCL) for filtering multibeam echo sounder (MBES) point cloud data. The cascading filter technique based on cross track distance by Sedaghat et al. (2013), and the segmentation-based methodology presented by Ware et al. (1990) provide further testament to the effectiveness of statistical measures.



(a) Visualization of outlier detection using standard deviation. The mean value is represented by a solid line (—), and the mean ± 3 standard deviations are represented by dotted lines (....).

(b) Visualization of outlier detection using median absolute deviation. The median is represented by a solid line (—), and the median ± MAD are represented by dashed lines (-.-).

(c) Visualization of outlier detection using interquartile range. The gray shaded region represents the bounds for outliers, with any data point outside of these bounds considered an outlier.

Fig. 2. Visualization of different methods for outlier detection in multibeam data.

Other notable approaches include outlier detection by computing differences in mean values among pings in 3 × 3 grids as presented by Bourillet et al. (1996) and the innovative approach by Ware et al. (2018) that classifies soundings into categories based on weighted mean and standard deviation depths.

Density and Nearest-Neighbor Based Methods: Recognizing patterns and relationships within data is crucial for outlier detection. Techniques centered on point density and nearest-neighbor relationships have proven effective for this cause. Yang et al. (2007), for instance, took a leap with their algorithm that detects outliers by examining point density in multibeam data. Similarly, CUBE introduced by Calder and Mayer (2003) brings forth a KALMAN filter, focussed on estimating the uncertainty of each sounding and passing the depth information to its vicinity. Relevant contributions include the studies by Guenther and Green (1982), Grim (1988), and Herlihy et al. (2018), which have presented improved versions of the Combined Offline Processing (COP) program and prefiltering procedures.

Machine Learning and Interpretability Approaches: The surge of machine learning, particularly the utilization of Convolutional Neural Networks (CNN) for multibeam data, has ushered in a new era of possibilities. A quintessential example is the work by Stephens et al. (2020) that achieved 97% classification accuracy using a 3D CNN model for noise removal in multibeam data. However, the lack of interpretability and reproducibility of such algorithms has often been their Achilles' heel. Efforts have been made to confront this limitation, with (Hughes Clarke, 2003) shedding light on the imperfections in MBES system integration, and Panjei et al. (2022) emphasizing the paramount importance of providing explanations for detected outliers. As we navigate through these methods, it becomes evident that the constant evolution and amalgamation of techniques might pave the way for more refined and accurate outlier detection in multibeam data.

A study from Zhou et al. on forward-looking sonar (Zhou et al., 2022) underscores the use of combined clustering techniques, like FCM and k-means, augmented by the PCNN for enhanced underwater target delineation. Additionally, with the evolution of MBES capabilities, there is a heightened emphasis on data processing automation, as highlighted in a recent research article from Le Deunf et al. (2019). This paper provides a retrospective on historical techniques while pointing towards the promising role of machine learning in refining bathymetric data.

Building upon these foundational works, our methodology presents a holistic solution for multibeam data outlier detection. By integrating both traditional and contemporary techniques, our approach addresses prevalent challenges such as the complexity and variability in the density and distribution of multibeam data and the ever-intricate task of balancing precision with explainability. Our system uniquely positions itself in the niche, offering enhanced accuracy and streamlined processing, paving the way for more robust geospatial data analyses in marine research.

2.4. Our approach

Our aim is to ascertain whether automated, data-analysis-based methods can equal or possibly exceed the performance of traditional manual data cleaning techniques both in terms of speed and accuracy. Using objective data analysis also allows reproducible and explainable techniques to be employed. We tested various data-analysis approaches on a data set which had previously been manually edited, comparing the manual flagging of outliers with the data-analysis results. The test multibeam bathymetry data were collected during the cruise MSM88 of RV MARIA S. MERIAN in the Atlantic (Devey et al., 2020) using a Kongsberg EM 122 multibeam system, which uses signal processing and beam forming techniques to measure the depth at 432 discrete locations on the seafloor from one acoustic ping. In total, depth data over an area of 153.121 square kilometers were collected (see Fig. 6 for more details), resulting in approximately 85.000.000 individual depth soundings. The area stretches from the edge of the Cabo Verde Exclusive Economic Zone (EEZ) in the East to the EEZs of Guadaloupe, Dominica and Martinique in the West and so covers a wide range of Atlantic seabed morphologies. These include flat sedimented plains, seamounts, smaller ridges and fracture zones and the Mid-Atlantic Ridge (MAR), which has shaped and still shapes the seafloor in the Atlantic Ocean.

We employed various outlier detection techniques, as categorized by Aggarwal et al. [1]:

Statistic-based techniques: These use statistical tests and metrics to detect outliers, such as the standard deviation (STD), median absolute deviation (MAD), and interquartile range (IQR) (Howell, 2005; Clark-Carter, 2005). The STD method compares soundings to the mean and flags them as outliers if they are a certain number of standard deviations away from it. In our approach we used three times the standard deviation. The MAD method identifies potential outliers by comparing the deviation of each data point from the median. Leys et al. (2013) propose the MAD as a new method for outlier detection, pointing out weaknesses in the STD method. The IQR method takes the difference between the upper and lower quartile of a data set and flags all soundings as outliers that fall outside of this range.

Distance- and Density-based outlier detection techniques: Distance-based techniques use a distance metric to identify outliers by comparing the distance between a given sounding and nearby soundings.

Initially, to standardize our data set, we transformed the raw latitude and longitude coordinates from the Universal Transverse Mercator (UTM) projection in Zone 26 to World Geodetic System 1984 (WGS84) coordinates. Subsequently, we computed the Euclidean distances between individual soundings within our data set, effectively capturing the spatial proximity of each sounding to its neighbors. To evaluate the immediate neighborhood of each sounding, we defined a radius and converted this radius into decimal degrees, assuming a conversion

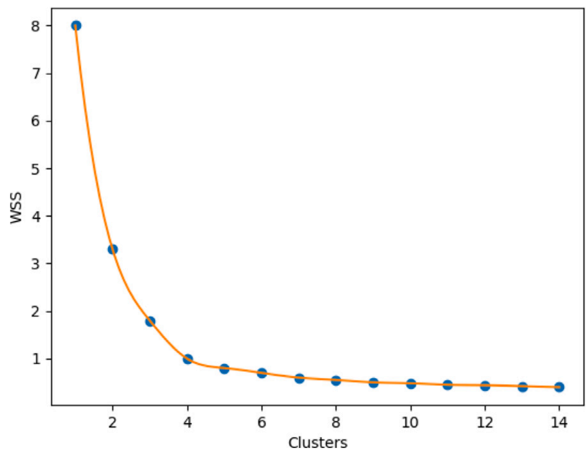


Fig. 3. Determination of optimal number of clusters for k-means clustering using the elbow method.

factor that equates one degree to 111.32 kilometers at the equator. This conversion was crucial for consistency, as our distance calculations were based on latitude and longitude coordinates.

As an example for a distance-based ODT, Zhang et al. (2009) propose a Local Distance-based Outlier Factor (LDOF) to measure the "outlierness" of objects in scattered data sets. Density-based techniques identify outliers as those with a lower density of nearby points compared to non-outlying observations. To analyze the structure and patterns within our data, we employed the k-means clustering method (Ahmed et al., 2020). This technique groups similar data points into clusters based on their proximity to each other. However, determining the optimal number of clusters is crucial for obtaining meaningful results. To identify the optimal number of clusters, we applied the k-means clustering method to the data using the elbow method (Syakur et al., 2018). The elbow point was identified at K=4, as shown in Fig. 3. Next, we applied the k-means clustering method to the full data using K=4 clusters. To identify outliers in the data, we calculated the distance from each data point to the center of its assigned cluster. To select the actual outliers, we calculated the median k-means distance and marked soundings with a k-means distance greater than the median as outliers.

