Use of a Robust Estimator for Automatic Detection of Isolated Errors Appearing in the Bathymetry Data

Nathalie Debese, SHOM, France

The Service Hydrographique et Océanographique de la Marine (SHOM) has been using multi-beam echo-sounders for about ten years to carry out bathymetric surveys. The principle of multi-beam echo-sounders is based on the sonar technique of channel formation. A sound wave-length, of short duration, is emitted towards the sea-bottom, and then must be recovered by the receiving beam. The propagation of the sound wave gives information on the bathymetry. The sounding not only occurs at the vertical point of the vessel, as for single-beam systems, but on a perpendicular crossing line from the vessel.

The use of multi-beam echo-sounders in bathymetric surveys increases the density of soundings and improves the data resolution. Nevertheless, experience has shown that these recordings include some isolated aberrant data. These errors can be explained by the phenomena of signal reflection, by bad weather conditions (inducing a low signal/noise ratio), or by the presence of aeration outside the transducers. Although there are not that many errors, they nevertheless must be removed, in order to provide reliable bathymetric charts to insure safety of navigation. For example, a SHOM study (DEB 97) indicates that for the SIMRAD EM12-Dual multi-beam echo-sounder the errors amount to less than 0.5 per cent.

The removal of the aberrant measurements is an essential point of the treatment of the bathymetric data. There are two approaches to this problem:

- The first approach is entirely manual. An operator must examine all the soundings. The detection of an abnormal value, in the local bathymetric context, is left to the appreciation of the operator
- The second approach, on the contrary, is entirely automatic. The aberrant values are identified by the treatment of algorithms. The object is to validate a certain number of rules, which have been previously defined

The SHOM has preferred a middle solution, combining those two approaches. The phase of validation is left to the operator, who chooses to invalidate or not, the outliers identified by the algorithms.

This hybrid approach answers the worry of consistency of treatment by the different operators, and offers a good compromise between the time of treatment and the quality of the data discarded.

All the algorithms for the detection of aberrant soundings, resulting from the different SHOM studies, are based on the hypothesis of continuous local bathymetry. Two categories of algorithms can be distinguished. The first category is composed of algorithms inherent to a classification defined '*a posteriori*' from a set of around five million soundings treated manually.

This specific study (DEB 97) has lead to the construction of three algorithms specifically dedicated to data coming from the SIMRAD EM12-Dual deep-sea echo-sounder. The algorithm described in this article belongs to the second category of algorithm. The detection of isolated errors based on a local modelling of the seabed. The retained model is a 'quadric'. As introduced in DEB 98, modelling is applied directly to the raw data through a robust estimator. The robust estimators, that are the easiest to set up, are the 'well-balanced', which are called W-estimators. Among them, we have chosen the estimator of Tukey, because of its adaptability characteristics. So far most of the robust methods, the estimator relies on the residual measure of the information, to identify the potentially aberrant points: a strong residual value indicates a strong deviation of the point compared with the expected model. The suspicious points are labelled as erroneous after comparison with the local results.
The algorithm for automatic detection of isolated errors found in bathymetric data is described below. Its evaluation has been performed on five sets of real data coming from different multi-beam echo-sounders, chosen for the variety of relief measured. The data sets are presented in paragraph 3, the criteria and results of the evaluation are provided in paragraph 4.

Description of the Algorithm

General Principle

The algorithm is based on the hypothesis that there is at least a scale of representation of the marine relief for which a 'quadric' model is possible. As this objective is not easily achievable over the complete geographic zone, it is necessary to split it apart. Under those circumstances, a splitting into square cells of identical sizes (L) has been decided. If the local quadratic model is statistically verified, it is then possible to use the residual information, measured from the real depth and the one estimated by the model, to control the coherence of each sounding against the next one.

$$y = a_5 x_1^2 + a_4 x_2^2 + a_3 x_1 x_2 + a_2 x_1 + a_1 x_2 + a_0 = A \cdot X \qquad (1)$$

When it comes to detecting isolated errors appearing in the bathymetric data, the residual value measured can be attributed to two noise sources, meaning (GAU 96):

- The noise of measurement of the sensor itself, which will be supposed as Gaussian
- The noise from the isolated errors, bound to erratic phenomena, of unknown distribution laws, but non Gaussian

$$\hat{A} = \underset{A}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - A \cdot X_i)^2 \qquad (2)$$

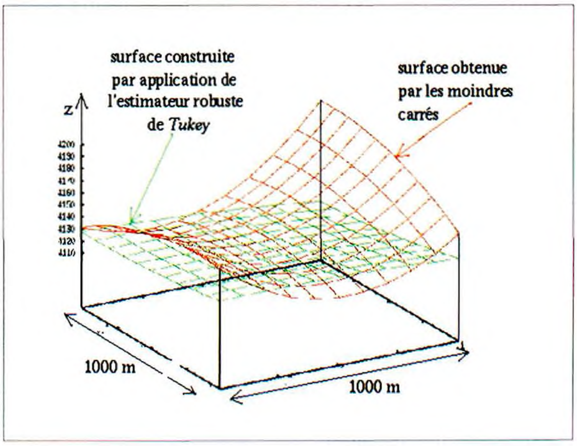


Figure 2.1: Comparison of surfaces, obtained with a robust estimator (Tukey), and with the method of the least squares

In the presence of aberrant points, as is the case here, a classical estimation technique such as the one of 'least squares' (2) cannot be used, as all the points contribute identically to the surface construction. The adjustment of the surface onto aberrant points could lead to discarding valid points, as the case of Figure 2.1. a. A robust estimation technique is necessary to be able to determine the model's parameters.

