Outlier detection for Multibeam echo sounder (MBES) data from past to present

Julian Le Deunf, Nathalie Debese, Thierry Schmitt, François Guibourt, Jenner Etienne, Vincent Lucas, Romain Billot

Outlier detection for Multibeam echo sounder (MBES) data: from past to present

Le Deunf Julian
Development bathymetry department
Shom
IMT Atlantique, Lab-STICC, F-29238 Brest, France
julian.le.deunf@shom.fr

Debese Nathalie
Hydrographic department
ENSTA Bretagne
Lab-STICC
Brest, France
nathalie.debese@ensta-bretagne.fr

Schmitt Thierry
Development bathymetry department
Shom
Brest, France
thierry.schmitt@shom.fr

Guibourt François
IMT Atlantique
Brest, France
francois.guibourt@imt-atlantique.net

Jenner Etienne
IMT Atlantique
Brest, France
etienne.jenner@imt-atlantique.net

Vincent Lucas
IMT Atlantique
Brest, France
lucas.vincent@imt-atlantique.net

Billot Romain
IMT Atlantique
Lab-STICC
F-29238 Brest, France
romain.billot@imt-atlantique.fr

Abstract—With the new data acquisition capabilities of latest MBES, the need of an automation data processing is more and more essential. The aim of this article is to present what kind of data processing we want to improve, what techniques were used in the past and how machine learning could help us now and in the future for a better bathymetric data processing.

Keywords— Outlier detection; Data processing; Multibeam echo sounder data; Machine learning

I. INTRODUCTION

For hydrographic offices, outlier detection is a critical and time-consuming task. This is inherent to their mission to map the ocean floor and ensure safety of navigation. High level of confidence is hence required throughout all the data acquisition and processing steps. For this reason, bathymetric data processing for nautical chart production has often been carried out manually. Such an approach is performed by trained operators visualizing one by one all the soundings of a survey, pointing out erroneous soundings from local validations of the bathymetry. Given the huge amount of data collected by the new generation of data acquisition systems (Multibeam Echosounder (MBES), bathymetric LIDAR, the autonomous surface/underwater vehicle (ASV / AUV), and crowdsourcing bathymetry), such a task is inevitably repetitive, fastidious and subjective. Moreover, with the use of fully automatic machines, the need to process data in near real time will become a challenge. Therefore, the use of automated algorithms for outlier detection is getting critical, to significantly reduce processing time, ensure objectivity, guarantee the cleaning procedure traceability and make sure that the desired data quality is achieved.

As an illustration of the previous paragraph, the French Hydrographic and Oceanographic Service (Shom) has recently renewed all the MBES equipping its fleet between 2011 and 2017. With an increase from 256 beams per ping to 800 beams, the volume of acquired data has increased by a factor of ten over this 6-year period as illustrated in Fig. 1. Even if the new MBES are showing to be more accurate, producing weaker erroneous sounding rates, the validation post processing step is still needed to ensure navigation safety.

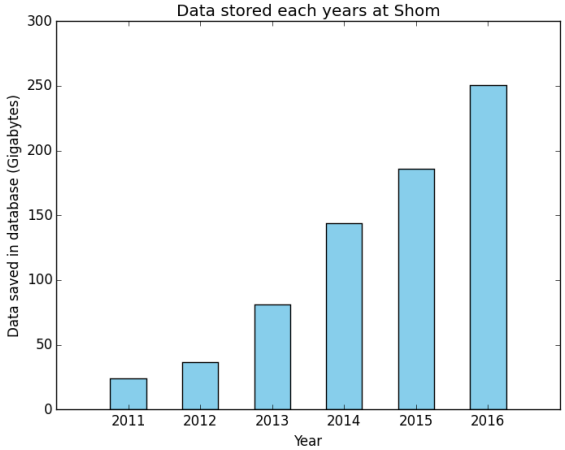

Fig. 1: Volume of bathymetric data post- processed at Shom between 2011 and 2016

We have chosen the term outlier to define the false information collected during a hydrographic survey (as in Hodge & Austin [1]). Although diverging definitions involving coherency and temporal stability have been discussed in Edgeworth's study [2], we have retained the following definition build on Grubbs' work [3] and quoted in Barnett & Lewis [4]:

"An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs."

In the hydrography field, it must be taken into account that what emerge from the seabed data may also be natural or human obstructions. And by nature these obstructions are difficult to discern. With respect to critical applications such as navigation safety, it is necessary to select the best detection algorithms used to find the real outlier observation.

In this paper we will first see the definition of the errors found in the hydrographic field. Then we will look at the classic errors detection method. Finally, we will study the methods linked to the machine learning to find these outliers

II. .HYDROGRAPHIC ERROR TYPOLOGY

In the field of hydrography one can define three types of errors: systematic errors, abnormal soundings and noise measure (random uncertainty) as explained by Debese in [5] and represented on the figure below.