Recent studies have corroborated the effectiveness of these techniques in multibeam bathymetric data. Wei et al. (2021) demonstrated the successful application of a clustering algorithm for automatic cleaning of outliers in multibeam bathymetric data (Wei et al., 2021). Similarly, a study by Wang et al. (2023) combined uncertainty and density clustering methods to effectively filter outliers in multi-beam bathymetry (Wang et al., 2023). Moreover, Michel et al. (2021) developed the ToMATo algorithm, which utilized clustering and topological persistence approaches for outlier detection in multibeam data, and importantly, their study included a comparison of this algorithm with the DBSCAN and LOF algorithms, showcasing its efficacy in different scenarios (Michel et al., 2021).

In addition to k-means clustering, we tested the LOF (Alghushairy et al., 2021) and DBSCAN (Hahsler et al., 2019) algorithms for detecting outliers in multibeam data. However, determining optimal parameters for these algorithms can be challenging. For LOF, we tested various combinations of the number of neighbors and degree of contamination and found the optimal parameters to be 82 neighbors and a degree of contamination of 0.1. For DBSCAN, we set min samples to 6 and used the nearest-neighbor method to determine that an eps value of 0.1 was optimal.

Machine-learning techniques: Studies have shown that applying machine learning algorithms, particularly deep learning approaches like multi-layered artificial neural networks, has great potential in analyzing multibeam data (Cun et al., 1997; Jain and Seung, 2008; Krizhevsky et al., 2012). We utilized three different supervised machine learning models: Multi Layer Perceptron (MLP) (Gardner and Dorling, 1998), Spatial Logistic Regression (LR) (Rahmatullah Imon and Hadi, 2008), and Random Forest (RF) (Breiman, 2001).

To ensure comparability, we used the same set of features for the training process of all three models, including latitude, longitude, depth and the following calculated features. We first clustered all soundings and calculated the local neighborhood for each sounding in a 100 m radius. We then calculated the standard deviation, mean depth, and normalized distance to mean depth in the local neighborhood. In the MLP model, we utilized a logistic activation function with a cross-entropy error function optimized by the SGD method (ichi Amari, 1993). We set the number of epochs to 50, with early stopping if the loss function did not improve by more than 0.002 over three epochs. The dropout parameter was fixed at 0.25, and the learning rate was adaptive. The MLP model consisted of three hidden layers, each with 32 perceptrons, and utilized the ReLu activation function for the hidden layer. The batch size was set to 32. For the RF model, we evaluated various parameters and determined that the best configuration was achieved with 40 decision trees with a maximum depth of 5.

In our analytical model development, we adopted a geospatial approach for data segmentation, dividing input data by geographical coordinates to differentiate between training and validation sets. Specifically, data between 31 and 35°W (see Fig. 6) were allocated to the validation set, with the remainder used for training. We further refined this division using scikit-learn's train_test_split function, assigning 80% of the training data for model training and 20% for internal evaluation. This geographically informed segmentation strategy ensures a rigorous evaluation of our model's predictive performance, as discussed in Section 5, by testing on entirely new geographical areas.

We compared the performance of the different techniques outlined above against the conventional manual approach, considering technical implementation and outcomes. By segmenting the calculated features into ranges, we identify feature ranges where the efficiency of the techniques varies, and provide recommendations on the optimal utilization of these methods. We also investigate the attributes that exert a more pronounced influence on the efficiency of outlier detection techniques. To evaluate the techniques, we use precision, recall, and F1 score, as recommended by Alimohammadi and Nancy Chen (2022) and Caroline Cynthia and Thomas George (2021).

3. Algorithm and implementation

The complexity and variability in the density and distribution of multibeam data can pose challenges for selecting the most suitable algorithm for automatic outlier detection in any particular setting. Our approach to dealing with this is shown in Fig. 5 and involves 6 steps:

1. Importing multibeam data and extracting latitude, longitude, and depth values.
2. Filtering, segmenting, and transforming the data involve recalculating the coordinates to UTM and calculating additional attributes, such as local neighborhood and standard deviation, depending on the outlier detection technique used. The goal is to prepare the data for algorithmic outlier detection.
3. Applying the outlier detection techniques presented in Section 2.4 to the prepared data.
4. Evaluating the performance of the applied outlier detection techniques using the standard metrics precision, recall and F1 score.
5. Constructing a weighting function based on the best performing outlier detection technique for each range of each attribute, using the outlier label and the F1 score.
6. Evaluating the defined weighting function using the same performance metrics as in Step 4. This step helps to verify the efficiency of the weighting function and determines whether it leads to improved outlier detection accuracy.

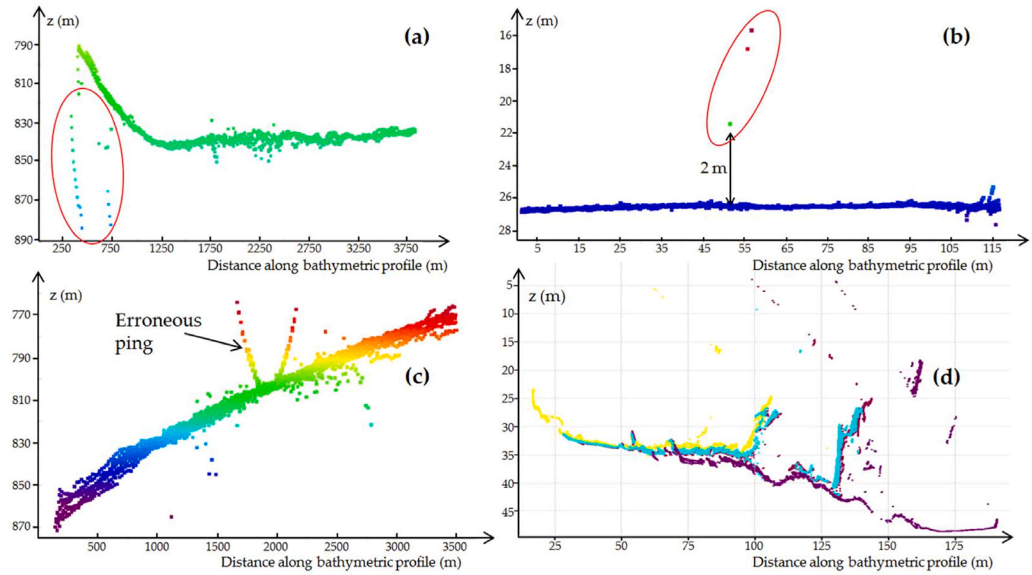


Fig. 4. Four examples of data quality issues in multibeam data. For cases (a) to (c) soundings are colored according to depth, for case (d) the colors represent different survey lines (Le Deunf et al., 2020).

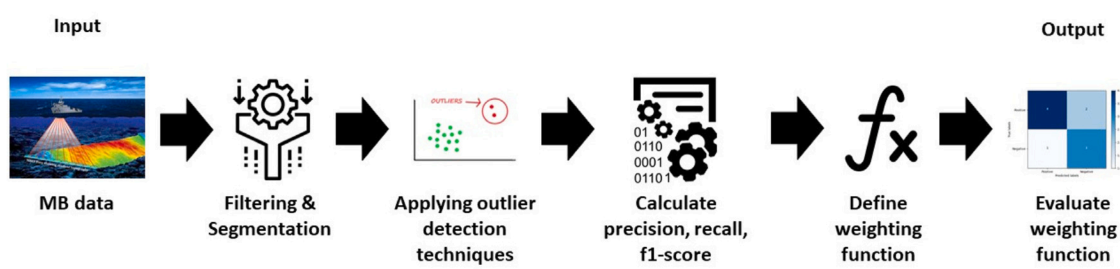


Fig. 5. Data-driven pipeline for identifying outliers in multibeam data. The process involves importing the data, filtering and segmenting it, applying outlier detection techniques, calculation of the evaluation metrics precision, recall and F1 score, defining a weighting function and evaluating the defined function.

To determine the suitability of the individual outlier detection techniques for different types of spurious data we assessed their performance in 4 typical real-world "use cases". These cases are illustrated by raw data examples in Fig. 4 and comprise:

- *Case (a)*: This is the simplest case. Erroneous soundings are located below the seabed as defined by the vast majority of data points.
- *Case (b)*: A few isolated erroneous soundings located far from the vast majority of the other soundings which themselves are tightly clustered (in this case showing a flat seabed).
- *Case (c)*: A group of erroneous soundings, connected to the seafloor, containing a larger number of data points. These soundings all belong to the same ping and this ping produced significantly different data from adjacent pings.
- *Case (d)*: Relatively noisy data acquired on a shallow and rough seafloor. Particular care must be taken to ensure that the bathymetric features are preserved.