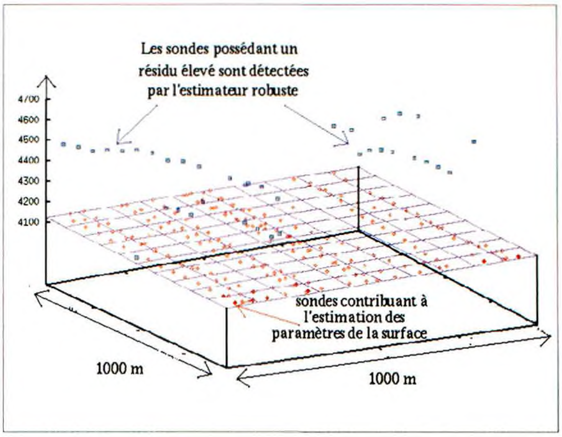


Figure 2.2: The soundings appearing in squares have been designed as erroneous by the Tukey estimator, the ones appearing as lozenges are taken for the parameters estimation of the quadric. The area represented is a square cell of 1000 m which contains about 290 soundings (of which almost 15 % are erroneous)

Main Characteristics of a Robust Estimator

The robust estimation imposes the set up of estimators influenced weakly by the presence of aberrant points – which are commonly called outliers. As outlined by Rousseeuw in (ROU 87), these estimators will not just try to discard the exterior points. On the contrary, from the calculation of the residuals by these estimators, it will be possible to pin point the outliers in the set of data. Contrary to the estimation method based on the least squares technique, a robust estimation procedure will not tend to adjust the whole set of data. In Figure 2.2, the erroneous soundings which stand above the adjusted surface of the valid soundings, do have a high residual value. However, the valid data must comprise more than 50 per cent of the data in order to let the robust estimator adjust them.

In the literature (PRES 92), the robust estimators are commonly grouped into three categories, each one with a specific construction mode.

- **The M-estimators**: are roughly a generalisation of the estimator of the maximum probability. It is generally the type of estimator used in problems of data modelling. The Tukey estimator is one of them
- **The L-estimators**: are constructed as a linear combination of statistics
- **The R-estimators**: are deducted from statistical tests

Definition

The purpose is to represent a set of points pi = (Xi, yi)i = 1, ...N by means of the class of function defined by (1). To take into account the characteristic uncertainty of the measurements, we say that at each abscissa Xi = (X1, X2)i generally fixed (which is true when the model is known in advance). Yi is a random variable, noted Yi. We than speak of sampling size N, to name all random vectors Yi. When the sample is gaussian, the estimator of least squares is the estimator of the maximum probability. Now, let us suppose that we have a sample Yn, with the Yi independent aleatory variable, but with any probability law. Let ρ be the opposite of the logarithm of the probability density, then the M-estimator is the estimator of the maximum probability deducted by minimising the following (3):

$$\sum_{i=1}^{N} \rho(y_i, y(X_i)) \qquad (3)$$

The measurement y and the predicted value y(Xi) being generally bound, the function r can be rewritten to depend on:

$$\frac{r_i}{\sigma_i} = \frac{y_i - y(X_i)}{\sigma_i} = z \qquad (4)$$

- Where the σi is the scale factor associated to the measure Yi.
- Then to minimise:

$$\sum_{i=1}^{N} \rho\left(\frac{r_i}{\sigma_i}\right) \qquad (5)$$

- The function ρ(y) is a function of cost (SOM 96) which must be continuous, symmetrical, and have a unique minimum of 0.

Properties

The class of M-estimators is composed of many elements. Some of the properties of these estimators are presented as follows, in the aim of differentiating them. The robustness of an estimator is appreciated by varying the size of the sample Y1...Yn. The random variable Ê –estimator of the model parameters – depends from Y as well as from N. This notion is close to the continuity one of N of the function Ê. This property (HAM 86) pp40-47 only enables to eliminate the classical estimation procedures. It does not allow the comparison of the robustness of several estimators. Physically, it means that a small disturbance in the sample size must have small effects on the estimator. The influence function of a robust estimator is used to measure the effects of infinitesimal disturbances on the estimator. For a relatively large set of distribution laws L –where L is the probability law of the sample - (HAM 86), the influence function Ψ reflects the disturbance of any point on the statistics E.

$$\psi(z) = \frac{d}{dz}\rho(z) \qquad (6)$$

The collapsing point of a robust estimator is the small fraction of contaminated points which disturbs the estimator. When the estimator Ê is applied to the realisation z = (y1...Yn) of the sample (Y1...Yn), the vector of regression coefficient is obtained Â. We then construct a new realisation t' of the sample (Y1...Yn), by replacing P points by set values, the other points being unchanged. It appears as though we had introduced on purpose P aberrant points. A second vector of regression coefficient Â' is then obtained. The slope is defined as B(P, E, z) by:

$$B(P, E, z) = \sup_{z'} \left| A' - A \right| \qquad (7)$$

The collapsing point corresponds to the minimum number of P points, for which the slope becomes infinite. For the estimator of least squares, only one aberrant point can contradict the obtained regression vector. The collapsing point of this estimator is worth 1/N if N represent the size of the sample. When N tends to infinity, the collapsing point tends to 0, which reveals the extreme sensitivity of the estimator of the least squares to the noise of the measurements. An estimator is as interesting as the collapsing point is high, meaning that it is not very sensitive to obvious errors. To obtain a high collapsing point, like 0.5 –which is the maximum value possible- is easy (ROU 87). In fact, as soon as it is tried to attenuate the sensitivity of the estimator to large errors, the efficiency problems arise: the more robust the estimators, the less efficient (for example, these estimators have no slopes and with minimum variance). In that way, the M-estimators are optimum.

