Multibeam Echosounder Data Cleaning
Through an Adaptive Surface-based Approach

Nathalie Debese

Hydrographic and Oceanographic Service of the French Navy (SHOM),
13 rue du Chatelier BP-30316 29603 BREST CEDEX – France
mail : nathalie.debese@shom.fr

Abstract – Acquisition of bathymetric data with multibeam echo sounder is currently of common use. These systems allow increasing the soundings density while improving the resolution of the measures. Sporadic punctual errors present in large datasets have forced Hydrographic services to modify the whole issue of their validation procedures which were still now essentially turned to manual control. This paper describes, through an hydrographic point of view, a generalization of the algorithm presented at the Canadian Hydrographic Conference in 1998. Local modelization of the seabed – issued from the application of an M-estimator – is robustified with the help of a hierarchical approach dedicated to the automatic search of the optimal size of the estimation area. Its dynamic construction, analog to the one of a descending quadtree, is based on subdivision rules. Their definitions are key points of the new algorithm. Rules are built on dynamic estimations of both statistical and spatio-temporal properties deduced from currently detected datasets. Moreover, this approach enables to reduce the processing time while simplifying the parameters setting. As a confidence level map comes with the algorithm outputs, the systematic control of its results – which is a requirement linked to the hydrographic point of view – is greatly simplified. The aim of this map is to focus the operator to litigious areas as well as to identify areas where hypothesis of the algorithm may not be verified. The performances of the algorithm were estimated over artificial and real datasets.

1 Introduction

Acquisition of bathymetric data with Multi Beam Echo Sounder (MBES) is currently of common use. Echo sounding is deduced from the measurement of the two ways travel time of an acoustic pulse emitted from the ship to the sea bottom and the measurement of the arrival angle of the returned echo. These systems allow acquiring accurate bathymetric data of high resolution along the ship’s track in a wide swath. Nevertheless, data post-processing shows that MBES data sets sporadically contain erroneous soundings. Such errors can come from multiple reflection paths, can be due to bad weather conditions (low noise level ratio), or can result from side lobes or bubbles in front of the transducers. Even if
erroneous data rates are usually weak – less than 0.5% in a previous study [1] – their detection and invalidation are essential in a hydrographic framework as hydrographic services are responsible for the navigation security.

As the time required by manual cleaning procedures can exceed the one spent to data acquisition, automatic cleaning algorithms were developed. In a generic framework, process involved in automatic algorithms can be divided into three principal steps:

1. depth estimation(s) in a given location;
2. optionally, evaluation of the quality of the estimation process used;
3. use of decision rules enabling, when it is possible, to deduce soundings status;

In the beginning of the ninetieths, erroneous detection was based on description of data distributions through weighted mean and standard deviation (C.f. [15] and [4]). In 1995, Eeg [5] proposed to build a statistical test in order to validate the size of the neighbourhood used to detect spikes present in the data sets. With the help of clustering techniques, as the one proposed by Du in [6], who suggests simulating the operator’s decision-making, the size of the neighbourhood was adapted to the data distribution. Following the example of Hou in [9], some approaches have proposed to accurately describe punctual erroneous soundings (for example erroneous pings detection …). Such approaches led them to apply a cascade processing scheme, built on basic data filters. As suggested by Bongiovanni in [10], cleaning process can start at swath level before examining data in geo registered space. Cleaning process was then mainly improved by robustifying the involved modelization process. Robust approaches as those based on M-estimators (C.f. [2], [3]), or kriging-filtering technique [7], or Kalman filter [8] were carried out in order to take into account the effects of the outliers on local estimation processes. However, due to its security aim, the cleaning process still requires a manual validation. In this context, the cleaning scheme, itself, can be robustify by providing a confidence map of the estimation process that lets, in some cases, the operator carrying on the validation (C.f. [8] [11]). The approach described in this paper comes within this scope. By enlarging the approach initiates in [1], it greatly improves manual controls by focusing the operator’s attention on litigious points.