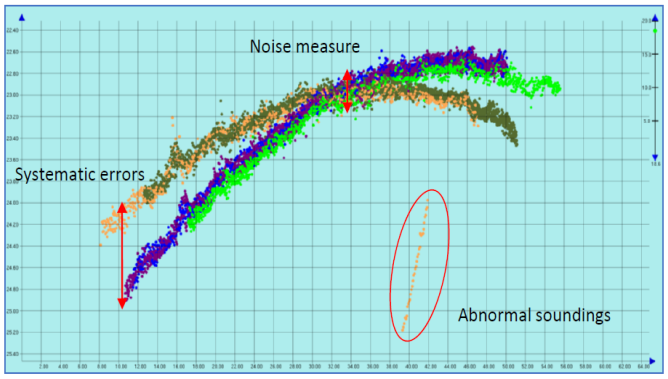

Fig. 2: Soundings within a slicing box colored by survey lines

A. *Systematic errors*

Systematic errors mainly come from poor control of the measurement environment such as improper calibration of devices composing the hydrographic system, the results of issues related to tide levelling or sound velocity measurements in the water column. For example, In the Fig. 2, the systematic error in the bathymetric data is most likely

due to a roll bias not adequately taken into account during the acquisition process.

Most of the time the systematic errors are resolved as part of the calibration procedure (e.g. the patch test for boresight angle), which is why outlier detection algorithms often make the assumption that the systematic errors are solved. In that case, remaining errors in bathymetric data are abnormal soundings.

B. Abnormal soundings

Remaining outliers are abnormal soundings which are not representative of the true seabed bathymetry. These can be the result of punctual malfunctions of the sounders, human errors at the time of acquisition or environmental phenomena such as an acoustic contrast linked for example to the presence of a school of fishes or a hydrothermal vent. As their origin can be multiple and poorly anticipated, these outliers are difficult to identify. Finding the perfect algorithm that could remove all types of abnormal soundings is a real challenge.

In this paper all the methods used for outlier detection in MBES data aim to detect this type of error.

C. Noise measure

Noise measurement error is linked to each of the individual sensor's inherent physical limitations. Hence, this noise measure error is accentuated as a result of the convolution related to the integration of all the instruments composing the bathymetric acquisition sensors.

This noise gives us pertinent information about the sensibility of the sensor's system. Therefore, we want to preserve this information and we can't suppress any sounding that would be in this noise (classically known as the the "soundings masttress").

In the context of navigation safety, the final bathymetric data production must respond to an international standard: the IHO standards for hydrographic surveys, special publication n°44, see [6].

III. OUTLIER DETECTION APPROACHES IN THE HYDROGRAPHIC FRAMEWORK

Fig. 3 presents a non-exhaustive classification of outlier detection that can be found in the hydrographic literature. We clearly see an unbalanced class of outlier detection with only one algorithm using supervised approach and all others using unsupervised classification/filtering. In this article we will focus on this unsupervised perspective. The second level of this diagram shows us that we can separate the algorithms according to the type of segmentation used. We will here use this typology to describe various algorithms, starting by describing these types of segmentation.
These different algorithms are just a sample of outlier detection techniques applied to hydrography. This chapter is not exhaustive but it gives us a good idea of what has been done in the field.

A. Data segmentation

In hydrography data can be handled with a dual representation: either in time series, in the referential frame of the acquisition system, known as a ping/beam view