Construction

It has been seen before that as M-estimator Ê minimises:

$$\sum_{i=1}^{N} \rho\left(\frac{r_i}{\sigma_i}\right) \qquad (8)$$

by writing:

$$\psi(z) = \frac{d}{dz}\rho(z) \qquad (9)$$

Ê is then the solution of the equation:

$$\sum_{i=1}^{N}\psi\left(\frac{r_i}{\sigma_i}\right)=0 \qquad (10)$$

Which brings back to the resolution of a M non-linear equations system, generally difficult to resolve. We must then reformulate the problem differently.
To define a M-estimator, each random variable Yi has been assumed based on the probability law

$$p(r_i)\propto\frac{1}{\sigma}g\left(\left(\frac{r_i}{\sigma}\right)^2\right) \qquad (11)$$

Where

$$r_i = y_i - y(x_i, a) \qquad (12)$$

$$r_i = y_i - \sum_{j=1}^{M} a_j x_{ji} \qquad (13)$$

The algorithm of the probability law for a sample of size N is written

$$L(a,\sigma|y_1\ldots y_N)=C-\frac{1}{2}N\,Ln\,\sigma+\sum_{i=1}^{N}Ln\left[g\left(\left(\frac{r_i}{\sigma}\right)^2\right)\right] \qquad (14)$$

Set:

$$w_i = -2\left[\frac{\partial}{\partial(u_i)}Ln\left[g(u_i)\right]\right] \qquad \text{avec} \qquad u_i=\left({r_i}/{\sigma}\right)^2 \qquad (15)$$

W is the weight function, then (16)

$$\frac{\partial}{\partial a_k}L(a,\sigma|y_1\ldots y_N)\Big|_{a=\hat a}=\sum_{i=1}^{N}\frac{\partial}{\partial a_k}Ln\left[g\left(\left(\frac{r_i}{\sigma}\right)^2\right)\right]=\frac{1}{\hat\sigma^2}\sum_{i=1}^{N}w_i\left(y_i-\sum_{j=1}^{M}\hat a_j x_{ji}\right)x_{ki}=0 \quad \text{avec} \quad k=1,\ldots M \qquad (16)$$

$$\frac{\partial}{\partial\sigma}L(a,\sigma|y_1\ldots y_N)\Big|_{\sigma=\hat\sigma}=-\frac{1}{2\hat\sigma^2}N+\frac{1}{2\hat\sigma^4}\sum_{i=1}^{N}w_i\left(y_i-\sum_{j=1}^{M}\hat a_j x_{ji}\right)^2=0 \qquad (17)$$

the usual form

$$\sum_{i=1}^{N}w_i\left(y_i-\sum_{j=1}^{M}a_j x_{ji}\right)x_{ki}=0 \qquad k=1,\ldots M \qquad (18)$$

of the least squares which can be solved by iterative method. The M-estimators still able to be under this form are called w-estimators.

$$\hat\sigma^2=\frac{\sum_{i=1}^{N}w_i\left(y_i-\sum_{j=1}^{M}a_j x_{ji}\right)^2}{N} \qquad (19)$$

(19) is the estimator without skew (or slope ?) of σ2 in the method of reweighted least squares, if, however, N is replaced by N-M.

Robust estimators, that are the easiest to set up, are the W-estimators (ROU 87), or the IRLS-estimators (for iterative reweighted Least Squares). Their iterative construction is based on the generalised technique of Least Squares.

- By applying the least squares a first time, an identical weight is affected to each point, we obtain a first estimation of the parameters:

$$\left(\hat{a}_j^{(0)}\right)_{j=1,\dots M} \qquad (20)$$

- From a first estimation of the residuals and of the weights

$$\left(w_i^{(0)}\right) \text{ et } \left(r_i^{(0)}\right) \qquad (21)$$

- The residuals estimation is used to calculate the next iteration

$$\hat{A}^{(j)} = \arg\min_A \sum_i w_i^{(j-1)} r_i^{(j-1)2}$$
$$r_i^{(j-1)} = \left|z_i - \hat{A}^{(j-1)} X_i\right|^2 \qquad (22)$$

And so on until the convergence criteria is verified.

A particular influence function corresponds to a given W-estimator (HAM 86). This function assigns to each point the corresponding weight at the next iteration, according to the value of the residual. The purpose of this article not being to describe the construction technique of W-estimators, we will simply mention that they are grouped into three classes according to the characteristics of their influence function Ψ:

- Descending
- Abruptly descending
- Re-descending

The usual W-estimators have decreasing influence function that is always positive.
Meaning that from one iteration to the other, there is no point discarded. On the following figure, the form of the weight function is presented for each corresponding class.