Availability of first deep-water MBES turned the SHOM to retain a hybrid approach while post processing its bathymetric datasets. The cleaning step laid on manual examinations of a set of doubtful soundings obtained from the application of a semi automatic procedure. To this aim, the SHOM (C.f. [1], [2]) has developed an automatic cleaning algorithm based on the analysis of residual values deduced from local modelization of the seabed through an M-estimator.

Tests carried on deep waters data assessed the algorithm’s capacities to reduce operator’s interventions. Nevertheless, tests also pointed out algorithm’s requirements to optimally adjust its parameters to the whole area depicted by the processed data set. It’s the reason why it was not so easy to
use it in very shallow waters featured by high volumes of data. As each algorithm has its own limits, taking advantages of automatic approaches meant focussing the operator’s attention to litigious areas where algorithm met difficulties. Moreover, tests conducted on particular sea bottom types showed the action of one of the two algorithm’s parameters. Such a difficulty can be avoided in deeper waters as the parameters settings can be adjusted using an experimental adjustment procedure. But high volumes of dataset dismissed this approach in shallow waters. On the opposite side, dividing the survey into areas of homogeneous relief, so as to limit the control of the parameters setting on a small part of data set, was not acceptable in a hydrographic framework.

The proposed automated cleaning algorithm addresses both shallow and deep waters bathymetric datasets. It generalizes the principle of the algorithm described in [2] and [3], already initiates in [11]. It takes advantage of the spatio-temporal coherence of the data to locally adapt the size of the window to the relief. In the remainder of this paper, basic principle of the first algorithm is recalled. The second section deals with the retained hierarchical approach. The partition of the survey into areas that fit together is dynamically driven by a set of rules combining both statistical and morphological features of the detected and validated soundings. Section 3 presents the implementation of the algorithm. It comes up to the operator’s expectations, that is to say: (i) automatic cleaning of the MBES data, (ii) identification of the litigious areas (i.e. areas of sensor dysfunctions, or areas of detected soundings clusters potentially assimilated to objects). Finally, the performances of the algorithm are estimated in section 4 over both a semi artificial and three real datasets chosen for their level of difficulty regarding to the manual cleaning process. Before reading the body of the paper some terminology keys points should be shortly underlined. The term detection will refer to a sounding identified as outlier during the cleaning process. The term connexity will refer to existing network edge linking two soundings. Finally, a cluster will point out a small group of detected and connected soundings.

2 The proposed outliers detection algorithm

2.1 Theory
Due to the measurement system, soundings have a random distribution with respect to the “true sea bed”. In general, this random error distribution is considered as a gaussian random process with zero mean and \( \sigma \) standard deviation. Gross errors occurring in MBES datasets come from another random process with unknown features. Considering that the sea bottom relief can be locally described by a given analytical model, robust interpolation techniques allow identifying outliers as soundings not absorbed by the model. Because the application is restricted to 2.5D problem, the retained analytical model expresses the depth \( z \) in \( (x, y) \) location as a polynomial function \( z = f_p(x, y; a) \) where \( p \) the polynomial order and \( a \) the vector
of the model’s parameters. In practice, $p$ is set to 2 or 3. As the local modelization relies on measured soundings, a common interpolation technique like the least-squares estimation can not be used due to the presence of outliers in the dataset. The estimation of the model’s parameters requires the use of a quadratic robust norm. This norm is minimized with the help of an M-estimator. The chosen M-estimator carries out the optimization process through a weighted iterative scheme whose construction relies on the generalized least squares.