4. Results

Table 1 summarizes the results of all applied outlier detection techniques and shows their suitability for each outlier case.

The consolidated metrics, reflecting the average values across all four outlier cases, demonstrate the efficacy of each ODT on the 20% segment designated for evaluating the models on the training data. This evaluation was conducted while maintaining the 80% of the data set solely for training and parameter optimization. Furthermore, Fig. 7 complements these metrics by visually depicting the cleaned data post-application of each Outlier Detection Technique (ODT), displaying only the data points not identified as outliers. This figure demonstrates the test data (upper left panel labeled "Input Data") derived from the geographic segmentation within longitudes −31 to −35, employing a color gradient from blue to yellow to represent depth transitions from deep to shallow regions.

This panel in Fig. 7 demonstrates the data set's complexity, encompassing both flat terrains and varied elevations, with outliers highlighted as anomalies. Importantly, we have highlighted the five outlier types in the input image to provide a clear visual representation of these anomalies, categorized into five types:

1. Uniformly spaced parallel lines with darker colors across the section, indicative of systematic anomalies.
2. Dark spots among these lines, signaling unusual depth variations.
3. Dark spots on the deep-sea plateau, especially between longitudes −34.8 to −33.8 and latitudes 14.81 to 14.69, denoting unexpected depth changes.
4. More dark spots between longitudes −32.9 to −32.3 and latitudes 14.6 to 14.48, indicating further anomalies.
5. A significant outlier group at longitude −34.3 and latitude 14.8 on the plateau's edge.

These outlier types are critical for assessing the efficacy of ODT algorithms in automatically detecting and cleaning the data.

Table 1
Comparison of all applied outlier detection methods in multibeam data: Precision, recall, F1 Score, and suitable cases with reasoning.

Method	Precision	Recall	F1 score	Cases	Reasoning
STD	0.552	0.006	0.012	a, b	Widely used, but can be sensitive to the presence of outliers, Effective in uniform areas, sensitive due to reliance on mean depth values.
MAD	0.508	0.141	0.221	a, b, (c)	Robust, using median depth; less affected by extreme outliers.
IQR	0.553	0.051	0.092	a, b	Insensitive to extremes, focuses on central data spread.
k-means	0.405	0.574	0.476	a, b, (c)	Can handle multi-dimensional data, and can detect isolated soundings based on their lower density compared to the other instances within the same swath. Clusters data, identifying outliers as distant from cluster centroids.
DBSCAN	0.712	0.041	0.076	a,b, (c)	DBSCAN is effective in handling multi-dimensional data and complex clusters, and can detect instances located below the seabed as outliers if they do not belong to any close cluster in the feature space. Forms clusters based on density; effective in dense soundings.
LOF	0.354	0.101	0.157	a,b	LOF can detect instances located below the seabed and isolated soundings located far from the flat seabed as outliers based on their higher distance to nearby points. Detects outliers based on local density deviation from neighbors.
LR	0.690	0.440	0.530	a, b, c	Interpretable, efficient for large data sets but sensitive to irrelevant features and limited in capturing complex patterns.
RF	0.740	0.650	0.690	a, b, d	Suitable for identifying acceptable soundings with important features. Utilizes ensemble decision trees; effective in diverse, noisy data.
MLP	0.770	0.720	0.750	a, b, c, d	Can handle complex data but requires a large volume of labeled data for training. Suitable for linearly separable data, identifying isolated soundings and soundings connected to the seafloor. Neural network approach; suitable for complex pattern recognition

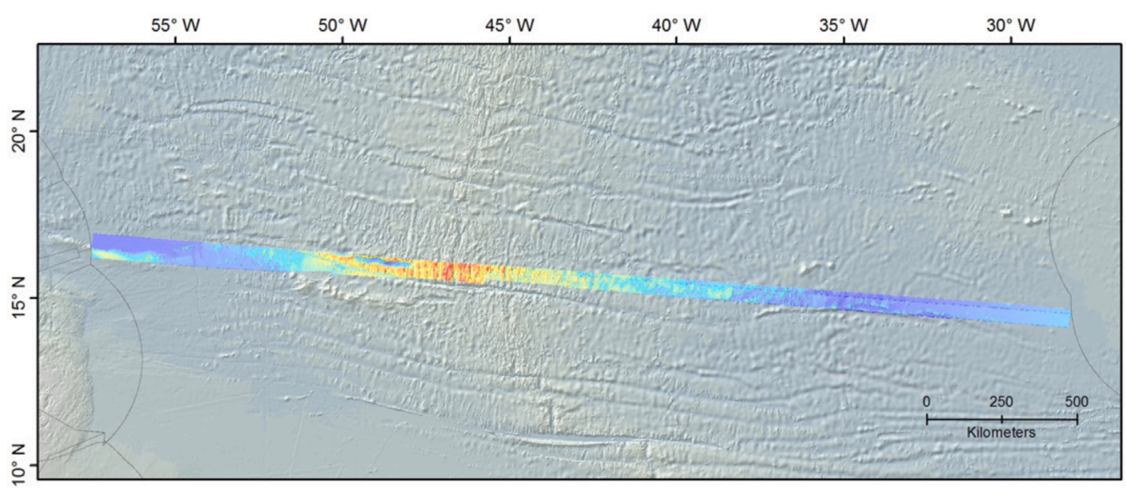


Fig. 6. Work area is situated between 14° and 17° North and stretches from the exclusive economic zone (EEZ) of Cabo Verde in the east to the Guadeloupean and Martinican EEZ in the west and covers only international waters. Cruise track during cruise MSM88 (Devey et al., 2020).

The second image in the first row demonstrates the manually cleaned data, while the third image shows the cleaned data using our weighted function, details of which are discussed in Section 5.

The standard deviation method (STD) shows limited effectiveness in detecting outliers in multibeam data, with evaluation results of 0.552, 0.006, and 0.012 for precision, recall, and F1 score, respectively. The method exhibits a moderate precision and misses many actual outliers (low recall), resulting in a low F1 score. The standard deviation method is sensitive to the proportion and extremity of outliers in the data set and is suitable for Cases a and b where outliers are relatively easy to identify based on their location with respect to the seabed (see Fig. 4). It can effectively detect outliers in flat seafloor areas where the strict threshold is able to differentiate between outliers and the expected seafloor. In noisy data and rough seafloor areas the method can potentially flag real bathymetric features as outliers, leading to a loss of valuable data and potentially erroneous interpretation of seafloor structure.

The median absolute deviation (MAD) method demonstrates moderate performance in detecting outliers, with a precision of 0.508, a recall of 0.141, and an F1 score of 0.221. This is slightly better than STD. MAD is generally considered to be a more robust alternative to standard deviation, as it is less sensitive to the influence of extreme values. The method is particularly suitable for Cases a and b and is effective in handling noisy data with a high density of outliers. However, MAD may misclassify valid measurements as outliers in rough seafloor areas, which is a notable limitation.

