

A SYSTEM FOR CLEANING HIGH VOLUME BATHYMETRY

by Colin WARE¹, Leonard SLIPP¹, K. Wing WONG², Brad NICKERSON¹, David WELLS³,
Y.C. LEE³, Dave DODD³ and Gerard COSTELLO⁴

Abstract

Multibeam bathymetry systems produce data volumes which have swamped the highly labour intensive cleaning procedures often used in hydrography. In this paper we describe a Hydrographic Data Cleaning System (HDCS) designed to speed up the data cleaning process. The principle of HDCS is to allow full user control and access to all data parameters, while also implementing algorithms that automatically highlight problem areas and display them graphically on screen. A highly interactive graphical user interface allows the rapid editing of data in those problem areas.

INTRODUCTION

High volume data acquisition techniques for mapping the seabed have recently become available and adopted for use. Techniques used in Canada include airborne laser bathymetry systems, such as the Canadian-developed LARSEN 500 system developed by Optech Systems, and its successor, the SHOALS system (depth capability to 30m); sweep systems, such as the Danish-developed Navitronics system installed on several vessels operated by the Canadian Hydrographic Service, Public Works Canada and the Canadian Coast Guard (depth capability to 100m); and swath mapping systems, such as the Norwegian-developed Simrad EM100 multibeam sounder (depth capability to 300m), in use on the CSS MATTHEW and the CSS CREED, and also mounted in the hull of a remotely controlled submarine platform called a Dolphin by Geo Resources Inc, of Newfoundland. These systems

¹ Faculty of Computer Science, University of New Brunswick, P.O. Box 4400, Fredericton, New Brunswick, Canada E3B 5A3.

² Universal Systems Ltd, Fredericton, New Brunswick, Canada.

³ Department of Surveying Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada.

⁴ Canadian Hydrographic Service, Department of Fisheries and Oceans, 615 Booth St., Ottawa, Ontario K1A 0E6, Canada.

have a number of features in common: A high data rate is one of them - typically 50,000 - 100,000 depth measurements per hour. More importantly they have considerable similarities in the structure of the data acquired.

Multibeam bathymetric surveys are typically done by specially equipped vessels moving over designated areas in parallel line patterns, taking depth measurements along the way (see Fig. 1). For purposes of identification, the ship's "track" (the vessel path over the entire survey) is divided into segments called "lines". As the ship moves along a line an array of transducers periodically generates sonar impulses towards the sea floor, and an array of depth values transverse to the ship's track is computed from the reflected signals. We call this array a "profile" and depending on the system this may contain from 12 to 132 individual depth measurements or "soundings". This structure of lines, profiles and soundings is common to all of the data acquisition systems, even though the methods for determining the depths may be quite diverse - and acoustic signals are not necessarily involved (lidar, for example, uses a laser pulse). It is this common structure which makes our general approach to the problem of data cleaning possible.