The reader will find in [2] and [3] the details of the implementation of the M-estimator applied to the robust bathymetric interpolation. Its algorithmic form is briefly described below in the case of an area filled with M soundings $(x_i, y_i)i = 1 \ldots M$. Let $\hat{z}_i^{(j)}(\hat{\beta})$ be the depth estimated at the iteration $j$ that minimizes the global quadratic error computed as the difference between the weighted observation depth $w_i^{(j-1)}z_i$ and the predicted depth. The robust norm is progressively computed by re-evaluating the weights deduced from the residual values $r_i^{(j-1)} = \hat{z}_i^{(j-1)} - z_i$. The M-estimator we have retained is the Tukey estimator. As an estimation of the standard deviation of the gaussian random process is made available, the weight $w_i^{(j)}$ at iteration $j$ is given by the following equation: $w_i^{(j)} = (1 - (r_i^{(j-1)}/\sigma^{(j-1)})^2) \text{ if } |r_i^{(j-1)}| < \alpha \sigma^{(j-1)}$ and $w_i^{(j)} = 0$ otherwise. This expression introduces a scale factor $\alpha$ that controls the adaptive rejection point of the M-estimator: $\alpha \cdot \sigma^{(j-1)}$. This point is usually obtained from a robust L-estimation of the standard deviation of the gaussian noise: e.g. the MAD estimator. If the mean of the residuals is close to zero, this estimated value is given by $\sigma^{(j-1)} = 1.48 \text{med}_{i}(|r_i^{(j-1)}|)$.

Such an implementation offers two advantages. The first one relies on the use of a weighting function that enables to explicitly give a status to each sounding. The second one is linked to the computation of the rejection point. The use of an L-estimator to adjust the rejection point contributes to turn the semi local robustness of the M-estimator into a global one (C.f. [12]). However, the starting weights are usually set to $w_i^{(0)} = 1$. Due to the semi local robustness of the M-estimator, this does not guarantee to limit the influence of gross errors onto the model. To limit the masking effect of gross errors, an initialisation step is introduced that fits a plane over the neighbourhood $N_{G},(i)$ defined around each sounding – this neighbourhood is defined as a 3x3 window in the beam-ping representation network. Let $\hat{z}_i^{(0)} = \text{med}_{j \in N_{G},(i)}(z_j)$ be the estimated plane. Then, the weighting function of the IRLS-estimator enables to initialise the weights $w_i^{(0)}$ from the residual value $r_i^{(0)} = \hat{z}_i^{(0)} - z_i$ - the starting rejection point is set so as to reject the 5 last centiles - the absolute residual values $|r_i^{(0)}|$ being sorted into increasing order.
In this way, by attaching to each sounding a weight $w_i$, accounting for the probability of the observation depth to follow a gaussian law, the optimization scheme lead to classify the soundings into two populations:

- Inliers are soundings with a non-null weight. These soundings are used to estimate the parameters of the model. They are supposed to follow a gaussian law with a higher probability as their weights tend to 1.
- Outliers are soundings whose weight is equal to 0. The process identifies them as sample of a non gaussian noise.

The estimation process ends with both an estimation $\hat{\xi}$ of the outliers rate and an estimation $\hat{\sigma}$ of the standard deviation of the gaussian noise that affects the validated soundings.

Except for the internal parameter introduced to guarantee the convergence of the process, this algorithm (C.f. [2]) only requires the setting of two parameters: the sensitivity factor $\alpha$ and the size $L$ of the observed area. In normal conditions, the choice of the sensibility factor is not critical. It can be set to a theoretical value of 6, which optimizes gaussian estimation if there is no outlier.

### 2.2 Hierarchical adaptative approach

The approach, described above, is implicitly based on the hypothesis that the relief of the observed area can be approximated by a second – or third - order polynomial while taking into account an *a priori* acceptable noise. To be more flexible, while minimizing the risk inherent to a non appropriate parameters setting, the initial approach must be generalized by integrating an automatic search of the optimal size of the observed area.