The interquartile range method (IQR) achieves a precision of 0.553, a recall of 0.051, and an F1 score of 0.092, which are lower than those of the MAD, but slightly higher than those of the STD. Despite its limitations, the IQR is generally robust and insensitive to extreme

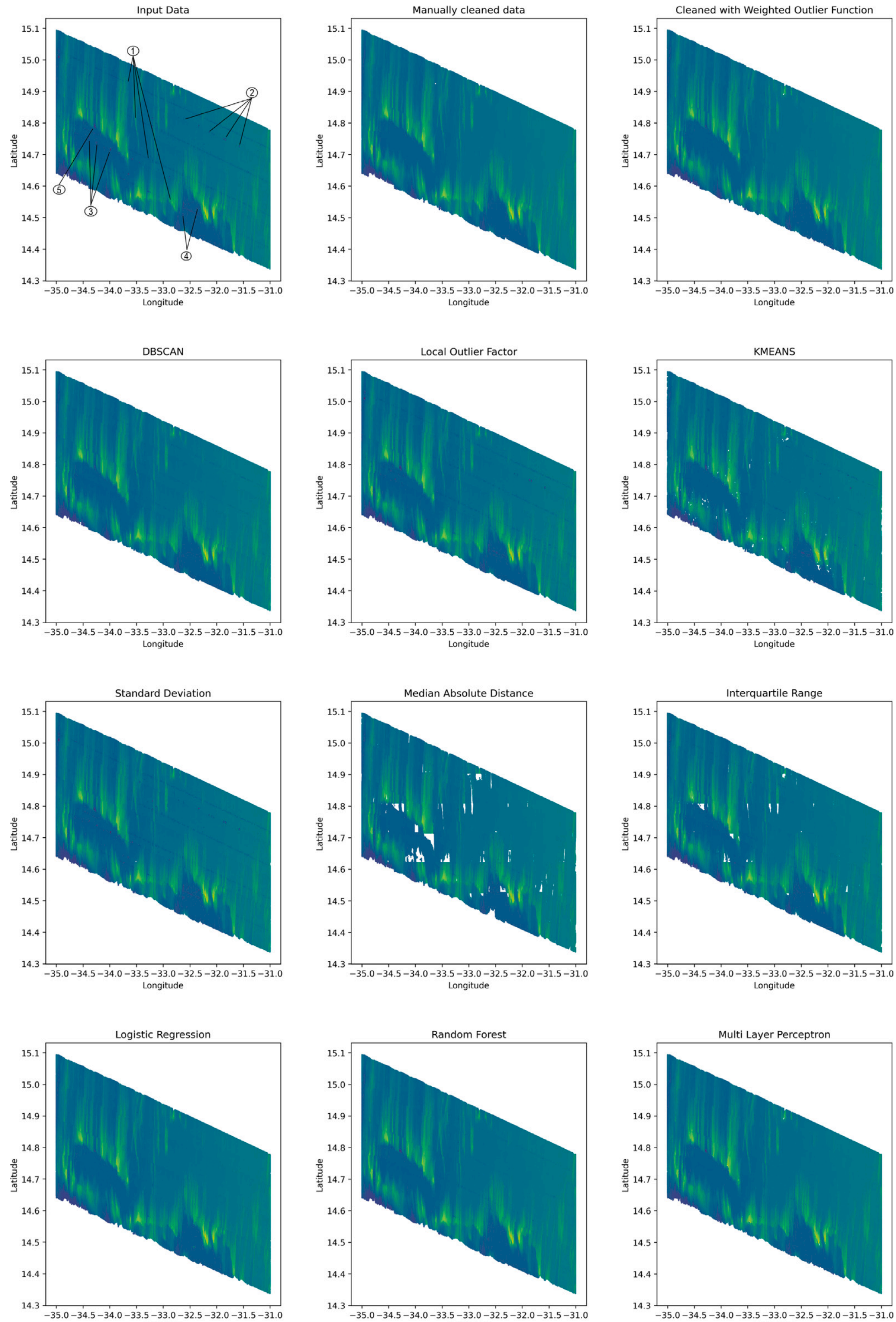


Fig. 7. Detailed comparative visualization of all applied Outlier Detection Techniques showing the cleaned multibeam data and displaying 'Input Data', which consists of all data points, 'Manually cleaned data', which shows the data cleaned by a domain expert and 'Cleaned with Weighted Outlier Function', which shows the cleaned data after applying our proposed weighting function, which will be discussed in Section 5. This figure highlights the distinct impact of each method on identifying and removing outliers, effectively illustrating their varying degrees of precision and recall. Regions labelled 1-5 on the Input Data panel refer to the five outlier types present in this region as discussed in the text.

values. The method is suitable for Cases a and b and has a strength in detecting extreme values. However, the method has a weakness in rough seafloor areas and data that is highly contaminated by outliers, leading to the weak evaluation scores for the selected study area.

STD method's low recall suggests it frequently overlooks true outliers, leading to their retention within the data set. This cautious approach minimizes outlier identification, leaving most outliers from categories 1–5 unaddressed, as evidenced by Fig. 7. Although some data points are overly removed, resulting in occasional white spots, the bulk of outliers persists.

In contrast, the Median Absolute Deviation (MAD) method tends to excessively flag data points as outliers, creating noticeable gaps or 'holes' in the data representation. This aggressive strategy effectively identifies outliers in categories 1–3 but leaves some from category 4 and extensively removes those in category 5, causing significant data voids.

The Interquartile Range (IQR) method, while more conservative than MAD, still introduces notable gaps, suggesting the removal of many valid data points. It achieves a balance with slightly higher precision but lower recall than MAD, effectively detecting outliers in categories 1 and 2, with persistent visibility over the second plateau and remaining outliers in categories 3 and 4. The aggressive removal in category 5 leads to pronounced white spaces.

These findings illustrate the challenges in using statistical-based methods for outlier detection in multibeam echosounder data, especially in maintaining a balance between eliminating true outliers and retaining genuine seafloor features. Fig. 7 visually substantiates the varied effects of STD, MAD, and IQR methods on data integrity, underlining their differential precision and recall impacts.

The low precision, recall, and F1 score of the three statistical methods when used with our data are attributed to their application to a noisy real-world data set that includes both flat and rough seafloor areas. Statistical outlier detection methods depend on a preset threshold to classify data points as outliers, resulting in high accuracy in identifying outliers in flat seafloor environments. Nevertheless, this fixed threshold approach is less effective in detecting outliers in environments with variable and rough seafloor shapes.

For the distance- and density-based algorithms, the k-means clustering algorithm showed low precision (0.405) and moderate recall (0.574), resulting in a moderate F1 score (0.476). DBSCAN showed higher precision (0.712) and lower recall (0.041), resulting in a low F1 score (0.076). The LOF method showed the lowest precision (0.354) and a slightly higher recall (0.101), resulting in an F1 score of 0.157.

For outlier Cases a and b, the k-means technique is suitable since it can identify distinctive features of erroneous soundings. However, it may struggle to capture complex relationships between groups of soundings and the seabed in Case c. k-means is strong at identifying outlier groups distant from the seafloor, but it assumes soundings within each cluster are homogeneous and have a spherical distribution, which may not be true in noisy and rough seafloor environments.

DBSCAN works well with complex clusters and multi-dimensional data. It can identify outliers located below the seabed in Cases a and b. However, it may misidentify a group of connected soundings with many samples as a normal instance rather than an outlier. DBSCAN is not ideal for Case d due to the noise present in the data set that may affect density-based calculations.

The LOF technique is suitable for Cases a and b, detecting far-off outliers based on their lower density relative to the surrounding soundings. However, it may not accurately detect outlier groups with a larger number of samples, and noise in the data set could affect the density-based calculations.

Fig. 7 highlights the k-means clustering algorithm's tendency for overly aggressive outlier removal, evidenced by significant data voids. It also evaluates the nuanced effectiveness of each Optical Digital Terrain (ODT) method against specific outlier types identified in our data set.

The DBSCAN algorithm achieves partial success in outlier detection, notably struggling with the regular parallel lines (outlier type 1) while effectively identifying most dark spots between lines (outlier type 2) and the majority of outliers on the deep-sea plateau (outliers type 3 and 4). However, it falls short of fully recognizing the significant outlier group at the plateau's edge (outlier type 5), leaving some outliers undetected.

Conversely, the Local Outlier Factor (LOF) method shows limited effectiveness, detecting few outliers within the parallel lines (outlier type 1) and only a subset of dark spots (outlier type 2). Its performance is modest in the deep-sea plateau areas (outliers type 3 and 4), and it fails to adequately address the significant outlier cluster (outlier type 5), which remains clearly visible.

The k-means algorithm's approach to outlier detection is marked by its overgeneralization, leading to the flagging of overly large regions as outliers, hence creating data voids. This indicates an imprecise strategy, particularly in handling outliers among parallel lines (outlier type 1) and in the deep-sea plateau regions (outliers type 3, 4, and 5), thereby compromising data integrity through excessive cleaning.