The Tukey Estimator: a Particular Case of W-estimator

The Tukey estimator is a particular case of W-estimator. One of its two main characteristics is the influence function of the re-descending type. The associated weight function being truncated, soundings can be discarded from one iteration to the next. The soundings to which are assigned a 0 weight value at the end of the iterative process, are designated by the estimator as potentially aberrant.
The second advantage of the Tukey estimator comes from its adaptive characteristics, giving it the name of 'biweight' estimator. The threshold for discarding the soundings varies from one iteration to the next (23). It depends linearly from the median $r_{median}^{(j-1)}$ value calculated over the total number of residuals, in absolute values, and from α the sensitivity coefficient of the estimator.

$$
\begin{cases}
w_i^{(j)} = \left(1 - \left(\dfrac{r_i^{(j-1)}}{\alpha \cdot r_{\text{median}}^{(j-1)}}\right)^2\right)^2 & \text{si} \quad r_i^{(j-1)} < \alpha \cdot r_{\text{median}}^{(j-1)} \\
w_i^{(j)} = 0 & \text{sinon}
\end{cases} \qquad (23)
$$

Algorithm Parameters

The proposed algorithm simply requires the adjustment of two parameters:

- The sensitivity coefficient of the estimator: α
- The size of the cells (for example the regions): L

The sensitivity factor is a parameter used in the calculation of the discard threshold. As mentioned earlier in 2.2, the soundings with a residual, in absolute value, of a times greater than the median value of the residuals, will not be taken into the surface estimation for the next iteration: they make up the set of the erroneous soundings.

The cells size, L, must be chosen, so that the relief of each cell allow a statistically quadric model.

The behaviour of the Turkey estimator has been evaluated with an artificial data set, constructed by superimposing a blank noise over a constant landform. This test has clearly shown that, for a blank noise, which is supposed to correspond to the noise of measurements of a multi-beam echo-sounder, and despite the adequate cells size - which was necessarily the case for a flat plane - the estimator has assigned valid soundings to the potentially erroneous batch. This behaviour is essentially due to its adaptive characteristics.

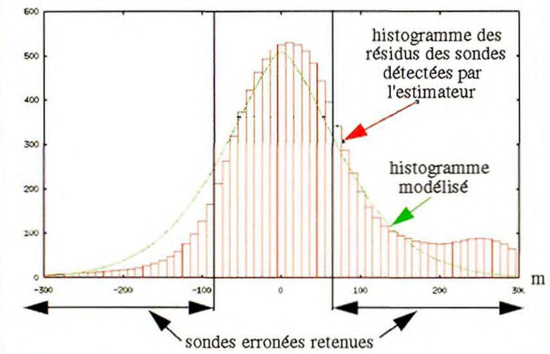


Figure 2.2 b: Illustration of the shape of the ponderation function for each class of W-estimator

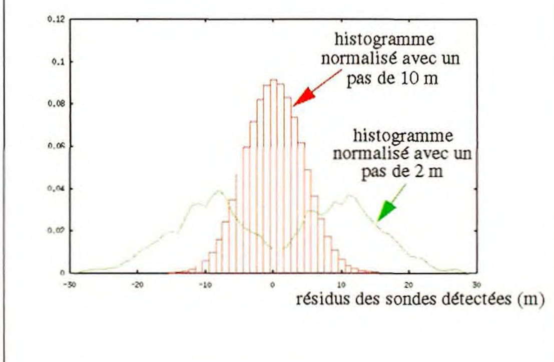


Figure 2.3.a : The automatic detection of the histogram mode, centered on the origin, is used for the erroneous soundings data set. In the example above, the width of the shape is set at 2/3 of the length

Therefore, even when the cell size, L, has been correctly defined, the soundings corresponding to the central mode of the residuals histogram, do not constitute real errors. It is the counter-balance of the adaptive characteristics of the detection process. It is then of prime necessity to introduce independently to the process, a global parameter defining the minimum residual value for the potentially erroneous soundings. This value can be more precisely defined from the intrinsic characteristics of the echo-sounder (for example the precision).

It is also possible, as illustrated by Figure 2.3.a, to adjust the threshold value more finely, by visualising the histogram for residuals of the soundings detected by the estimator. For this, the discretisation step of the histogram must be properly chosen (Figure 2.3.b). It must be big enough compared to the echo-sounder imprecision, in order to get an histogram with a principal mode centred on the origin. In certain conditions, a probable value can be proposed to the operator. This value, if the operator agrees, can be automatically calculated by the modelling of the central mode of the histogram (on Figure 2.3.a, the histogram is figured in dash line).

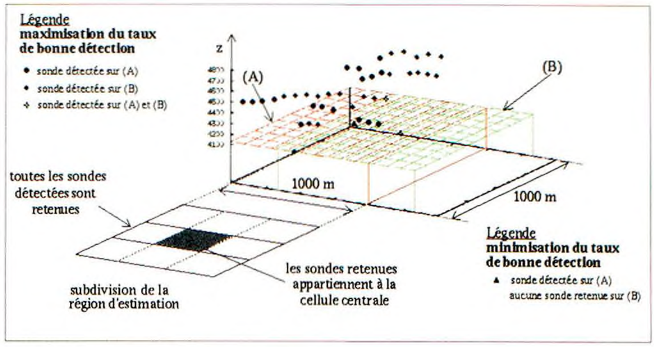


Figure 2.4.b: Illustration of the covering modes

Functioning Mode of the Algorithm

The described algorithm works following 2 modes:

- A fast mode, in which each sounding is looked at only once
- An overlapping mode, allowing several tests on each sounding

The fast mode operates by a sequential and separate investigation of the adjacent cells LxL.