Requirements for implementing an adaptative approach are depicted on figure 1. The artificial surface was built on a raw dataset by preserving the spatial distribution of the data when the depth of each sounding was analytically re computed. The sea bottom was obtained by adding four functions of the following form $z = k \exp(-((x-x_0)^2 + (y-y_0)^2) / \lambda)$ - parameters $(k, \lambda)$ were set to $(10, 100)$, $(20, 100)$, $(30, 100)$ and $(40, 100)$ in the example bellow. Inadequacy of the cell size is conveyed by soundings that are statistically detected as outliers and clustered. Clusters of detected soundings – which are mainly grouped on cell’s border – come from the intrinsic limitations of the model. Clusters on figure 1 are obtained by applying the approach described in §2.1 and setting an inadequate cell size of 10 meters.
Figure 1: Inadequacy of a non adaptive approach

The adaptative approach aims at finding the optimal cell’s size $L$ for each part of the surveyed area. Finding the optimal cell size is dictated by the two following and opposite requirements:

- Wider cells improve statistical criteria used to classify the soundings into two populations;
- On the opposite side, the cell size has to be limited in roughly seabed so as to guarantee the adequacy of the quadratic model. Otherwise, valid soundings will be detected as outliers regarding to their high residual values.

It is the reason why the initial approach is now coupled with a multiresolution strategy to account for the level of details allowed by the model. The retained strategy is based on a descending approach. The observed area is divided as long as the model can not absorb morphological details of the seabed. Due to its building way, it mainly retains wider areas. For a given area, the subdivision process is stopped as soon as rules controlling some working hypothesis are validated. Thus, the main difficulty is to define reliable rules that allow deciding on the validity of the robust modelization. Description of these rules will be discussed in the following section. The subdivision strategy lay on two points:

1. The finest partition of the survey area is defined by the mean of a regular 2D grid of square cells whose size are set to $L_0$;
2. These primary cells are fit together so as to build a hierarchical partition composed of $n = 1, \ldots, N$ levels (i.e. in this quadtree structure a mother cell is divided into four child cells of equal size). The partition of level $n = 1$ is defined as the grid of $L_0$ cell size. The size of the cells at level $n$ is equal to $L_n = 2^{n-1} L_0$;

This adaptative procedure only introduces two new parameters: $L_0$, the cell size of the primary grid and $N$, the number of levels of the hierarchical partition. Let $M_0$ be the minimal number of soundings
required to robustly estimate the parameters of the model fitted over a given observed area. \( L_0 \) can be empirically deduced from the features of the statistical distribution of the soundings (C.f. §3.1). A guess value for \( L_0 \) can be proposed to the operator. The number of hierarchical levels is not crucial as soon as it can be set by excess. These excess drawbacks will be confined to the processing time - because of useless estimations perform at higher levels.

### 2.3 Rules used to validate the modelization

This section focuses on the contextual rules used to *a posteriori* control the validity of the parameters estimated over the current observed area. These rules are based on information issued from the robust modelization of the seabeds the current observed area contains. Whenever the set of available rules do not end to take a reject decision, while the hierarchical level \( n \) is greater than 1, the area will be systematically divided keeping on, recursively and independently, estimating the relief of the four child areas. In the opposite case, the examination of the area will be considered as complete. The decision rules retained gather several aspects:

- Purely statistical aspects based on estimations of \( \hat{\sigma} \) and \( \hat{\tau} \), that respectively represent the noise standard deviation and outliers rate;
- Temporal aspects that takes advantages of covering swath areas and thus convey the concomitance of the detected soundings;
- Spatial aspects that evaluate both size and distribution of the clusters of detections;

The necessary and sufficient condition presiding over the rejection of the current estimation is that, at least, one of the five invalidation rules described above is positive. In practice, so as to reduce the processing time, this invalidation rules will be applied by increasing computation cost. As stated above, these rules assume that each of the preceding one is negative.

#### 2.3.1 Invalidation rule I

This statistical rule is based on soundings validated by the current modelization. This rule will give a positive answer if the standard deviation of the noise \( \hat{\sigma} \) is greater than a maximal value \( \hat{\sigma}_{\text{max\ a\ priori}} \) defined.