In summary, the suitability of each density- and distance-based technique depends on the outlier case. Distance-based methods are efficient and scalable, while density-based methods can handle complex data. However, when choosing the appropriate method, it is crucial to consider the sensitivity of the parameter choices and the complexity of the outlier groups.

Based on the evaluation results presented in Table 1 and shown in Fig. 7, the machine learning-based outlier detection techniques showed good precision and recall scores in detecting outliers in multibeam data. However, their performance may vary depending on the specific outlier case. Logistic Regression (LR) is a computationally efficient model suitable for identifying outliers in Cases a and b, but its performance may not be as good in Cases c and d due to the non-linear relationships between features and target variables. It is a good choice for users who value interpretability and ease of implementation. However, LR may be sensitive to irrelevant features. The evaluated metrics showed that LR achieved a precision of 0.690, a recall of 0.440, and an F1 score of 0.530. Random Forest (RF) may be suitable for Cases a and b, but it may not capture the relationship between the soundings and the seafloor in Case c. However, it could still be used in Case d with caution to ensure that bathymetric features are preserved. Based on the evaluated metrics, RF achieved a precision of 0.740, a recall of 0.650, and an F1 score of 0.690. The evaluation results for the MLP demonstrate that it outperforms all applied ODTs in detecting outliers in multibeam data. MLP is suitable for detecting outliers in all four outlier cases. Its strengths include its ability to learn non-linear relationships and handle high-dimensional data. Based on the evaluated metrics, the MLP achieved a precision of 0.770, recall of 0.720, and an F1 score of 0.750.

Logistic Regression (LR) shows high effectiveness in outlier detection, accurately pinpointing the majority of outliers in categories 1–4 without significant data voids. It leaves a few anomalies, such as at (−33.25, 14.6) and (−32.49, 14.51), with the category 5 cluster partially detected, indicating precise yet occasionally overzealous removal at specific sites like (−34, 14.74).

Random Forest (RF) masks linear outlier patterns well, diminishing the visibility of category 1 lines and effectively identifying anomalies in categories 2–4, with the category 5 cluster remaining noticeable. This highlights RF's broad detection capability, though it does not fully eliminate all anomaly types.

Multi-Layer Perceptron (MLP) mirrors LR and RF in efficacy, identifying outliers in categories 1–4 with minimal exceptions and occasionally over-deleting near edges, like around latitude 14.78. The notable category 5 cluster is also partially addressed, showcasing a selective sensitivity to complex outlier formations.

These machine learning approaches illustrate a delicate balance between accurately detecting outliers and preserving data, with LR, RF, and MLP offering distinct advantages in minimizing data loss while ensuring high precision. Their visualized performance suggests machine learning's potential to improve outlier detection in multibeam echosounder data, though refinement is needed to perfect data cleaning.

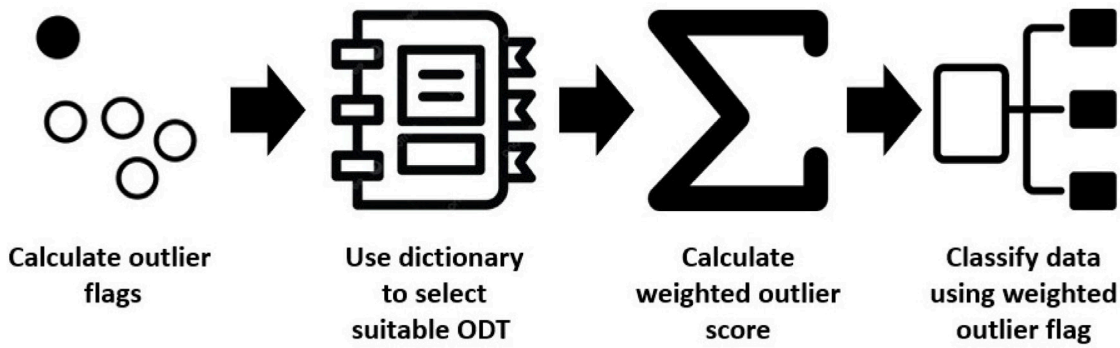


Fig. 8. The flow chart of the Weighted Outlier Detection Function demonstrates the systematic approach for identifying and classifying outliers in multibeam echo-sounder data: (1) Calculation of outlier flags for each data point using different Outlier Detection Techniques (ODTs); (2) Selection of suitable ODTs for each data point based on a predefined dictionary and specific attributes; (3) Computation of a weighted outlier score by multiplying flag values with the F1 score of the corresponding ODT; (4) Classification of data points using the weighted outlier flag, where the algorithm determines the final outlier status based on the accumulated scores.

5. Discussion

Our evaluation in the previous section has shown that there is no "one size fits all" ODT for all types of seafloor morphology and that different ODT perform better or worse (as shown by their F1 scores) in different situations. Therefore, we implemented a weighting function to improve the performance and robustness of outlier detection in multibeam data by selecting the most appropriate ODT for each data point considering the particular local situation of that point. The workflow of our weighted outlier detection function is illustrated in Fig. 8. Our weighting function accumulates the flagging results (0 = good data, 1 = outlier) across all ODT techniques for each point to determine the most appropriate ODT for that point relative to its local environment. The local situation of the point is characterized by calculating three derived attributes: "mean_depth_100m" (the mean of the depth values of all soundings within 100 m radius of the point being considered), "normalized_distance_100m" (the difference in depth between the point being considered and the average depth of all points within 100 m radius) and "std_deviation_depth_100m" (the standard deviation of the depths of all points within 100 m radius of the point being considered). To give more emphasis to flagging results from methods which have a good performance (high F1 score) in outlier detection, we multiply the flag value by the F1 score of that particular ODT. To prepare for the weighted outlier detection function, the three attributes were analyzed and segmented into different ranges, and the F1 score was calculated for each outlier detection technique for each range and the best-performing ODT for that particular situation identified and compiled in a dictionary (see Table 2 for detailed information). The resulting algorithm calculates a weighted score for each data point, indicating the likelihood that the point is an outlier, by performing the following steps:

1. Calculate the 0 = good data, 1 = outlier flags for the data point using each of the different ODT described above.
2. Based on the values of the three attributes "mean_depth_100m", "normalized_distance_100m", and "std_deviation_depth_100m" for the data point, the algorithm uses the pre-defined dictionary to determine which ODT should provide the flags (one for each attribute-designated ODT) for this point. This results in three flag values for each data point derived from the attribute-determined preferred ODTs.
3. The three flag values for the point are each multiplied by the F1 score (for the particular attribute range) for the ODT which provided the flag. This adds increased weight to the flags from better-performing ODT (higher F1 score) in the final function. Based on the range, the algorithm selects the outlier label and the corresponding F1 score of the best performing outlier detection technique for this specific range.
4. The algorithm then sums the results of all three techniques.
5. If the weighting score equals 0 (according to all three attributes, the best ODT all classified the point as good data and gave it a flag = 0), the 'weighted outlier' column is set to 0, which indicates that no outlier is detected.
6. If the weighting score is equal to the sum of the three F1 scores, the 'weighted outlier' label is set to 1, which indicates that an outlier is detected based on all three attributes
7. In all other cases, the 'weighted outlier' score is set to 2, indicating that the number of ODT that flagged the point as an outliers was greater than 0 but less than 3 and so the automatic outlier classification was not definitive and requires manual intervention by a domain expert.

The above steps are defined in the function 'calculate weighted score', which takes a row of data (all measured and calculated attributes of one depth measurement point) as input and returns the updated row with the 'weighted outlier' score and the sum of F1 scores for the row. The algorithm loops over each row of data and applies the 'calculate weighted score' function to each row, resulting in a weighted outlier score for each data point.