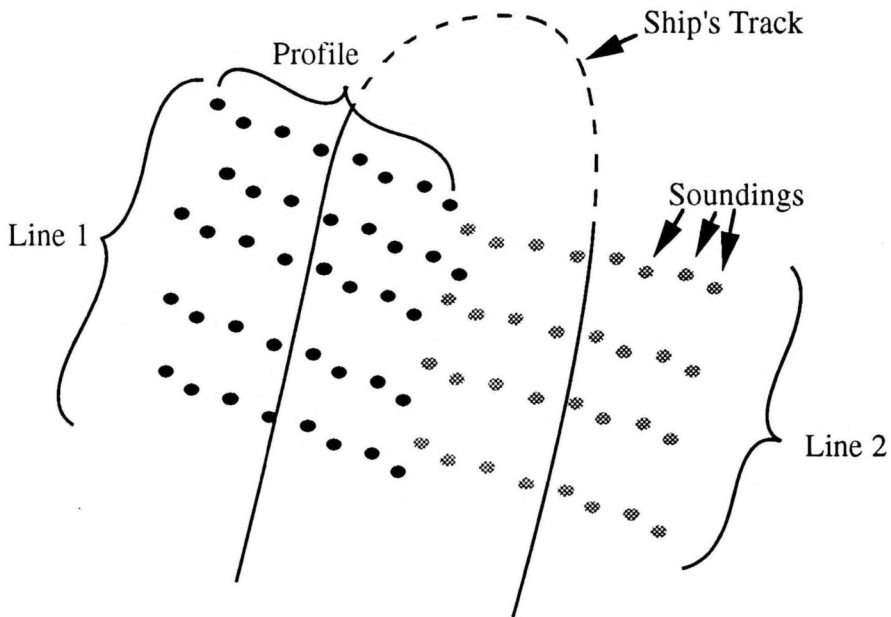


FIG. 1.- The anatomy of a multibeam survey. The ship's track is divided into survey lines, each of which contains multiple transverse arrays of soundings.

The system we describe in this paper evolved over a three year development cycle starting at the beginning of 1989. Our first steps were to establish a preliminary data format for storing the data and to work on visualization tools to allow us to interpret samples of data from the different systems. However, it became apparent that these tools could be integrated with some on the statistical analysis of surface data (WARE, KNIGHT and WELLS, 1991) to create a data cleaning system for the removal of depth blunders (WARE, FELLOWS, and WELLS, 1990). The statistical work resulted in an immediate transfer of visualization tools into industry by Universal

Systems Ltd. of Canada for Krupp Atlas Elektronik of Germany and Holming of Finland. The present Hydrographic Data Cleaning System (HDSC) is a complete reworking and integration of ideas developed in a number of studies on editing and visualization tools, filtering techniques and advanced data structures (WELLS et al, 1990). It supports rapid data access using an integrated object oriented design with a consistent graphical user interface.

DESIGN PRINCIPLES

Over the course of the system evolution described above a small set of ideas became fundamental principles on which the design is based. For clarity we set out these ideas below in point form.

1. Preservation of data

We believe that it is important not to decimate the data in the process of cleaning them. Instead, it has been our policy to flag bad data and preserve their values. This is important because different applications have different requirements, some of which may be impossible to anticipate. For navigation charts shoal biasing is a standard practice, whereas for the computation of dredge volumes all the soundings may be factored into a running average. The noise in one application may be the signal in another.

2. General purpose file formats

A crucial element in our effort to make the data cleaning system as general purpose as possible has been the development of a general purpose file format designed to accommodate data from a wide variety of multibeam systems. This is only possible because of the basic similarities in high level structure of the data as illustrated in Figure 1. Once data from a system has been translated into the standard file format it can be processed by the system. Thus most of the differences between data acquisition systems can be accommodated at a very early stage by the software which translates data into the standard file format.

3. Algorithmic blunder detection

We have designed algorithms to detect navigation anomalies and depth anomalies. The purpose of this detection is primarily to alert the hydrographer to areas where the data has problems. In some cases the hydrographer will accept the criteria employed by the algorithm and cause the data to be rejected; in other cases the hydrographer may wish to reject a portion of the anomalies detected by the algorithms.

A guiding principle of our design of algorithms for automatic cleaning is their simplicity. It should be easy to understand the basis of statistical decision

making. Simplicity is not to be derided, it is important that as many people as possible understand how decisions are made concerning the data.

4. Power tools approach

Full automation of the data cleaning process requires (a) that a decision process acceptable to users be implemented, and (b) users become confident that these automated decisions reliably meet their criteria. How a hydrographer decides that a particular area contains errors due to fish or kelp or equipment malfunction is poorly understood and algorithms to automate this are beyond our current modelling capability; too many factors are involved.