#### 2.3.2 Invalidation rule II

This statistical rule is based on soundings validated by the current modelization. This rule will give a positive answer if the outliers rate \( \hat{\tau} \), is greater than the maximum rate assumed in normal survey’s conditions, \( \hat{\tau}_{\text{max}} \).
This spatial rule deals with the relative size of the outlier’s clusters. Its introduction is linked to the remarks made on figure 1. When the analytical model is not appropriate to the seabed, soundings detected as outliers by the M-estimator are gathered in clusters. This rule consists in explicitly identifying the clusters of outliers. Locating the clusters means finding the connected components. In other words, this means introducing an oriented graph between soundings – as it operates on point clouds such a network was not recovered by the previous rules.

An efficient detection clustering requires to take into account both spatial and temporal connexities. The temporal dimension relies on the acquisition time of each measure. This comes down to build a graph that connects the soundings regarding both their number of beam and ping. Due to navigational aspects (yaw or roll effects...) the spatial proximity of the soundings can be linked to the one inferred by the temporal connexion. Taking into account the temporal connexion enables to introduce virtual soundings corresponding to sensor’s blanks. As these soundings reflect sensor perturbations during depth measurements, their introduction will catalyse building of significant connected components. The algorithm that builds the connected components makes use of a recursive procedure based on a front propagation technique in an adjacency-graph (see [13] to a detailed description of a similar approach).

The process that controls the cluster building is depicted on figure 2. The clusters of the detected soundings are firstly built from the connexion implicitly defined by the sensor. As mentioned above, the nodes of this graph include the blank soundings. During the connected components building process, some of these virtual nodes contribute towards the same way as those soundings detected by the modelization. In this aim, the virtual nodes, that are located at less than $d_v$ edges from the nearest native detected sounding, are temporarily labelled as detected soundings too. By taking into account difficulties encountered by the MBES to measure depth, connections are established between clusters. This procedure reinforces the consistency of clusters appearing during acquisition difficulties.

In a second step, these detections – labelled according to their belonging connected component - are projected into a classical DTM built over the current observed area. The sole objective of this DTM is to introduce a spatial connexion that potentially extends the previous clusters. The DTM resolution $l_0$ is defined as a fraction of $L_0$, the cell size of the primary grid. This second step, as the previous one, does not link soundings belonging to different swathes. This rule will give a positive answer if, at least, a cluster of more than $A_{max}$ detections is present.

This spatial rule is applied only if soundings issued from different swathes belong to the observed area. Clusters defined above (C.f. §2.3.3) were created from temporal and spatial relationships established
between soundings of the same swath. This rule intends to take advantage of possible spatial coincidences that may exist between clusters issued from distinct swaths. Clusters of detections from distinct swaths that partially match reveal either an object or a modelization error – both requiring a closer lookup.

This rule will give a positive answer if, at least, two clusters are superimposed. That is a pixels-based mapping rule.

---

**Figure 2**: Clusters building
2.3.5 **Invalidation rule V**

If the preceding rules haven’t invalidated the modelization, all the clusters, if some exist, contain less than $A_{\text{max}}$ detected soundings. Unlike real objects or model errors, there is no reason that outliers will occur at a fixed location. Thus, as far as these clusters do not intersect clusters belonging to other swathes, the model must not be rejected. On the opposite side, clusters belonging to areas outside of intersection swathes are still doubtful. One must then reject the model whenever such clusters of noticeable size exist because they may potentially describe an object.

This rule will give a positive answer if, at least, one doubtful cluster, which size is greater than $A_{\text{min}}$, exists.

3 **Inputs and outputs managing**

3.1 **Algorithm parameters**

The parameters of the algorithm belong to three distinct classes: those controlling the hierarchical partition, those controlling the robust modelization, and finally, those validating the modelization.

- $N, M_0, L_0$ : hierarchical partition building
  - $N$, the number of hierarchical levels. So as to have no effect on the quality of the results, this parameter has to be defined by excess (*typ. val. 10*).
  - $M_0$ The minimal number of soundings contained in a primary cell of the quadtree. This value is theoretically fixed by the degrees of freedom of the retained regression model – 6 (*resp. 10*) in the case of a polynomial model of order 2 (*resp. 3*). In practice, while working with an order-2 model, $M_0$ is set to 10 soundings so as to guarantee a minimal redundancy to fit the model.
  - $L_0$, the size of the cells of the primary grid. If an estimation of the mean distance $l_0$ between the can be made available, a guess value of $L_0$ can be computed while taking into account the constraint $M_0$.