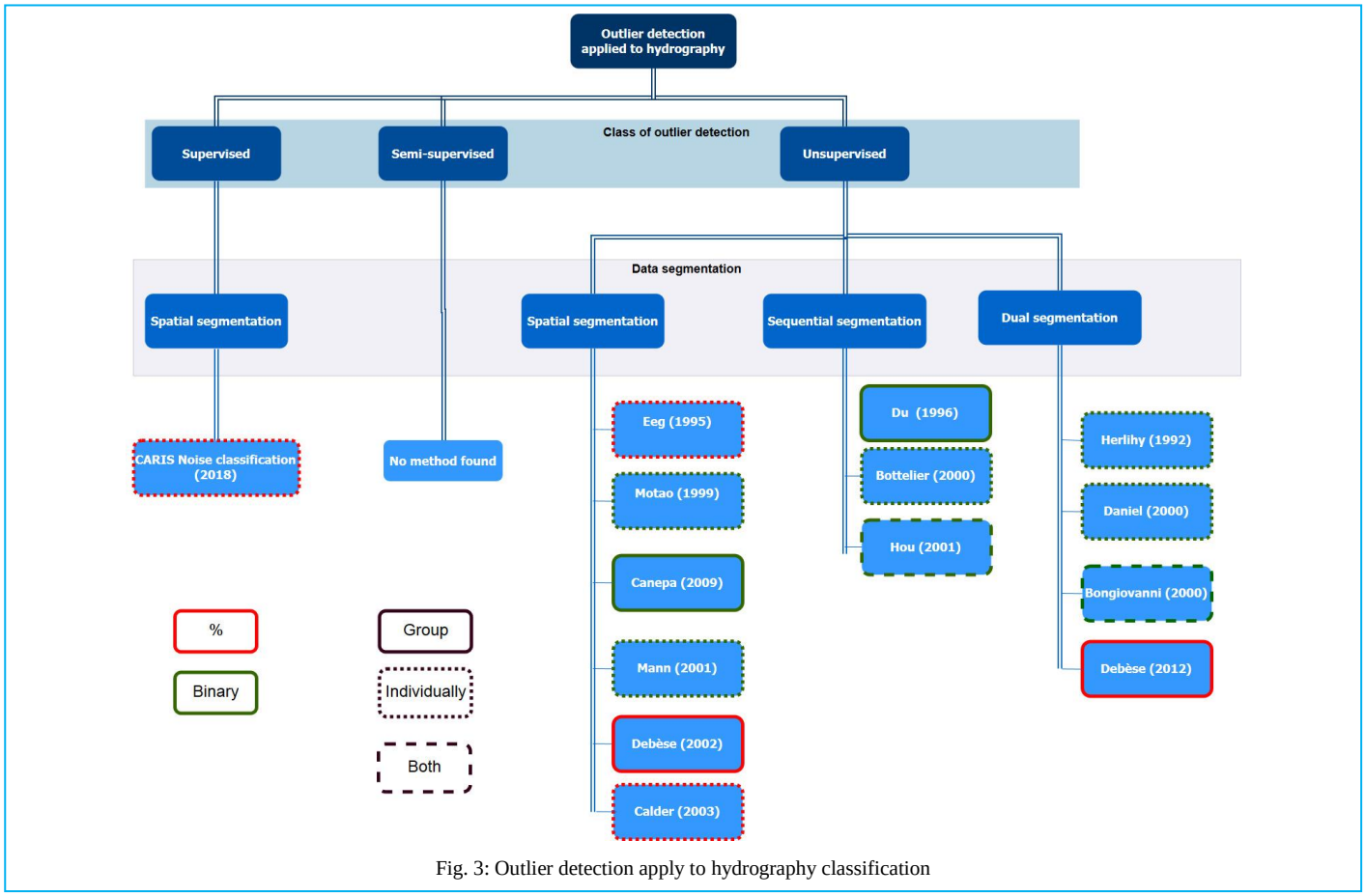

Fig. 3: Outlier detection apply to hydrography classification

(sequential view); or in an absolute georeferenced data frame (spatial representation)

1) Sequential representation

In the sequential representation, data are stored in a matrix, with the beam number along the line axis and ping number along the column axis (Fig. 4). This approach is based on the ping/beam point of view. It is studing data swath by swath, it can't be used when we search the data overlaying but the density is fixed and very similar to a matrix.

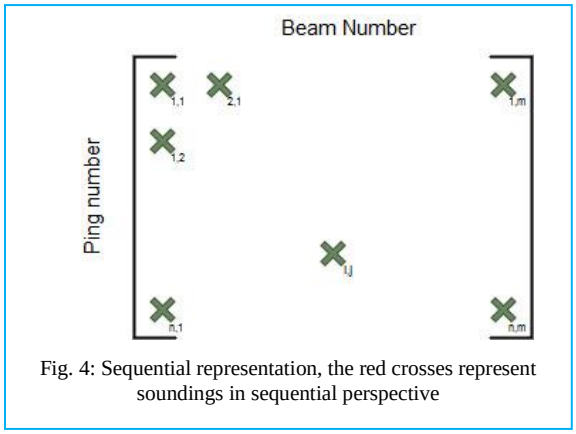

Fig. 4: Sequential representation, the red crosses represent soundings in sequential perspective

2) Spatial representation

This bathymetric data represents each sounding as a triplet where x and y are the geographical coordinates (Fig. 5). This classical representation is particularly useful to control the consistency of the soundings on superimposed parts of adjacent swathes. The major issue with this data representation is that, due to sensors' geometry and data acquisition conditions the density of this representation is variable.

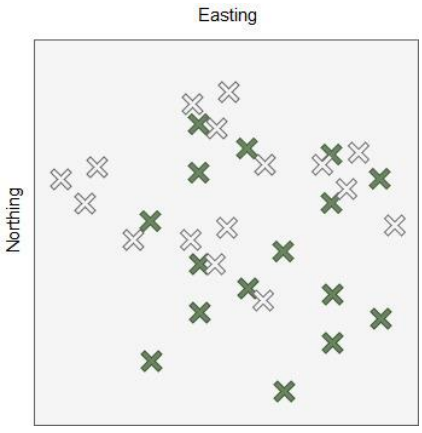

Fig. 5: Spatial representation, the crosses' color depends on the data swath.

B. patial segmentation based outlier detection methods

In the context of outlier detection, it is common practice to consider one or the other representation for outlier detection. We will present some algorithms based on these different types of representation.

We will first examine the spatial representation.

1) CUBE (2002)

The Combined Uncertainty and Bathymetry Estimator (CUBE see [7]) algorithm was developed by Brian Calder at CCOM-UNH. This method is an error-model based on the computation of a digital bathymetric model (DBM) which estimates the depth associated with a confidence interval directly on each of the node points of a bathymetric grid (see Fig. 6).The algorithm works in 3 steps. The first one is making the data selection for each grid node, which will be used for the hypothesis computation; it is based on the total propagated uncertainty (TPU). The second one is building the hypothesis for each grid node. The third one is the disambiguation of the previous hypothesis. For each node, the algorithm proposes to the hydrographer alternative seabed hypotheses if the sounding's dispersion (with regard to the parameter setting used) is too important.

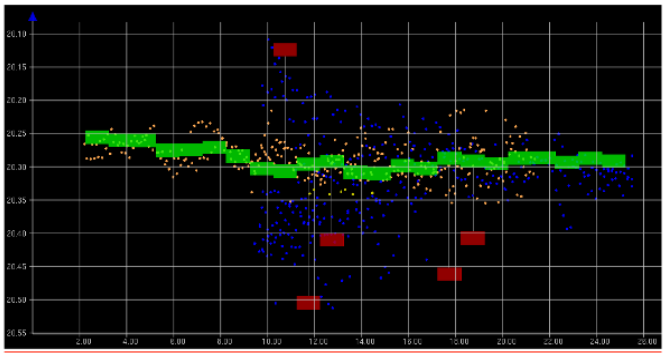

Fig. 6: Superimposition of CUBE hypothesis (in green: CUBE selected hypothesis in red: CUBE alternative hypothesis) to soundings within a slicing box (2D display from CARIS/HIPS software)

Today this algorithm is widely used, at the National Oceanographic and Atmospheric Administation (NOAA) and at the the Canadian Hydrographic Service (CHS). There are still some issues with this algorithm: In the case of chaotic seafloor (rocky area or obstructions), the number of hypotheses significantly increases requiring the intervention of the hydrographer. It is strongly recommended to perform a quick manual pre-filtering on the data. An improved version of CUBE is proposed with CHRT (CUBE with Hierarchical Resolution Techniques) including the multi-resolution, multi-processing and taking into account the quality factor developed by Ifremer (Institut français de recherche pour l'exploitation de la mer) which could resolve the issues explained above.