Thus the hydrographer will be an important component of the data cleaning process at least for the short and medium term. Accordingly, the approach we have adopted is the implementation of productivity tools which allow the users to rapidly identify problem areas (highlighted through the preprocessing step) and deal with the problems rapidly and with as much information as possible. The hydrographer does what people do best: sophisticated pattern recognition and decision making. The computer does what it does best: careful bookkeeping and number crunching to isolate problem areas, rapidly implement decisions, and present the information in an accessible, easy to understand fashion.

5. Graphical User Interface

The best way to comprehend the large volume of data generated by multibeam system is through graphical representations. We find that providing highly interactive graphical editing tools provides a natural and efficient interface to the data. We use the Motif Toolkit with the X11 window standard in creating a consistent and uniform user interface to all of the different cleaning and visualization operations.

6. Data structures designed to support interactivity and data processing

A highly interactive graphical approach requires that the system should respond to common inputs (such as data flagging) in a fraction of a second, and other operations should be as fast as possible; the penalty is a severe loss in user productivity. Given millions of data points it is necessary to design data structures to support rapid data selection operations.

THE HIGH LEVEL DESIGN

The typical data flow in processing consists of the following four steps:

- Step 1. The raw data are reformatted into the HDCS file formats.
- Step 2. Navigation is checked and bad positions are flagged as necessary.

- Step 3. Soundings are cleaned and flagged in an algorithm-assisted editing process.
- Step 4. Cleaned data are moved out of HDCS for database storage, chart making, or further analysis.

From the user's perspective, the interface to the data is through an interactive windowed environment with pull down and pop up menus and, wherever possible, direct mouse manipulation of graphical data objects and viewing parameters. Thus, for example, a line of soundings can be selected by point and click mouse interactions; the zoom windows can be resized by dragging the corners.

The basic screen layout is as illustrated in Figure 2 (colour plate). The three windows labelled 0, 1 and 2 represent views of the data at three different scales. Thus it is possible to show an entire survey region of, say, 5 km² in window 0, a 500m x 500m subset in window 1 and a 50m x 50m subset of window 1 in window 2. In this way a highly magnified small region of data can be viewed without losing the context in which that data exists. The scaling ratios between the three windows are under the user's control and can be changed either by direct manipulation of the yellow field-of-view indicators (in windows 1 and 2) or by means of dialogue boxes. Updates made to one window are automatically reflected in the other two.

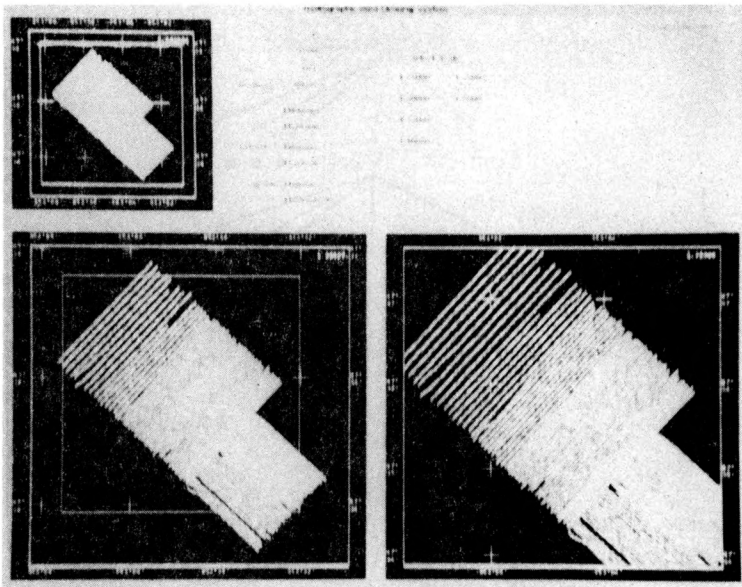


FIG. 2.- Basic Screen layout. The basic screen layout consists of four windows, which we number 0, 1, 2 and 3 counterclockwise from the upper left. Windows 1, 2 and 3 are for graphical data display. Window 3 contains textual information or dialogue boxes depending on the task at hand.