In the overlapping mode, the cells LxL partially overlap (2/3 in x, 2/3 in y). Each sounding is examined several times by the investigation. This mode partially deviates the weaknesses of the quadratic model, like where the landforms need a higher order model. This increases the probability of finding a quadratic form including more soundings, meaning we increase the probability to adjust a quadric to the relief. Therefore, the probability to find erroneous soundings in areas of strong relief increases too. Also, this mode gives a grade to each erroneous sounding translating the degree of doubt. The highest grade is given to the soundings overlooked N times and N times detected as erroneous. In practise, the implementation of this idea implies a subdivision of each region. Each cell which defines the estimation surface is made of 9 sub-cells coming from a finer grid (cf Figure 2.4.b).

There are several possible options with the overlap method. They depend on the reweighting of each cell. By only retaining soundings of the central cell, the excessive detection rate is minimised: only soundings which have an isotropic information distribution are retained. On the contrary, if for each of the nine sub-cells all the detected soundings are retained, the rate of proper detection is maximised, the number of observed erroneous soundings being potentially increased.

Description of the Data Sets

Five multi-beam profiles have been chosen to set up and evaluate the performances of the proposed purification algorithm. The data were provided by three multi-beam echo-sounders:

- The SIMRAD EM12-Dual, deep-sea echo-sounder (200 - 1200 m):characterised by 162 beams and an aperture of 128 °
- The Thomson-Lennermor shallow-water echo-sounder (5 - 300 m): characterised by 16 beams, and an aperture of 75 °
- The SIMRAD EM3000, very shallow water (0 - 150 m): characterised by 127 beams, and an aperture of 140 °

The data coming from the first three profiles are provided by the EM12-Dual echo-sounder onboard the vessel *l'Espérance*. The first profile corresponds to a sub-marine mount, reaching 2,400 m. South of this mount is a plain lying at a depth of about 2,700 m. North the slopes get steeper to reach depths of 3,900 m. The erroneous soundings are very outstanding. They are mainly found in the central beam of the port-side echo-sounder. Difference measurements represent more than 10 per cent of the water column.

The data of the second profile come from a transit route between Brest and the Azores. Water depths vary form 3,600 to 4,400 m. Here again, the erroneous soundings are easily picked out. The external and central beams of the port-side sounder show the highest error rates.
The data of the third profile come from a zone with water depths varying between 1,800 and 3,300 m. The bottom form makes the purification process delicate. Excepting the beams corresponding to the change in detection mode phase amplitude, the error rate for each beam is fairly constant.
The fourth profile has been provided by the Lennermor echo-sounder. The relief examined revealed a submarine dune with varying depths of 28 to 35 m. This echo-sounder has a maximum error rate for the lateral beams. The data have been treated among a set of 5 profiles of 180,000 soundings with overlapping passes, to balance the small sampling rate of the echo-sounder in its transverse axis.
The data of the last profile have been supplied by the very shallow water EM3,000 echo-sounder. The 180,000 soundings have been acquired during less than 2 minutes over a zone having very little slope between 3 and 9 m. In this profile, the beams corresponding to the change in detection mode phase amplitude, are here again where the highest error rates are found.
These data sets have been treated manually to make up a reference set: each sounding has been systematically examined.

Data Characteristics of the Abyssal Plain (First Profile)
On this profile 0.71 per cent of the soundings are erroneous or questionable. The hydrographer has not found any difficulty in treating this data set: less than 15 per cent of the erroneous soundings are classified as dubious. Errors of classes 3-4 concern almost 5 per cent of the erroneous soundings of the central channels of the port-side sounder. On the contrary to the data taken over the mount, the lateral beams of the port-side echo-sounder show a high error rate, close to 30 per cent.

Characteristics of the Data from the Mount
The main characteristics of the profile are presented in the following table.
The erroneous soundings (including the dubious soundings) represent 0.39 per cent of the total soundings of the profile. They are mainly found in the beams 60 to 81 of the port-side sounder, for which the error rate reaches 5 per cent. The essential characteristics of the aberrant soundings is that they are statistically deeper. On the example of the cycle 15 609, the difference represents 11 per cent of the water depth. As shown by the small percentage of dubious soundings, being less than one third of the erroneous soundings, the manual treatment of this profile was easy. Relief of the profile are presented on Figure 3.1.a. Each sounding is figured with a point with a colour code. The erroneous soundings are figured in black, the dubious soundings in blue. As illustrated on this figure, the erroneous are all clustered.

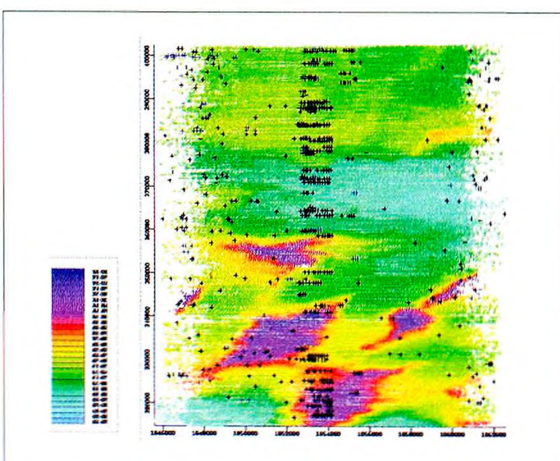


Figure 3.1.a: Qualitative description of erroneous and uncertain soundings in profiles acquired from EM12