2) RMQMP(2018) method

The Robust Multi-quadric Method and Median Parameter Model (RMQMP) is described in [8] and in Fig .7. At first a fitting trend surface model is built, a median parameter method is used to obtain a first value of residual error which is applied as an initial value within an iterative process to weaken soundings' weights (considered as outliers) in the DBM generation.

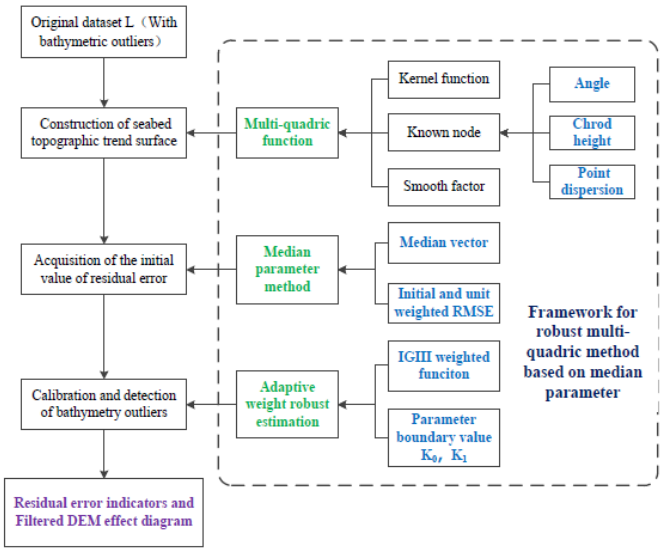

Fig. 7: Bathymetry Outlier Detection Procedure of RMQMP method from [8]

This method is really sensitive to the initial value of robust estimation (contamination of this value by outliers). This initial value is computed through the median parameter method using a fix constant (used in the weight function) so we can question the performance of the method on a very wide seabed with large depth amplitude.

C. *Sequential segmentation based outliers detection methods*

Following the spatially based outlier detection methods, we will now introduce outliers detection methods based on the sequential representation of the hydrographic data.

1) Du (1996) algorithm

This method is based on a data clustering approach (namely Dixon test). Data are bundled in modes [9]; these aggregations are formed on data with same characteristics (distance characteristics in this case). After applying a depth data thresholding, we only keep data in between a minimal and maximal depth. The depth histogram is computed (see Fig. 8), and analysed to find the main mode and the secondary modes. When a secondary mode is detected at a vertical difference which is considered too far from the the main mode, then the secondary mode is flagged as an outlier. This algorithm is a recursive method, so the working window of the algorithm changes with the processing (it decreases in size). It starts from a large number of pings and becomes smaller and smaller (10 pings in the first implementation).

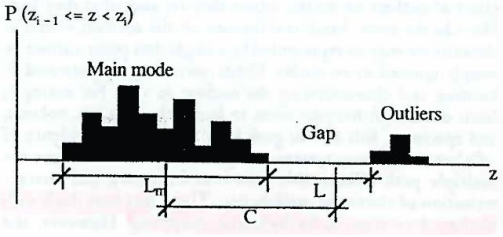

Fig. 8: Taken from [9], clustering by mode seeking.

2) Hou (2001) algorithm

For this algorithm [10] all the data are stacked over 60 pings. The outliers' detection successively applies three filters, from a global perspective to single bad ping detection. The first filter tests the data heterogeneity by computing a global and local variance. The second filter tests the sounding depth contribution to the local standard deviation, with the concept that a more distinctive outlier will have a major impact on the local standard deviation. If this impact is over a predefined value, the sounding will be rejected. The third filter operates ping by ping. Erroneous beams are detected: if the beam values are too different from the neighborhood mean depth, the neighborhood used in the second filter is re-arranged in three sub-neighborhoods in order to compute their respective standard deviation and see if a beam is altering too much the standard deviation.
Note that this algorithm will be used as the support for the generation of sequential features in our Machine Learning workflow (see section IV of this paper).

D. *Hybrid segmentation based outliers detection methods*

In this section we will focus on the methods that work with both the spatial and sequential data representation.

1) CHARM algorithm Debese (2012)

This algorithm [11] performs a Cleaning of a MBES dataset through a Hierarchic Adaptive Robust Modeling approach. The seafloor is constructed as an assemblage of surface elements with the help of a robust statistical approach. The local parameters model is a priori chosen, its scale is driven through a quadtree descending approach using subdivision rules based on both statistical and spatio-temporal inferences. This multi resolution approach provides, with

the algorithm outputs, a classification map that notes areas of concern.