Other aspects of the user interface will be described in the context of the various data cleaning operations described here.

FILE FORMATS

This section outlines the file structures and data formats we have developed to support data cleaning tools for large multibeam bathymetric data sets. The basic principle followed here is to save all the data necessary to make proper data cleaning decisions. Normally, this means saving all of the observed data from a variety of sensors such as the positioning system, gyro, vessel roll, pitch and heave, tide gauges, and salinity/temperature/depth profiles along with the multibeam data. The advantage of our standardized format is that once data from a particular system is converted the cleaning tools can be applied without modification. System differences are, as far as possible, encapsulated in the code which transforms the data into the standard format.

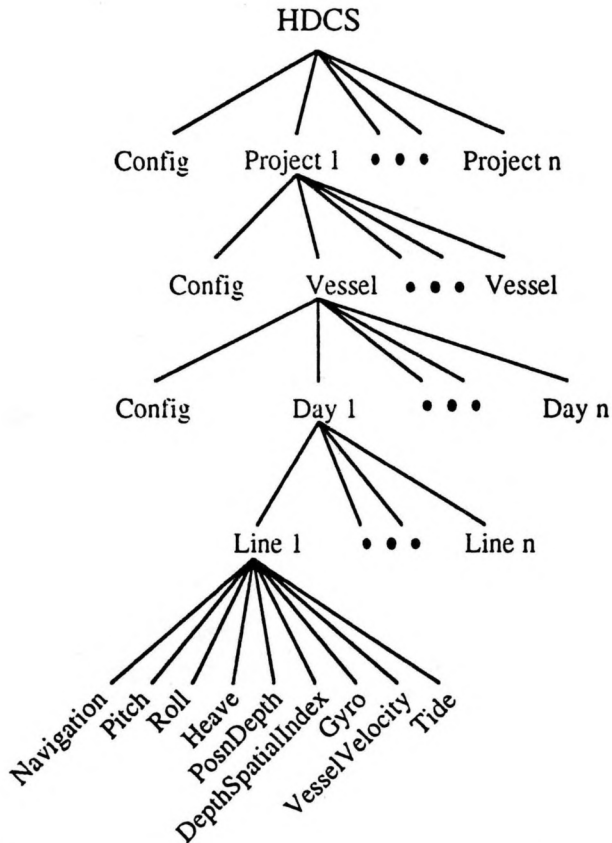


FIG. 3.- The HDCS file structure. This structure is based on the UNIX file system. Each entity is either a file or a directory. Projects, Vessels, Days and Lines are all directories. The bulk of the data is stored in the set of files associated with each line.

The HDCS standard file structure is illustrated in Figure 3. Each object in this figure represents either a file or a directory. The highest level is the Project, which can consist of the data from a number of vessels. The data from a particular

vessel is divided up into days and days are divided into lines. Each line generates 9 components which separately store the Navigation, Pitch, Roll, Heave, Depth, Gyro, Vessel Velocity, and Tide. Standardized formats have been defined for each of these files. The PosnDepthIndex file is a special file designed to allow fast access to the data relating to a particular sounding when that sounding (or set of soundings) has been queried through the graphical user interface.

NAVIGATION FILTERING

For the detection of bad navigation data the HDCS employs a form of Kalman filtering. The purpose of the filter is to detect and flag bad observations and also to interpolate in time, to synchronize data from various sensors (navigation, gyro, depth). It is not intended to adjust the position in any way. If a position is determined to be bad then it is flagged for rejection and its effects on the filter are removed.