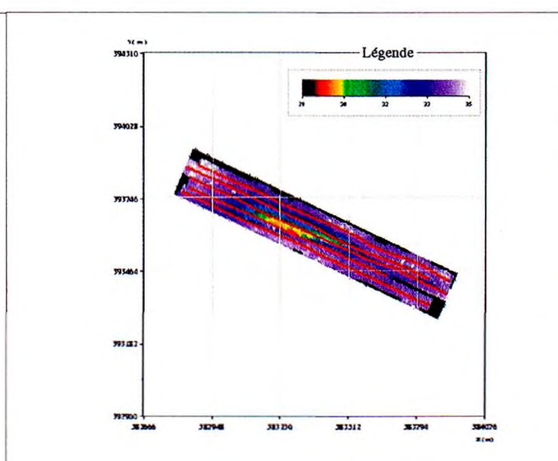


Figure 3.1.e: Description of the relief with the shallow-water echosounder

Characteristics of the Data Acquired on Zone

The rate of erroneous soundings is equivalent to the one of the two preceding ones (about 0.5 per cent). However, it seems that almost 60 per cent of the erroneous soundings have been classified as dubious. The example of cycle 4049, presented on figure 3.1.d illustrates particularly well the difficulties arising to the hydrographers attempting to validate this data set. The curve representing the rate of erroneous or dubious soundings as a function of the beam index does not outline any difference between the beams from the port-side and beams from the starboard side, as was the case for the data coming from the plain. Only looking at the erroneous soundings, we notice that the beams with the highest error rate are the ones corresponding to the change of detection mode: phase - amplitude.

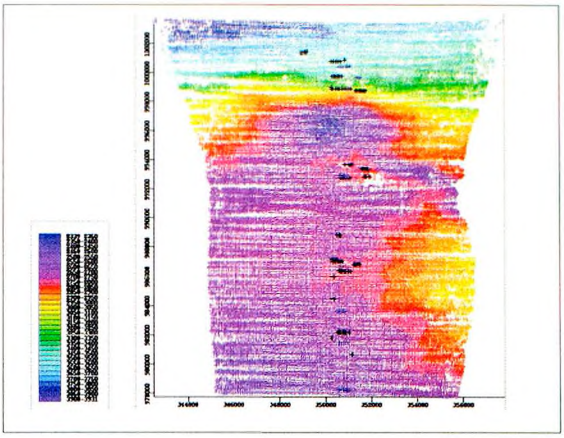


Figure 3.1.b: Characteristics of the erroneous soundings from the abyssal plain profile

If the erroneous or dubious soundings of this profile are not as characteristic as the ones of profile 1, they appear clustered as shown on Figure 3.1.a.

Characteristics of the Dune Data

The following table presents the main characteristics of the data provided by the Thomson-Lennermor echo-sounder. The relief of this zone corresponds to a submarine dune lying at about 30 m of depth (cf Figure 3.1.e). For this Lennermor echo-sounder the error rate is maximum on the lateral beams, as indicated on Figure 3.1.f. Because of the under-sampling along the transverse axis, the manual and automatic treatments have been performed on 5 parallel profiles, in order to increase the number of erroneous soundings detected.

Characteristics of the EM3000 data

The bathymetry recorded by EM3000 are presented on Figure 3.1.e.

On Figure 3.1.g, the error rate has been represented as a function of the beam index. As for the EM12-Dual, the central beams of the EM3000 are less reliable, with almost 3 per cent of erroneous soundings detected- and this particularly at the change of detection mode (ie amplitude - phase).

The evaluation of the algorithm is based on the two following criteria:

- the rate of good detection, defined as the ratio of detected erroneous soundings to the number of soundings to be detected.

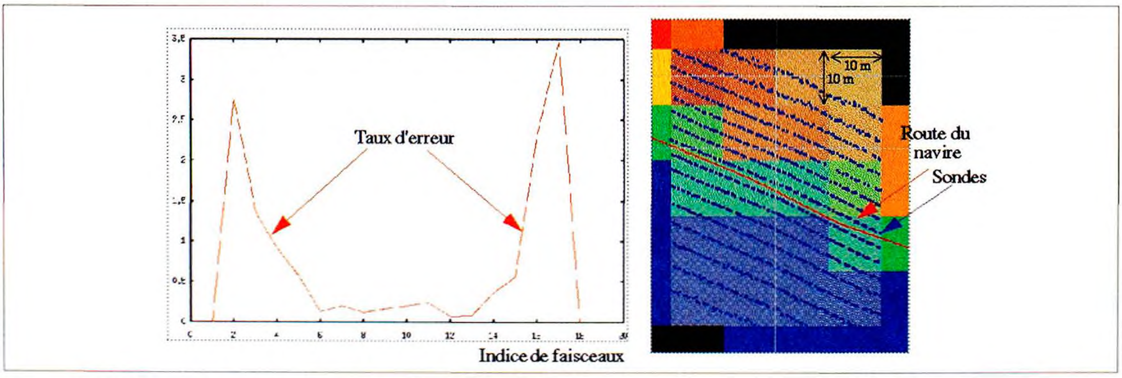


Figure 3.1.f: Characteristics of the erroneous soundings and soundings distribution of the dune profile, and with the Lennermor echo-sounder

- the rate of excessive detection, defined as the ration of valid soundings detected to the number of soundings to be detected.
- These error rates are calculated in comparison to the reference set.