2) Herlihy (1992) algorithm

This algorithm will scan soundings one by one, for each sounding three criteria will be tested to classify if the sounding is an outlier or not as presented in [12]. The first step works on a spatial perspective, it measures the distance between the longitudinal axis and the sounding position. This helps finding far outliers. The second criterion is a similarity measure between the sounding depth observed and a weighted mean computed with soundings in close neighborhood in a sequential perspective (the neighborhood is computed on a swath perspective). The last criterion checks the validity of near-neighborhood used during the second step of the algorithm.

The process proposed here is really simple to implement and the criteria sequencing is very interesting. This type of technique struggles to deal with very dense groups of outliers that will be considered as pertinent soundings.

3) Bonjiovani (2000) algorithm

This algorithm works with a ping stack that will pass through 5 filters. The first four filters work on the sequential approach and the final one on the spatial approach, see [13]. The first pass filters out all the data not included in a predefined depth range. The second filter is based on the covariance value applied on paired ping. The third filter computes the stacked ping roughness of the seabed from the variance and a gradient computed locally; it gives us a 2D histogram (of variance and gradient). From these 2 histograms, the filter defines the bounds of a confidence interval in which all the data will be accepted. The fourth filter continues this validation from local criteria of the previous filters. The final filter is constructed on a regular spatial grid; the status' sounding depends on the depth data standard deviation in the cell.

In this algorithm there are also different steps proposed in the workflow. This iterative procedure seems very efficient when multiple scenari of outliers are presented (isolated, dense, very distant…).

All algorithms proposed above are always working with fixed heuristics. Hence, their use might not be consistant all over the same dataset, where waterdepth, density and rugosity of the seafloor might be varying. Hence it would be preferable to have methods that will use information already contained in the data to generate filtering parameters.

IV. MACHINE LEARNING (ML) METHODS

The previous section explain us that classical outlier detection methods are based on a static heuristic method which is often fast to compute (for simple filtering algorithm) but might not be valid for all the dataset or all the types of data. In light of these findings, the goal of the following sub-sections is to introduce and test some conventional machine learning techniques and look at the advantages and drawbacks when applied to hydrographic data.

A. Data selection and description

The way bathymetric data is selected and represented is essential in our ML algorithm. This choice will affect the process data neighbourhood (density and spacing).

1) Data selection

a) Spatial segmentation

This is the most common way to select the data in the hydrographic point of view. It is very often used in the digital elevation model (DBM) computation, which is a 2.5D representation of the bathymetric information. The data will be selected depending on their affiliation cell, as illustrated on Fig. 9.

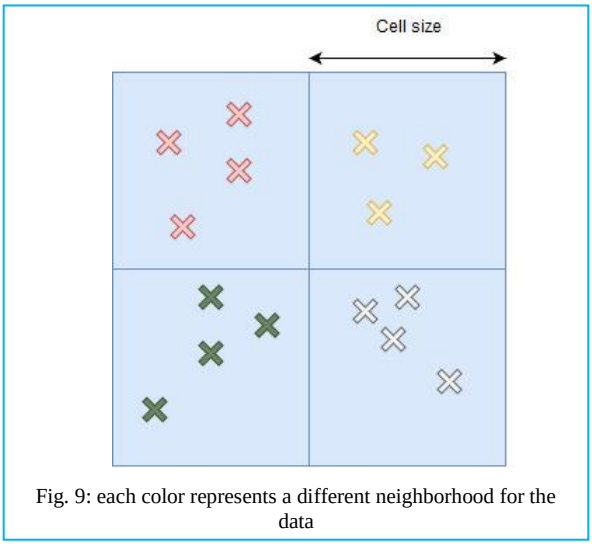

Fig. 9: each color represents a different neighborhood for the data

This is the easiest implementation for data selection. It only needs one parameter which is the cell's resolution, which needs to be carefully selected with respect to the density of soundings and the variability of the seafloor.

b) Moving window

This type of selection is data focused; it is always centered on the data studied. Because the window is always changing position and centred on one sounding, this method can be time consuming but the local neighbourhood is more representative than the previous data selection. It also needs only one parameter which is the search radius around the data.

Fig. 10 shows an example of a moving window data pattern, with the red cross being the centre of the moving window and the yellow plus sign being points selected.

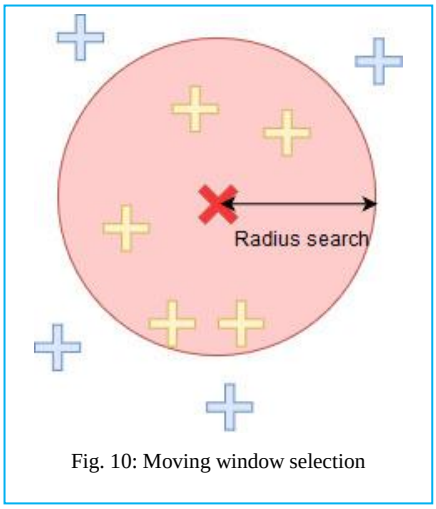

Fig. 10: Moving window selection

c) Quadtree structure

The quadtree is a structure used for partitioning horizontal two-dimensional space by recursively subdividing an initial square it into four quadrants. Each subdivision generates a relationship with the initial square such as in a tree. The terminating condition is often a condition set on the density of the data. This means that the resolution of each patch is adapted to the data. For each quadrant, we test a criterion. If the test is successful, we stop the quadtree. Else we divide the quadrant and try again in the four smaller quadrants. In

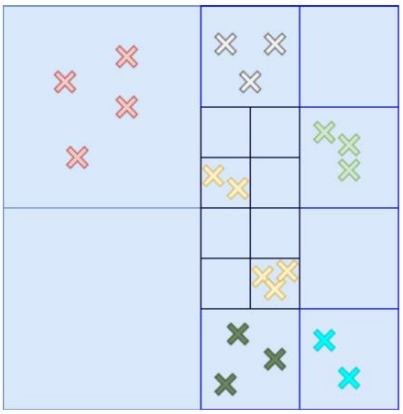

Fig. 11: Quadtree partitioning

Fig. 11 we can see an example of quadtree data selection.