- $\alpha$ : control of the robust modelization
  - $\alpha$, the tuning constant used to scale the adaptive rejection point of the M-estimator. In the lack of outliers or model’s shortcomings, its theoretical and optimal value (*i.e. the 95% asymptotic efficiency on the standard normal distribution*) is equal to 7 (the default one). This value refers to the value used to scale the median absolute deviation from the median residual. According to the cleaning requirements, increasing this value will minimize the false detection rate. On the opposite side, its decreasing will maximize the outliers detection rate.
- \( \sigma_{\text{max}}, \tau_{\text{max}} (\text{resp. } A_{\text{min}}, A_{\text{max}}, d_v) \) statistical (resp. spatio temporal) validation of the modelization

\( \sigma_{\text{max}} \) The maximal value of the standard deviation of the noise affecting the valid soundings. This value is set regarding to an \textit{a priori} estimation of the sensor’s precision or an \textit{a priori} estimation of the processing (\textit{typ. val.} the criteria defined by the S44 hydrographic norm).

\( \tau_{\text{max}} \) The maximal detection rate. The retained value is based on the admissible outliers rate in normal conditions survey. As the previous one, this parameter is set from sensor’s knowledge (\textit{Typ. val} 5%).

\( A_{\text{min}} \) The minimal number of samples belonging to a connected component that allows labelling it as cluster (\textit{Typ. val} 3).

\( A_{\text{max}} \) The size from which a cluster becomes an object (\textit{Typ. val} 9 soundings [14]).

\( d_v \) The number of edges in the temporal graph describing the influence radius of detection. This radius allows selecting temporarily virtual nodes as detection while the clusters building (\textit{Typ. val} 1).

In fact, the only parameters requiring a particular attention are both \( \sigma_{\text{max}}, \tau_{\text{max}} \), which magnitudes represent the survey conditions, and \( A_{\text{min}}, A_{\text{max}} \), which setting implicitly conveys the morphological details that need operator attention.

### 3.2 Cartographic cleaning track

The proposed algorithm provides, together with the set of the valid soundings, a map depicting the results of the cleaning process. This map enables to fully take advantage of the automatic procedure by focussing the operator attention to litigious areas – which typically described either objects or shortcomings of the model.

On this map, observed areas defined from hierarchical levels of size, \( n \), greater than 1 have received a valid status during the cleaning process. All of the soundings detected on these areas were identified as outliers. The algorithm diagnostic can thus be considered as confident. A color table is used (\textit{C.f.} table 1 cases 1, 2, 3) to classify the valid observed areas depending on their numbers of detections and swaths contained.

An area, rejected from the deeper hierarchic level (\textit{i.e.} \( n = 1 \)), is presented to the operator together with the rule leading to this decision. The retained color table (\textit{C.f.} table 1 cases 4, \ldots, 9 and 10) enables to distinguish two populations:

- Areas rejected by rules I or II (\textit{C.f.} table 1 case 10). An operator must systematically control these areas as they potentially may contain punctual errors that were not detected. This class also
includes areas that were not observed because of their insufficient number of soundings (i.e. less than $M_0$).

- Areas labelled as doubtful by rules III, IV or V C.f. table 1 cases 4, ..., 9). Depending on the followed scenario, these areas may require manual validations. Keeping in mind that the main objective of the algorithm is to simplify the operator’s task, an accurate classification was proposed. On one hand, this classification identifies areas according to their content (i.e. cluster, object or clusters superposition), and on the other hand according to the sign of the outliers residual value. Moreover, the set of the detected soundings attached to these areas is thresholded before being presented to the operator. The threshold $S_{det}$ is built from the detection rate computed, for each sounding, from the results of the N previous estimations.