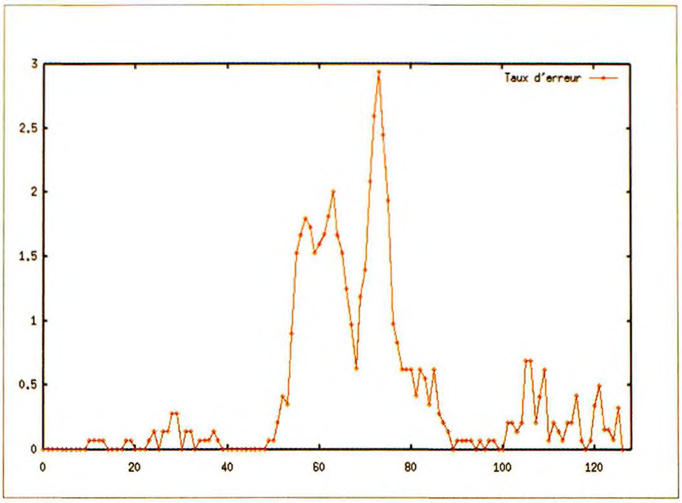


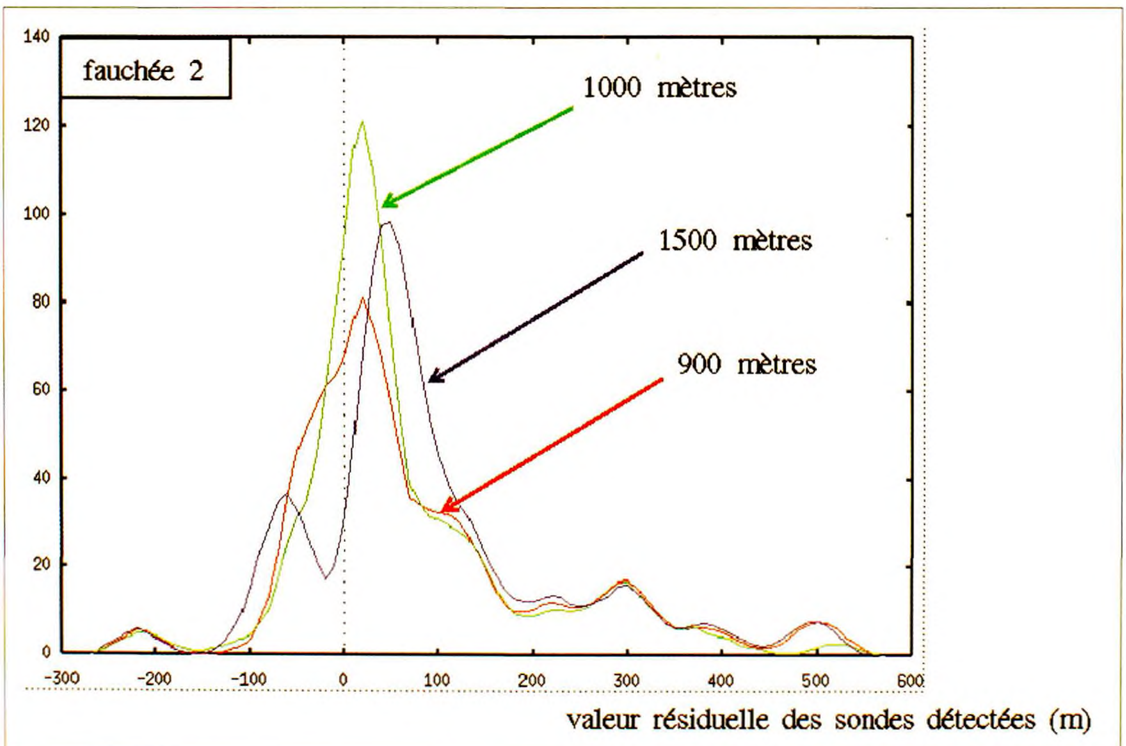
Figure 3.1.g: Characteristics of the erroneous soundings from the profile with the EM3000