This type of data partitioning is often used in bathymetry (as in [11] and [14]), this method gives an adapted resolution at any place. The choice of different criteria and thresholds give us a very flexible segmentation (depending on the application). It can easily be used in a 3D perspective with the Octree data structure.

In our first tests we have used the regular gridded structure because it was the simplest implementation for bathymetric data selection.

2) Data description and classification

As seen in III.A.a, we are given different ways to represent bathymetric data. Depending on the features we want to generate, we will use the spatial or the sequential representation. A feature is a particular description of the data. It can be obtained by measure or computed from data characteristics. All the ML algorithms are based on these features for data discrimination.

In order to run a ML algorithm, we need bathymetric data but also an accurate description of this data to train the algorithm classification task.

In the perspective of outlier detection, we will compute three different types of features:

- raw soundings features,
- spatial features
- sequential features,

as listed in Table 1.

TABLE I. TYPES OF FEATURES

Raw soundings features	*Spatial features*	*Sequential features*
➢ Emission Angle Across ➢ Emission Angle Along ➢ TPU ➢ Backscatter	➢ Median Absolute deviation (MAD) ➢ Local Outlier Factor	➢ Global variance estimation ➢ Local variance estimation ➢ Bad Ping Detection

Raw soundings features are based on bathymetric raw data such as the emission angle of acoustic ray tracing, the TPU which is computed both for the vertical and horizontal axis as detailed in [15] and the backscatter which is a measure of the intensity of the acoustic return. All this information is gathered directly from the raw datagrams.

Spatial features rely on spatial statistical dispersion of the

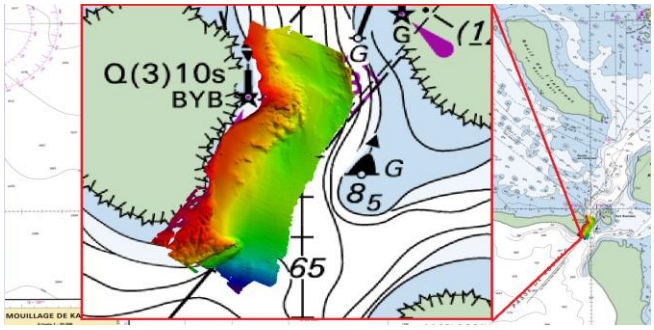

Fig. 12: Data sample on CM6985 @Shom

bathymetric data. To compute these features, we used methods given by [16] for MAD and [17] for Local Outlier Facto.
The sequential features depend on the beam and ping representation (see A.2.b), these features are all based on the article [18].

All these features are just a set of the various features used in our ML workflow.

B. ML workflow

For the initial tests using ML algorithms applied to bathymetric data, we have chosen to test the supervised classification perspective. Our labelled datasets at Shom are massive and made with our empirical outlook. Evaluating this work is a great opportunity permitted by ML algorithm.
The data used was acquired with an EM2040p from Kongsberg in New Caledonia; the depth amplitude goes from 9.93m to 99.25m depths. Fig. 12 shows the location of the survey in the Koumac pass.

1) Data analysis

A data analysis of the different features built was carried out before starting our different ML algorithms. The evaluation of the backscatter features (a raw soundings feature) has been carried out to compare the status: accepted, rejected at the conversion, and rejected by the hydrographer (see the Fig. 13). Results clearly show that this feature is currently used by manufacturers to reject data (during the detection process). On the data sample the mean backscatter level of data rejected by the hydrographer is lower than the accepted one. This feature seems discriminant for the soundings rejected by the manufacturer.

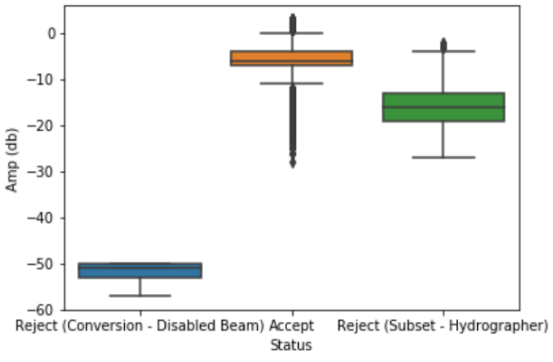

Fig. 13: The mean basckatter level depending to the status

Regarding the MAD (Median Absolute Deviation) feature (a spatial feature) we observe that the MAD distance is greater for the rejected data than for the accepted one (fig. 14). This behaviour was expected because the MAD computes the dispersion of the data, and the outliers are information widely scattered around the median. This feature is discriminant for our problem.