This classification acts on the operators requirements. Litigious areas must be quickly identified. Plotting the ship’s tracks over this map is helpful to reveal systematic errors. Moreover, the localisation and classification of doubtful areas will help the operator to localise clusters potentially depicting objects.

<table>
<thead>
<tr>
<th>Accepted cells</th>
<th>The number of outliers is less than the one require to define a cluster (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No cluster detected – one swath (2)</td>
</tr>
<tr>
<td></td>
<td>Include more than one swath – No superimposed clusters – No clusters contained in a valid area (3)</td>
</tr>
<tr>
<td>Doubtful cells</td>
<td>A cluster detected – residual values are below the estimated surface (4)</td>
</tr>
<tr>
<td></td>
<td>A cluster detected – residual values are above the estimated surface (5)</td>
</tr>
<tr>
<td></td>
<td>An object detected – residual values are below the estimated surface (6)</td>
</tr>
<tr>
<td></td>
<td>Two superimposed clusters - residual values are below the estimated surface (7)</td>
</tr>
<tr>
<td></td>
<td>Two superimposed clusters - residual values are above the estimated surface (8)</td>
</tr>
<tr>
<td></td>
<td>An object detected – residual values are above the estimated surface (9)</td>
</tr>
<tr>
<td>Rejected cells</td>
<td>The Detection rate or the estimated standard deviation are abnormally high (10)</td>
</tr>
</tbody>
</table>

*Table 1: Features of the processed real datasets- algorithm settings and results.*

### 4 Evaluation of the algorithm

#### 4.1 Artificial datasets

The algorithm was tested over three artificial datasets built from a real dataset of about 500 000 soundings acquired by an ATLAS Fansweep 20 MBES in 1 to 7 meters water depths.
Figure 3: Visualization of the coverage areas between adjacent FANSWEEP20 swathes.

Figure 3 depicts the ship’s tracks and coverage areas of adjacent swathes. Points attributes – that’s to say their coordinates (x, y), beam, ping and line numbers – were kept. Only the depth were re-computed either using an analytical expression or from a bilinear interpolation of a smooth DTM describing a real seabed.

<table>
<thead>
<tr>
<th>S0</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadric</td>
<td>Surfaces built from 4 exponential functions</td>
<td>Bilinear interpolation of real seabed</td>
</tr>
</tbody>
</table>

Figure 4: Artificial datasets

The MBES noise was modelized using a gaussian distribution with a 0 mean and $\sigma^2$ standard deviation and added to the surfaces. Punctual errors were injected to the datasets using a uniform law with variable rates.

Tests are based on:
• The error detection rate: defined as the ratio of the number of detected erroneous soundings to the total number of erroneous soundings;
• The false alarm rate: defined as the ration of the number of valid detected soundings to the total number of detected soundings;
• The accepted cells rate: depicts the ratio of the task automatically performed by the algorithm that is no longer of the operator’s scope;
• The rate of the cells rejected: and labelled according to the classification described above (§3.2).

A gaussian noise ($\sigma = 5\text{cm}$) was firstly added to S0, S1 and S2 (C.f. Fig. 4). Punctual errors were then added according to a rate that was increased from 1 to 9%. The running parameters were set to: $L_0 = 1\text{m}$, $N = 10$, $\alpha = 4$, $\sigma_{\text{max}} = 12\text{cm}$, $\tau_{\text{max}} = 10\%$, $A_{\text{min}} = 4$, $A_{\text{max}} = 9$, whatever the surface tested. The cell size of the primary grid was set according to the density map depicted on figure 5.