Evaluation of the Algorithm

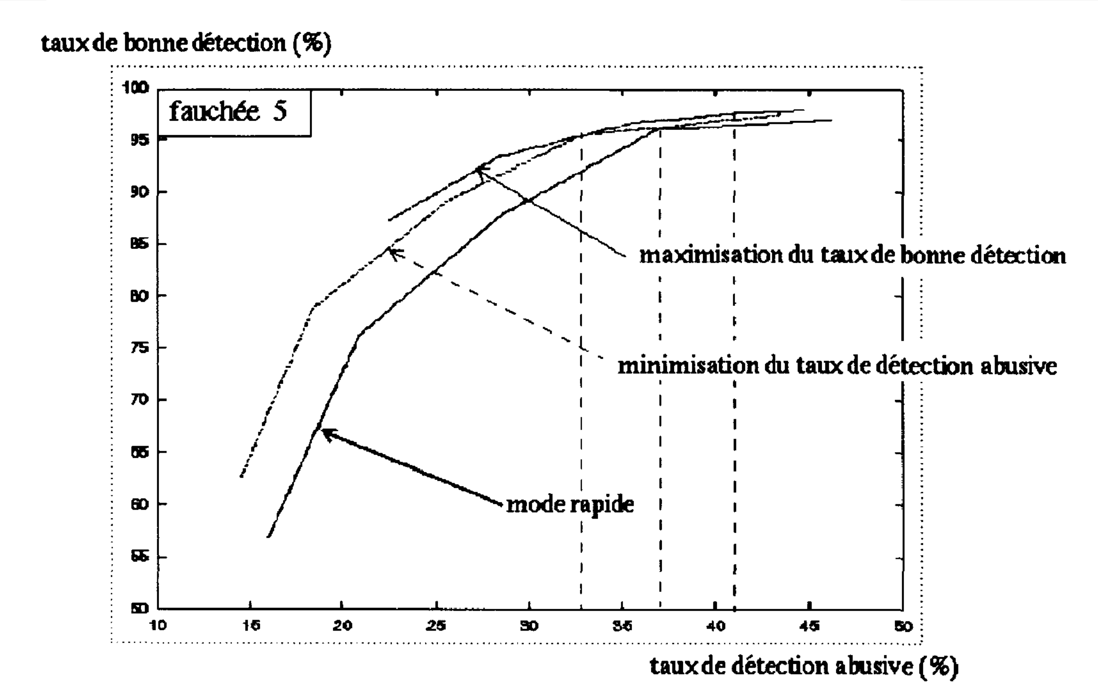
Parameters Adjustment

The optimum size of the cells (ie the maximum scale for which the local approximation of the bathymetry by a quadric is valid) has been determined for each data set. For profile 2, a cell a 1000 m side has been estimated as statistically correct (Figure 4.1.a). Indeed, for this size, the histogram for the residuals of the detected soundings by the estimator, has a narrow principal mode, centred on the origin, provided that an adequate step had been taken for the construction. Considering the variety of relief studied here, the results available in Table 4 can in practise be used to adjust the size of the cells.

A visual analysis of the histogram must *a posteriori* allow to control of (and validate) the size of the cells in order to seize a wider range of relief. The consequences of an inappropriate cell size can be observed on Figure 4.1.a, with the shape of the main mode of the histogram. A too small size of cell increases the



4.1.a.: Histogram of the residual soundings for different cell sizes, detected by the estimator



4.1.b.: Representation of the rate of good detection, function of the overestimated detection. The vertical dash lines are the results from profile 5 on Table 4

probability to retain errors due to the sounder's noise. In the opposite case, a too large cell size results in a smoothing of the relief. The inadequacy of the gridding results in a subdivision of the principal mode of the histogram. However, this heuristic criterion becomes inoperative when it is for the Lennermor echosounder, because of the under-sampling along the transversal axis. The second parameter of the algorithm is the sensitivity factor of the Tukey estimator. Figure 4.1.b presents the rate of good detection versus the rate of poor detection. The three lines correspond to each operating mode of the algorithm, and have been obtained for different values of sensitivity factor a (ie a varies between 6 and 14 with a step of 2). The smaller the sensitivity factor, the bigger the rate of poor detection (ie the treatment becomes sensitive to small disturbances). In practice, the sensitivity factor is set between 6 (shallow waters) and 10 (deep sea). The third parameter is the global threshold applied to all the residuals of the detected soundings of the estimator.

Results

The detection rates of the algorithm for the 5 reference profiles is presented in table 4. The thresholds applied to the residuals of the detected soundings of the estimator have been automatically determined from the central mode of each histogram (ie the width being equal to 2/3 of the height); the resulting thresholds are roughly the same as the ones which would have been visually deducted from the central mode of the histogram. For the EM3000 and the EM12-Dual systems, the choice of operating mode of the algorithm only depends on the constraints defined by the operator. For the profiles 1 to 3, the overlapping mode results in an increase of 5 per cent of the rate of good detection, while the rate of poor detection remains at 25 per cent (except for profile 3). The 108,000 soundings of profile 1 have been acquired during 4 hours with the EM12-Dual in depths of 4,000 m. The algorithm used in overlapping mode results in the detection of 93 per cent of the erroneous soundings in 2 min. 25s. (on a SUN Ultra-sparc station). For profile 5, the detection rates remain the same. The algorithm used in the fast mode does detect 96 per cent of the erroneous soundings inside the set of

178,000 data, and this in less than 30 seconds. On the contrary, it is absolutely necessary to use the algorithm in the overlapping mode for the Lennermor sounder, in order to detect 88 per cent of the erroneous soundings. The spatial disparity of the data is then partially mitigated.

Conclusion

The algorithm proposed here is based on the robust estimator of Tukey to detect the isolated errors present in the bathymetric data.
The adjustment of only two parameters is the undeniable advantage of this estimator. More, the pertinence of one of them, the cells size, can be controlled a *posteriori.*
The second advantage of the algorithm is that it can be operated in different modes. In the fast mode, for very shallow waters data, a factor of 4 is maintained between the time of acquisition and the time of treatment. In overlapping mode, it is possible to detect 98 per cent of erroneous soundings, with 25 per cent of poor detections. If the purpose is to minimise the rate of excessive detection, it is possible to detect 94 per cent of erroneous soundings with less than 10 per cent of abusive detections. At last, the algorithm can detect 88 per cent of erroneous soundings for the Lennermor echo-sounder, despite its characteristics.
For all these reasons, the algorithm has been recently incorporated into the treatment software developed by the SHOM, in differed time.

Biography

Nathalie Debese, engineer in mathematics, has graduated from UTC (University of Compiègne) in 1989. Her research work, undergone at *Ifremer*, during her phD thesis, defended in December 1992, is focused on the navigation readjustment by knowledge of the bathymetric data. From March 1995 to September 2001, she has been involved into the "Multi-beam echosounder" project undergone within the SHOM. She is conducting studies successively on the errors typology of a multi-beam sounder, on the automatic detection of the punctual errors, on the quality control of the arithmetic data, on condensing of the bathymetric data, and on the display and extraction of the characteristics of the marine landforms. She received in May 2000 the medal Commodore Cooper of the IHO for her work on the detection of punctual errors, and became in October 2001 Study and Manufacture Engineer of the SHOM, assistant to the Chief of multi-beam sounder.