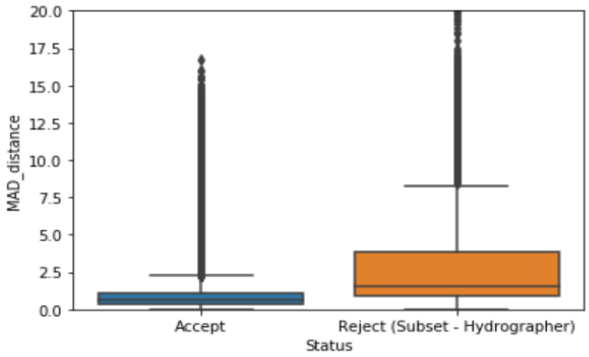

Fig. 14: The MAD distance depending to the status

As for the bad ping detection feature (a sequential feature, see [18] for computation), we observe that the bad ping detection feature is greater for the rejected data than for the accepted one (fig. 15). This behaviour is also logical since this feature is actually used as a part of a classical outlier detection algorithm (see [18]). This feature is clearly discriminant for our problem.

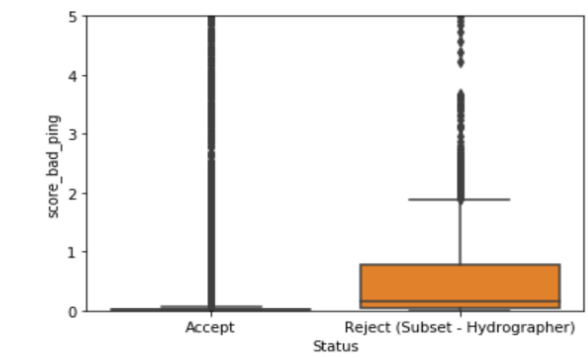

Fig. 15: The bad ping detetection feature depending to the status

2) ML models

After this data analysis, we have selected the more discriminating features for our detection problem, as working hypotheses in classical ML algorithms. We also gave training data from hydrographer editing as seen in the Fig. 16. We chose three different techniques, often used in machine learning literature, as potential classifiers:

- Logistic regression;
- Random forest ;
- Gradient boosting (XGBoost).

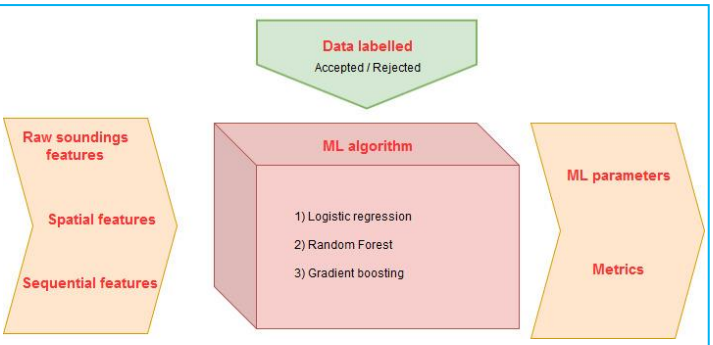

Fig. 16: Global workflow of our supervised classification

The logistic regression aims to predict a binary target by estimating the parameters of a logistic model which will be a linear combination of our features [19]. This method is simple to implement, efficient with small or big data, but the risk of underfitting is important.

The random forest is an ensemble learning method used for classification; it works by computing combinations on decision trees [20]. Although it is harder to implement than regression logistic model, while needing much more data, the algorithm is still explainable because it is based on decision trees.. It needs a proper balancing between the data accepted and the rejected ones [21]. Yet in bathymetric MBES data we have clearly much more accepted data than rejected.

The gradient boosting is a gradient descent (iterative method used in optimisation for finding minimum in a mathematical function) combined with boosting method (machine learning ensemble meta-algorithm). The idea is to iteratively combine weak learners into a single strong learner [22]. This technique is very fast, powerful but the method provides results that cannot be easily interpreted.

3) Metrics and results

These metrics and results were applied on the same dataset with a train/test ratio of 70%/30%.

The different algorithms were compared using the F1-score (as presented in [23]). In the ML workflow the aim is minimizing false positives results, because it is most important to assure that a sounding accepted by a hydrographer will not be rejected by the ML algorithm. Typically, we don't want to filter out any isolated pertinent information that could be a wreck or an obstruction. For that reason, we want the accepted/rejected prediction (pred) score to be the smallest (number in red in table 2, 3 and 4).

The table 2 shows the metrics and results for the logistic regression model. It shows that the logistic regression model performs poorly and has a high risk of underfitting. This model works well when there are very few outliers in the data. The accepted/rejected pred score is more than 2% of data, this score appears to be too important for safety of navigation. A discussion about an acceptable maximum score needs to be conducted within this scope of navigation safety.

TABLE II. LOGISTIC REGRESSION RESULTS

F1-Score = 0.58	**Accepted pred**	**Rejected pred**
Accepted	269 162	5 858
Rejected	17	570

Table 3 shows the metrics and results for the random forest model. The results given by this algorithm is much better than the logistic regression. The accepted/rejected pred score is around 0.001% of data which gives us a greater trust in the prediction. The rejected/accepted pred is greater than the previous algorithm.

TABLE III. RANDOM FOREST RESULTS

F1-Score = 0.97	**Accepted pred**	**Rejected pred**
Accepted	275 016	4
Rejected	55	532

Table 4 shows the metrics and results for the XGBoost model. The results given by this algorithm are very close to the random forest algorithm. The accepted/rejected pred score is around 0.001% of data. The rejected/accepted pred is lower than random forest algorithm. The F1-Score is the higher for this algorithm.

TABLE IV. XGBOOST RESULTS

F1-Score = 0.98	**Accepted pred**	**Rejected pred**
Accepted	275 015	5
Rejected	35	552

In the literature another approach is found in [24], CARIS has implemented a deep learning method for soundings classification and filtering. This technic needs to be tested on a benchmark dataset to measure performance.