The following is a "worked example" for one loop through the algorithm in which a point with the following attributes is analyzed: a measured depth of −3700 m, a normalized distance to the mean depth in 100 m radius of 3 and a standard deviation of depth within a 100 m radius of 250:

1. The values of the three attributes for the point being considered are read from the data file. Each row in this data file contains all the attributes for a particular depth measurement and also the flag values generated by the different ODT for this depth measurement.
2. Using the dictionary "range dict" (which correspond to the second and third columns in Table 2) the ODT with the best performance for data points with these attributes is determined. For the data point being considered here, its mean_depth_100 m of −3700 lies in the range −4000 - −3500, making mlp_outlier the ODT of choice. For its normalized_distance_100 m of 3 and std_dev_depth_100 m of 250, the corresponding ODT are mlp_outlier and lr_outlier, respectively.
3. Having identified which ODT are relevant, their flags for this data point and the corresponding F1 score for the ODT in this particular attribute range (fourth column in Table 2) are extracted. For each attribute, flag and F1 score are multiplied and the sum of these three multiplications are added together to make a new attribute "weighted score". For the data point considered here, the best-performing ODT for each attribute returned a flag of 1 (the point is an outlier) which results

Table 2
Table showing the attribute, ranges, and corresponding ODT and F1 score for each range. The first column describes the attribute which was calculated. The second column shows the different ranges of this attribute, the third column shows the ODT with the highest F1 score, and the fourth column shows the reached F1 score. MLP is best suited for most of the ranges, except for high and low normalized distances and standard deviations > 200 m.

Attribute	Range	ODT	F1 score
mean_depth_100m	−6500 – −6000	mlp_outlier	0.7735
	−6000 – −5500	mlp_outlier	0.7356
	−5500 – −5000	mlp_outlier	0.7413
	−5000 – −4500	mlp_outlier	0.7774
	−4500 – −4000	kmeans_outlier	0.9101
	−4000 – −3500	mlp_outlier	0.7002
	−3500 – −3000	None	0.0
normalized_distance_100m	−20 – −14	db_outlier	1.0000
	−14 – −11	kmeans_outlier	1.0000
	−11 – −8	kmeans_outlier	1.0
	−8 – −5	mlp_outlier	1.0
	−5 – −2	mlp_outlier	0.9833
	−2 – 0	mlp_outlier	0.7525
	0 – 2	mlp_outlier	0.6953
	2 – 5	mlp_outlier	0.9790
	5 – 8	mlp_outlier	1.0
	8 – 11	db_outlier	1.0000
	11 – 23	kmeans_outlier	1.0000
std_dev_depth_100m	0 – 100	mlp_outlier	0.7405
	100 – 200	mlp_outlier	0.8137
	200 – 300	lr_outlier	0.8609
	300 – 400	lr_outlier	0.9217
	400 – 500	mad_outlier	0.9543
	500 – 600	lr_outlier	0.9892
	600 – inf	mad_outlier	1.0000

in the following calculation for "weighted score": 1*0.7002 + 1*0.9790 + 1*0.8609 = 2.5401.

4. The value of "weighted_score" is compared to the sum of the F1 scores ("F1_score_sum") for the three selected ODT. In our example, this sum is also 2.5401. This data point is an outlier and its weighted outlier flag is set to 1.

The results of applying the 'calculate weighted score' algorithm to the testing set are summarized in Table 3. For the evaluation, all soundings with an outlier flag of 0 or 1 are used. All soundings labeled with 2 should be labeled by a domain expert (in our study area this is the case for 3382 out of approximately 18.648.877 soundings). The precision value of 0.841 indicates that 84.1% of the samples that were predicted as outliers by our function had been manually labeled as outliers, while the recall value of 0.823 indicates that our function was able to correctly identify 82.3% of the true outliers in the data set. The F1 score value is 0.832.

Compared to the results for individual ODTs presented in Table 1, these results demonstrate that the weighting function can significantly enhance outlier detection performance in multibeam data, surpassing any particular ODT.

The Weighted Outlier Function stands out as an effective method, closely mirroring the results of manual data cleaning (see top middle and top right-hand panel in Fig. 7). It excels in removing outliers of all types, notably diminishing the visibility of significant clusters without creating the data voids common to other methods. Its precision and adaptability mark it as a robust solution for complex multibeam sonar data, significantly improving data quality for further analysis. In conclusion, while various techniques offer differing effectiveness in multibeam sonar data cleaning, the Weighted Outlier Function distinguishes itself with its thorough and precise approach to outlier mitigation.

MLP is predominantly used in our weighted outlier detection (see column 3 in Table 2) due to its high precision, recall, and F1 score in our evaluation, while MAD, a statistic-based method, is utilized for soundings with a high standard deviation of the mean depth. The results of the evaluation confirm the suitability of mapping the

Algorithm 1 Weighted Outlier Detection Algorithm

```
function CALCULATE_WEIGHTED_SCORE(row)
    // Initialize the weighted score for the row
    weighted_score ← 0
    // Initialize the sum of F1 scores for the row
    F1_score_sum ← 0
    // Loop over the three columns
    for each col in ['mean_depth_100m', 'normalized_distance_100m', 'std_dev_depth_100m'] do
        // Extract the value of the column for the current row
        val ← row[col]
        // Find the range in the dictionary to which the value belongs
        for each key in range_dict[col] do
            range_min, range_max ← key.split('|')
            if float(range_min) ≤ val < float(range_max) then
                // Multiply the value with the F1 score and add it to the weighted score
                outlier_type, F1_score ← range_dict[col][key]
                weighted_score ← weighted_score + row[outlier_type] * F1_score
                F1_score_sum ← F1_score_sum + F1_score
                break
            end if
        end for
    end for
    // Add the weighted score and the sum of F1 scores to the row
    row['weighted_score'] ← weighted_score
    row['F1_score_sum'] ← F1_score_sum
    // Return the updated row
    return row
end function
```

Table 3
Evaluation results of our 'calculate_weighted_score' algorithm showing improved performance for applying it to the complete data set.

Method	Precision	Recall	F1 score
Weighted outlier detection	0.841	0.823	0.832

outlier detection techniques to the four use cases presented earlier. For Cases a, b, and potentially c, with a high standard deviation in depth, MAD is the ODT of choice as it is less affected by outliers and has a fixed threshold suitable for regions with significant spread of nearby soundings. MLP showed overall the best performance for all Cases a–d, as it can identify isolated soundings, outliers connected to the seafloor and outliers in a noisy data set with a rough seafloor. The density-based methods, DBSCAN and KMEANS, have the highest F1 scores for soundings with very low or very high normalized distance to the mean depth and KMEANS shows the highest F1 score for mean depth between −4500 - −4000, indicating their effectiveness in detecting outliers located below the seabed (Case a) and far from nearby soundings (Case b) due to lower cluster density. Lastly, Logistic Regression outperforms all other ODTs for soundings that have a standard deviation of the depth between 200–400 and 500–600.

In considering broader implications, our enhanced multibeam sonar data outlier detection not only refines seafloor mapping accuracy but also underscores the importance of precise mapping for sustainable geoscientific applications. For instance, (Mohammed et al., 2022) emphasize the significance of accurate seabed mapping in understanding sediment dynamics pivotal to aquatic ecosystem health. Furthermore, the work of Koley (2023) on groundwater arsenic contamination suggests that improved seabed mapping could better identify arsenic-prone regions, bolstering mitigation efforts. Such intersections highlight our research's potential role in reinforcing sustainable geoscientific endeavors. In geo-exploration, the efficacy of numerical modeling, augmented by advanced computing methods, is paramount. These methodologies enhance the precision and scope of geophysical analysis across varied geographic spectrums. Such computational approaches not only facilitate a deeper understanding of complex geological formations but also extend their applicability to diverse field conditions, thereby supporting comprehensive environmental and resource assessments, like resource management (Kuhn et al., 2020), environmental monitoring (Tassetti et al., 2015), and hazard assessment (Federici et al., 2019).

However, our method also has some limitations. One of the weaknesses of our method is that it requires a domain expert to manually classify outliers when the weighted score is neither 0 nor equal to the sum of the F1 scores, effectively when the preferred ODTs for the three attributes of the data point to not agree on a classification. This is time-consuming and may require additional expertise. Our weighting process also makes the implicit assumption that the best-performing outlier detection technique for each range is always accurate, which may not always be the case. Finally, we are presently unable to quantify the effectiveness of our method on different types of multibeam data, such as data collected in different ocean environments or from different types of multibeam sensors.