![Density maps](image)

**Figure 5: Density maps**

Results of the tests are shown on figure 6. As S0 corresponds to the model assumed, the blue curves convey the optimal running of the algorithm. Increasing the punctual errors rate increases the subdivision’s number as decisions are taken on lower hierarchical levels. When the punctual errors rate is lower than 5% of the whole dataset, the rejected cells rate – that’s to say the cells requiring a posteriori manual control – is lower than 5%. Moreover, tests have shown that the detection rate initially equals to 98.95% still reaches 96.05% for a punctual errors rate set to 9%. On the opposite side, the false alarm rate increases from 0% to 1.59%. As shown on figure 6, the algorithm’s behaviour is different for S2 which contains a rough seabed. The detection rate increases with the punctual errors rate increasing. This is a
consequence of the observation of the lower hierarchical levels induced by the spatio-temporal detections analysis (C.f. figure 7).

Figure 6: Evaluation of the algorithm over S0, S1, S2 datasets. Abscissa axis represents the punctual errors rate.

Figure 7: Classification of the cells by hierarchical levels. Abscissa axis represents the punctual errors rate.
4.2 Real datasets
The examination of the confidence map (see §3.2 for the retained classification) produced by the algorithm is illustrated over three MBES datasets chosen for their seabed features. The main feature of the seabed described by the first dataset is built up by the coralline peaks acquired in 3 to 13 meters water depth. The second dataset is a rocky area which is common near the French Britain coast. The last dataset was acquired during a survey run over deep water depth and selected because of its morphological diversity (a rift in the north, pockmark in the middle).

The sea bottom and associated confidence level maps are depicted on figure 8. The cells size represents the estimation area corresponding to a model’s validation. Cells having a $L_0$ size have been completely observed through the N hierarchical levels. The color table (C.f. §3.2) gathers the cells into three classes: Accepted cells (in grey), rejected cells (in red) and doubtful cells. The morphological features contained in each of these three datasets induce a cell’s subdivision in the vicinity’s of their location. If the slope changes of the abrupt coralline peak require operator’s a posteriori controls, more than 96% of the global survey area was automatically processed. The confidence levels map associated to the second dataset outlines the presence of systematic errors: rejected cells are all located in the coverage area of two adjacent swathes. Despite its wide morphological spectra, the algorithm enables to reduce to 1% the manual control required by the last dataset. As for the second dataset, the rejected cells are due to systematic errors.

<table>
<thead>
<tr>
<th>Soundings number</th>
<th>MBES used</th>
<th>Mean depths</th>
<th>Parameters settings</th>
<th>Number of detected soundings</th>
<th>Accepted cells rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SIMRAD EM3000</td>
<td>3m – 13m</td>
<td>$L_0 : 1m$ Special order cleaning</td>
<td>2255 (0.38%)</td>
<td>96.81%</td>
</tr>
<tr>
<td>2</td>
<td>SIMRAD EM1002</td>
<td>37m -52m</td>
<td>$L_0 : 3m$ Order 1 cleaning</td>
<td>4971 (0.16%)</td>
<td>95.80%</td>
</tr>
<tr>
<td>3</td>
<td>SIMRAD EM120</td>
<td>1790m -4690m</td>
<td>$L_0 : 100m$ Order 2 cleaning</td>
<td>1836 (0.31%)</td>
<td>99.22%</td>
</tr>
</tbody>
</table>

Table 2: Features of the processed real datasets- algorithm settings and results.

5 Conclusions and perspectives
Even if the tests presented above have to be confirmed over a large number of real datasets, the first results are promising. They outline the robustness of the algorithm. Moreover, as a working strategy is provided for its parameters setting as well as the interpretation of its results, it can be easily managed by any operator. The slight over cost induced by the hierarchic modelization validation step is greatly balanced by the reduction of the operator manual control.
<table>
<thead>
<tr>
<th>Digital Terrain Modeling (a)</th>
<th>Confidence levels map (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>depth</td>
<td></td>
</tr>
<tr>
<td>minimal</td>
<td>Accepted cells</td>
</tr>
<tr>
<td></td>
<td>Doubtful cells</td>
</tr>
<tr>
<td>maximal</td>
<td>Rejected cells</td>
</tr>
</tbody>
</table>

**Figure 8:** Confidence levels maps (b) coming from the automatic post processing of three real datasets (a).
References