V. CONCLUSIONS

Throughout this paper we have presented different outlier detection algorithms. Many of them used fixed heuristic to apply their methods. These kinds of technics are hence difficult to generalise on all data, due to the inherent variability of the dataset (as of the morphology, density, acquisition system…). However, these methods can easily generate features that can be used in classical ML algorithms.

This paper also shows initial tries to apply ML algorithms on bathymetric data. In the literature one will find other approaches as who has implemented deep learning methods for soundings features classification and filtering. The binary classification (accepted or rejected predication) perspective may be too strict. The metrics we are trying to minimize seems relevant for our safety of navigation needs. A filter to determine the mean depth position of the outlier could be built to be tighter for outlier below the seabed. Indeed the information below the soundings masttress is not critical for safety of navigation, it is why we can be tougher in this case.

Another perspective would be to change the algorithm or the parametrisation function to the scene described by the metadata or the global morphology of the seabed. The ML workflow would be in two steps; first one we will have a scene detection algorithm and second one we will used the best algorithm and parametrisation depending on the scene described.

ML algorithms associated with outlier filtering algorithms applied to bathymetric data have proven to generate promising results. This combination of tools will surely become parts of the hydrographic data processing tool box. Much more work is needed to reach this level. Amongst others, we need much more testing on many different data sets to ensure robustness and understand the limits of these methods.

ACKNOWLEDGMENT

The authors want to warmly thank all people who have worked on this article. And Shom (New Caledonia antenna) for providing the data used in this article.

REFERENCES

[1] V.J. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies" pp. 1–43, 2004.

[2] F.Y Edgeworth, "On discordant observations" Philosoph Mag. 23, 5, pp. 364-375, 1887.

[3] F.E. Grubbs, "Procedures for Detecting Outlying Observations in Samples" Technometrics, vol. 11, no. 1, pp. 1–21, 1969.

[4] V. Barnett and T. Lewis, "Outliers in Statistical Data" John Wiley & Sons, 3 edition, 1994.

[5] N. Debese, "Bathymétrie - Sondeurs, traitement des données, modèles numériques de terrain" pp.88-125, 2013.

[6] IHO, "IHO Standards for Hydrographic Surveys", Special Publication N°44,5 edition, 2008.

[7] B.R. Calder, L. Mayer, "Automatic processing of high-rate, high-density multibeam echosounder data", Geochemistry Geophysics Geosystems Volume 4, Number 6, 2003.

[8] S. Wang, P. Zhou, Z. Wu, J. Li and Y. Wei, "Detection and Elimination of Bathymetric Outliers in Multibeam Echosounder System Based on Robust Multi-quadric Method and Median Parameter Model", Journal of Engineering Science and Technology Review 11, pp 70-78, 2018

[9] Z. Du, D. Wells, L. Mayer, "An Approach to Automatic Detection of outliers in Multibeam Echo Sounding Data", The Hydrographic Journal N°79, 1996.

[10] H. Hou, L. C. Huff, L. Mayer, " Automatic Detection of Outliers in Multibeam Echo Sounding data", US HYDRO 2001, 2001

[11] N. Debese, R. Moitié "Multibeam echosounder data cleaning through a hierarchic adaptive and robust local surfacing", Computers & Geosciences, vol. 46, pp.330-339, 2012.

[12] D.R. Herlihy, T.N. Stepka, T.D. Rulon, "Le filtrage des sondes erronées dans les sondages multifaisceaux", International Hydrographic Review, LXIX (2), 1992.

[13] K. Bongiovanni, " Outlier Detection for Swath Bathymetric Data Sets", Oceanic Imaging Conference 2000, 200.

[14] R. Toodesh and S. Verhagen, "Adaptive, variable resolution grids for bathymetric applications using a quadtree approach" Journal of Applied Geodesy, Vol. 12, No. 4, 2018.

[15] R. Hare, "Depth and Position Error Budget for Multibeam Echosounding", International Hydrographic Review, Monaco, LXXII(2), 1995.

[16] C. Leys et al, "Detecting outliers: Do not use standard deviations around the mean, do use the median absolute deviation around the median", Journal of experimental social Psychology, 2013.

[17] M. Breunig, H. Kriegel, R. Ng, and J.Sander, "LOF: identifying density-based local outliers" SIGMOD Rec. 29, pp 93-104, 2000.

[18] T. Hou, L. Huff and L. Mayer, "Automatic Detection of Outliers in Multibeam Echo Sounding Data", US HYDRO 01, 2001.

[19] S. Menard, "Applied Logistic Regression", SAGE, 2nd Edition, 2002

[20] T. Ho, "Random decision forests", Proceedings of the 3rd International Conference on Document Analysis and Recognition, pp 278-282, 1995

[21] C. Drummond, R. Holte, "C4.5, class imbalance, and cost sensitivity:why undersampling beats over-sampling". ICML-2003 Workshop: Learning with Imbalanced Data Sets II, 2003

[22] J.H. Friedman, "Greedy Function Approximation a gradient boosting machine", Ann. Statist. 29, no. 5, 1189—1232, 2001.

[23] D.M.W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", Journal of Machine Learning Technologies 2, pp 37-63, 2011.

[24] B. Foster, "Applications of Machine Learning in Hydrographic Data Processing", US HYDRO 2019, 2019.