6. Conclusion

Multibeam echosounder mapping of the seafloor is essential for understanding, for example, the geology, ecosystem distribution and physical oceanography of the ocean basins. Data quality control is essential and has traditionally been carried out by hand, although some automated methods are available. In this paper we have attempted to improve the performance of these methods by developing a weighted outlier detection function that takes into account multiple attributes of the input data to determine which of a range of different outlier detection techniques (statistical, distance- and density-based, machine learning) is most appropriate for assessing the validity of each individual data point. An evaluation shows that this weighting method outperforms the individual outlier detection techniques and other state-of-the-art methods in terms of precision, recall, and F1 score. The weighting method is versatile, performing well across different ranges of data attributes. In particular, the method is effective for detecting outliers in areas with high standard deviation of the mean depth, as well as for identifying outliers that are far away from the rest of the soundings.

CRediT authorship contribution statement

Tobias Ziolkowski: Writing – original draft, Software, Methodology, Conceptualization. **Agnes Koschmider:** Writing – review & editing, Supervision, Methodology. **Colin W. Devey:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tobias Ziolkowski reports financial support was provided by Helmholtz Centre for Ocean Research.

Data availability

The data set is publicly available on https://doi.pangaea.de/10.1594/PANGAEA.918716.

Acknowledgments

We acknowledge the financial support provided by MarDATA | Helmholtz School for Marine Data Science, Germany for this research project. The RV Maria S Merian is financed by the Deutsche Forschungsgemeinschaft, Germany. We thank Captain Björn Maaß and his crew for support at sea while collecting the training data set. Anne-Cathrin Wölfl, Mia Schumacher and numerous students undertook the arduous task of manually flagging the outliers in the data set.

Code availability section

Name of the code/library: An-optimized-outlier-detection-function-for-multibeam-echo-sounder-data

Contact: Tobias Ziolkowski, tziolkowski@geomar.de

Hardware requirements: Software was developed on a computer with 40 CPU cores (2.4 GHz each) and 256 GiB RAM. A PC with 8 CPU cores and 32 GiB RAM installed is required for basic running of the software. Running example can be executed with a modern laptop.

Program language: Python 3.10.6

Software required: No external software. Python dependencies are listed in the source code.

Program size: 1.24 MB

The source codes are available for downloading at the link: https://github.com/TZiolkowski2910/An-optimized-outlier-detection-function-for-multibeam-echo-sounder-data.git.

References

Aggarwal, C.C., Yu, P.S., 2001. Outlier detection for high dimensional data. In: Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data. SIGMOD '01, Association for Computing Machinery, New York, NY, USA, pp. 37–46. http://dx.doi.org/10.1145/375663.375668.

Ahmed, M., Seraj, R., Islam, S.M.S., 2020. The k-means algorithm: A comprehensive survey and performance evaluation. Electronics 9 (8), http://dx.doi.org/10.3390/electronics9081295, URL https://www.mdpi.com/2079-9292/9/8/1295.

Alghushairy, O., Alsini, R., Soule, T., Ma, X., 2021. A review of local outlier factor algorithms for outlier detection in big data streams. Big Data Cogn. Comput. 5 (1), URL https://www.mdpi.com/2504-2289/5/1/1.

Alimohammadi, H., Nancy Chen, S., 2022. Performance evaluation of outlier detection techniques in production timeseries: A systematic review and meta-analysis. Expert Syst. Appl. 191, 116371. http://dx.doi.org/10.1016/j.eswa.2021.116371, URL https://www.sciencedirect.com/science/article/pii/S095741742101664X.

ichi Amari, S., 1993. Backpropagation and stochastic gradient descent method. Neurocomputing 5 (4), 185–196. http://dx.doi.org/10.1016/0925-2312(93)90006-O, URL https://www.sciencedirect.com/science/article/pii/092523129390006O.

Bourillet, J., Edy, C., Rambert, F., B., L., 1996. Swath mapping system processing: Bathymetry and cartography. pp. 487–506. http://dx.doi.org/10.1007/BF00286091.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. http://dx.doi.org/10.1023/A:1010950718922.

Calder, B.R., Mayer, L.A., 2003. Automatic processing of high-rate, high-density multibeam echosounder data. Geochem. Geophys. Geosyst. 4 (6), http://dx.doi.org/10.1029/2002GC000486, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002GC000486, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002GC000486.

Caris, T., 2023. CARIS Mira AI. URL https://www.teledynecaris.com/en/products/caris-mira-ai/. (Accessed: 13 March 2023).

Caroline Cynthia, P., Thomas George, S., 2021. An outlier detection approach on credit card fraud detection using machine learning: A comparative analysis on supervised and unsupervised learning. In: Peter, J.D., Fernandes, S.L., Alavi, A.H. (Eds.), Intelligence in Big Data Technologies—beyond the Hype. Springer Singapore, Singapore, pp. 125–135.

Clark-Carter, D., 2005. Interquartile range. In: Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd, http://dx.doi.org/10.1002/0470013192.bsa311, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470013192.bsa311, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa311.

Cun, Y.L., Bottou, L., Bengio, Y., 1997. Reading checks with multilayer graph transformer networks. In: 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing. Munich, Germany, pp. 151–154. http://dx.doi.org/10.1109/ICASSP.1997.599580.

Devey, C., Wölfl, A.-C., Augustin, N., Besaw, M., Damaske, D., Evora, D., Gray, A., Hübscher, C., Le Saout, M., Lux, T., 2020. Bathymetric mapping of the seafloor - a german contribution to completing the map by 2030, cruise no. MSM88/1 + MSM88/2. http://dx.doi.org/10.2312/cr_msm88.

Enderlein, G., Hawkins, D.M., 1987. Identification of outliers. Chapman and hall, London – New York 1980, 188 s., £14, 50. Biom. J. 29, 198.

Farr, H.K., 1980. Multibeam bathymetric sonar: Sea beam and hydro chart. Marine Geodesy 4 (2), 77–93. http://dx.doi.org/10.1080/15210608009379375.

Federici, B., Corradi, N., Ferrando, I., Sguerso, D., Lucarelli, A., Guida, S., Brandolini, P., 2019. Remote sensing techniques applied to geomorphological mapping of rocky coast: the case study of gallinara island (western liguria, Italy). Eur. J. Remote Sens. 52 (sup4), 123–136. http://dx.doi.org/10.1080/22797254.2019.1686957.

Gardner, M., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32 (14), 2627–2636. http://dx.doi.org/10.1016/S1352-2310(97)00447-0, URL https://www.sciencedirect.com/science/article/pii/S1352231097004470.

Grim, P., 1988. The COP algorithm for selecting soundings within a PUA. Natl Ocean. Atmosph. Admin..

Guenther, G.C., Green, J.E., 1982. Improved depth selection in the bathy- metric swath survey system (BS3) combined offline processing (COP) algorithm. Natl Ocean Atmosph. Admin. OTES 10, URL https://repository.library.noaa.gov/view/noaa/23480.

Hahsler, M., Piekenbrock, M., Doran, D., 2019. Dbscan: Fast density-based clustering with R. J. Stat. Softw. 91 (1), 1–30. http://dx.doi.org/10.18637/jss.v091.i01, URL https://www.jstatsoft.org/index.php/jss/article/view/v091i01.

Herlihy, D.R., Stepka, T.N., Rulon, T.D., 2018. Filtering erroneous soundings from multibeam survey data. Int. Hydrogr. Rev. 69 (2), URL https://journals.lib.unb.ca/index.php/ihr/article/view/26178.

Howell, D.C., 2005. Median absolute deviation. In: Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd, http://dx.doi.org/10.1002/0470013192.bsa384, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470013192.bsa384, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa384.

Hughes Clarke, J.E., 2003. Dynamic motion residuals in swath sonar data: Ironing out the creases. Int. Hydrogr. Rev. 4 (1), URL https://journals.lib.unb.ca/index.php/ihr/article/view/20600.

Jain, V., Seung, S., 2008. Natural image denoising with convolutional networks. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (Eds.), Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference. In: Advances in Neural Information Processing Systems, vol. 21, Curran Associates, Inc., pp. 769–776, URL https://proceedings.neurips.cc/paper_files/paper/2008/file/c16a5320fa475530d9583c34fd356ef5-Paper.pdf.

Knorr, E.M., Ng, R.T., 1998. Algorithms for mining distance-based outliers in large datasets. In: Very Large Data Bases Conference. New York City, pp. 392–403.

Koley, S., 2023. Arsenic calamity in India's west bengal: a critical review of mitigation scenarios. Proc. Inst. Civil Eng. Eng. Sustain. 176 (4), 198–213. http://dx.doi.org/10.1680/jensu.21.00066.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems. vol. 25, Curran Associates, Inc., pp. 1097–1105, URL https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Kuhn, T., Uhlenkott, K., Vink, A., Rühlemann, C., Martinez Arbizu, P., 2020. Chapter 58 - manganese nodule fields from the northeast Pacific as benthic habitats. In: Harris, P.T., Baker, E. (Eds.), Seafloor Geomorphology As Benthic Habitat (Second Edition), second ed. Elsevier, pp. 933–947. http://dx.doi.org/10.1016/B978-0-12-814960-7.00058-0, URL https://www.sciencedirect.com/science/article/pii/B9780128149607000580.

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 (7), http://dx.doi.org/10.3390/geosciences10070254, URL https://www.mdpi.com/2076-3263/10/7/254.

Le Deunf, J., Nathalie, D., Thierry, S., François, G., Etienne, J., Lucas, V., Romain, B., 2019. Outlier detection for multibeam echo sounder (MBES) data: from past to present. In: OCEANS 2019 - Marseille. pp. 1–10. http://dx.doi.org/10.1109/OCEANSE.2019.8867321.

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49 (4), 764–766. http://dx.doi.org/10.1016/j.jesp.2013.03.013, URL https://www.sciencedirect.com/science/article/pii/S0022103113000668.

Lirakis, C., Bongiovanni, K., 2000. Automated multibeam data cleaning and target detection. In: OCEANS 2000 MTS/IEEE Conference and Exhibition. Conference Proceedings (Cat. No.00CH37158), Vol. 1. pp. 719–723 vol.1. http://dx.doi.org/10.1109/OCEANS.2000.881336.

Lu, D., Li, H., Wei, Y., Zhou, T., 2010. Automatic outlier detection in multibeam bathymetric data using robust LTS estimation. In: 2010 3rd International Congress on Image and Signal Processing, Vol. 9. pp. 4032–4036. http://dx.doi.org/10.1109/CISP.2010.5648184.

Michel, M., Deunf, J.L., Debese, N., Bazinet, L., Dejoie, L., 2021. Multibeam outlier detection by clustering and topological persistence approach, ToMATo algorithm. In: OCEANS 2021: San Diego – Porto. pp. 1–8. http://dx.doi.org/10.23919/OCEANS44145.2021.9705930.

Mohammed, S.J., Zubaidi, S.L., Ortega-Martorell, S., Al-Ansari, N., Ethaib, S., Hashim, K., 2022. Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective. Cogent Eng. 9 (1), 2143051. http://dx.doi.org/10.1080/23311916.2022.2143051.

Monterey Bay Aquarium Research Institute (MBARI), 2023. MB-system installation guide. URL https://www.mbari.org/technology/mb-system/installation/.(Accessed: 13 March 2023).

Panjei, E., Gruenwald, L., Leal, E., Nguyen, C., Silvia, S., 2022. A Survey on Outlier Explanations, Vol. 31, No. 5. Springer-Verlag, Berlin, Heidelberg, http://dx.doi.org/10.1007/s00778-021-00721-1.

QPS, 2023. Qimera Download Page. URL https://qps.nl/downloads/qimera/#. (Accessed: 13 March 2023).

Rahmatullah Imon, A.H.M., Hadi, A.S., 2008. Identification of multiple outliers in logistic regression. Comm. Statist. Theory Methods 37 (11), 1697–1709. http://dx.doi.org/10.1080/03610920701826161.

Sedaghat, L., Hersey, J., McGuire, M.P., 2013. Detecting spatio-temporal outliers in crowdsourced bathymetry data. ISBN: 9781450325288, http://dx.doi.org/10.1145/2534732.2534739.

Šiljeg, A., Marić, I., Domazetović, F., Cukrov, N., Lovrić, M., Panđ a, L., 2022. Bathymetric survey of the st. Anthony channel (Croatia) using multibeam echosounders (MBES)—A new methodological semi-automatic approach of point cloud post processing. J. Marine Sci. Eng. 10, 101. http://dx.doi.org/10.3390/jmse10010101.

Souiden, I., Brahmi, Z., Toumi, H., 2017. A survey on outlier detection in the context of stream mining: Review of existing approaches and recommadations. In: Madureira, A.M., Abraham, A., Gamboa, D., Novais, P. (Eds.), Intelligent Systems Design and Applications. Springer International Publishing, Cham, pp. 372–383.

Stephens, D., Smith, A., Redfern, T., Talbot, A., Lessnoff, A., Dempsey, K., 2020. Using three dimensional convolutional neural networks for denoising echosounder point cloud data. Appl. Comput. Geosci. 5, 100016. http://dx.doi.org/10.1016/j.acags.2019.100016, URL https://www.sciencedirect.com/science/article/pii/S2590197419300163.

Syakur, M.A., Khotimah, B.K., Rochman, E.M.S., Satoto, B.D., 2018. Integration K-means clustering method and elbow method for identification of the best customer profile cluster. IOP Conf. Series: Mater. Sci. Eng. 336 (1), 012017. http://dx.doi.org/10.1088/1757-899X/336/1/012017.

Tassetti, A.N., Malaspina, S., Fabi, G., 2015. Using a multibeam echosounder to monitor an artificial reef. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XL-5/W5, 207–213. http://dx.doi.org/10.5194/isprsarchives-XL-5-W5-207-2015, URL https://isprs-archives.copernicus.org/articles/XL-5-W5/207/2015/.

Wang, J., Jin, S., Bian, G., Cui, Y., Long, Z., 2023. A multi-beam outlier automatic filtering algorithm combining uncertainty and density clustering method. Acta Geodaetica Cartograph. Sinica 52 (10), 1669. http://dx.doi.org/10.11947/j.AGCS.2023.20220579, URL http://xb.chinasmp.com/EN/abstract/article_13128.shtml.

Ware, C., Fellows, D., Wells, D., 1990. Feasibility study on the use of algorithms for automatic error detection in data from the FCG smith. Contract Rep..

Ware, C., Slipp, L., Wong, K.W., Nickerson, B., Wells, D., Lee, Y.C., Dodd, D., Costello, G., 2018. A system for cleaning high volume bathymetry. Int. Hydrograph. Rev. 69 (2), URL https://journals.lib.unb.ca/index.php/ihr/article/view/26179.

Wei, Y., Li, S., Jin, S., Ji, H., 2021. Automatic cleaning of outliers in multibeam bathymetric data with clustering algorithm. Abstract ICA 3, 314. http://dx.doi.org/10.5194/ica-abs-3-314-2021, URL https://ica-abs.copernicus.org/articles/3/314/2021/.

Yang, F., Li, J., Chu, F., Wu, Z., 2007. Automatic detecting outliers in multibeam sonar based on density of points. In: OCEANS 2007 - Europe. Aberdeen, UK, pp. 1–4. http://dx.doi.org/10.1109/OCEANSE.2007.4302202.

Yang, F., Xu, F., Fan, M., Bu, X., Tu, Z., Yan, X., 2022. An intelligent detection method for different types of outliers in multibeam bathymetric point cloud. IEEE Trans. Geosci. Remote Sens. 60, 1–10. http://dx.doi.org/10.1109/TGRS.2022.3209344.

Zhang, K., Hutter, M., Jin, H., 2009. A new local distance-based outlier detection approach for scattered real-world data, CoRR abs/0903.3257. arXiv:0903.3257.

Zhou, T., Si, J., Wang, L., Xu, C., Yu, X., 2022. Automatic detection of underwater small targets using forward-looking sonar images. IEEE Trans. Geosci. Remote Sens. 60, 1–12. http://dx.doi.org/10.1109/TGRS.2022.3181417.