The observations in this case are northing, easting and gyro readings. In the filtering process the state of the vessel is represented by a state vector containing: northing (N), easting (E), drift angle (β), velocity north (δN), velocity east (δE), velocity of the drift angle ($\delta\beta$), acceleration north ($\delta^2 N$), acceleration east ($\delta^2 E$), and the acceleration of the drift angle ($\delta^2\beta$). The drift angle is the difference between the ship's heading and the course made good (ship's track). The observation models (y) are:

$$\begin{array}{lll} \text{North} & \text{East} & \text{Azimuth} \\ y_1 = N + \epsilon & y_2 = E + \epsilon & y_3 = \tan^{-1}(\delta E / \delta N) - \beta + \epsilon \end{array} \quad (1)$$

There are two basic stages to this navigation filter. The first stage is the prediction process, and the second stage is the filter process. The prediction process is the use of past observations to predict the state at a given time. The filtering process is the use of the predicted state and the current observation to estimate a state, at the time of the observation (HOUTENBOS, 1982).

For the prediction process the HDCS uses a constant acceleration state space model (SCHWARZ, 1989). The prediction equations are:

$$\begin{array}{l} X_{k/k-1} = F_k(X_{k-1/k-1}), \\ C_{k/k-1} = F_k C_{k-1/k-1} (F_k)^t + Q_k, \end{array} \quad (2)$$

where:

$X_{k/k-1}$	predicted state at time k,
$X_{k-1/k-1}$	estimated state at time k-1,
$C_{k/k-1}$	predicted state covariance at k,
$C_{k-1/k-1}$	estimated covariance at time k-1,
F_k	transition matrix (k-1 to k),
Q_k	process noise covariance matrix.

The transition matrix and the process noise covariance matrix are derived using the methods described in SCHWARZ (1989).

To estimate the state vector t HDCS uses a least squares adjustment in the form of Kalman Filtering. The estimation equations are :

$$\begin{aligned} X_{k/k} &= X_{k/k-1} + G_k[Z_k - H_k(X_{k/k-1})], \\ C_{k/k} &= (I - G_k H_k)C_{k/k-1}, \\ G_k &= C_{k/k-1}(H_k)^T[R_k + H_k C_{k/k-1}(H_k)^T]^{-1}, \end{aligned} \quad (3)$$

where:

$X_{k/k}$	estimated state at time k ,
$C_{k/k}$	estimated state covariance at time k ,
Z_k	observation vector for time k ,
G_k	Kalman gain matrix,
R_k	observation covariance matrix,
H_k	design matrix defining observation and state space relationship. Made up of the first derivatives of the observation models with respect to each of the state vector elements.

One of the advantages of using Kalman filtering techniques is that it is possible to process one observation at a time (BIERMAN, 1977). As a result $[R_k + H_k C_{k/k-1}(H_k)^T]^{-1}$, R_k , and Z_k are scalar quantities, and H_k and G_k are vectors. This eliminates the need for matrix inversion and reduces the computational strain on computing resources. It also enables the observations to be recorded at different times and from different sources, as is the case with many bathymetry systems.

The positions and gyro readings are recorded at different times and with different time intervals. Because the gyro observations are recorded with the depth record the filter estimates a state for every depth record, as well as for every position. As a result this filter may be used to produce estimated positions for all depth records if desired.

When an observation is passed through the filter the time of that observation is used to determine elapsed time from the last update. This time difference, and the previous state, are used to predict the state of the vessel at the time of the observation via the transition matrix and the system noise covariance matrix. If the observations are recorded at the same time and have the same time interval then the transition matrix and process noise covariance matrix would be determined once and would remain the same throughout the process.

The predicted state is now used to determine the final estimates. The observations are represented in the form of misclosures in the observation vector. The Kalman gain is used to proportion the influences of the predicted state and the observation. The lower the gain the less emphasis is put on the observation, and greater emphasis is put on the prediction. The final estimates are used in the prediction process of the next observation.

The rejection criteria are set by the operator in the form of thresholds. The thresholds are the maximum acceptable along track, across track, and gyro misclosures. The observation misclosures are defined as the differences between the observed values and the predicted values. The position misclosures (northing and easting) are converted to along track and across track misclosures for comparison to

the threshold values. Any observation that exceeds this threshold is considered bad and is flagged for rejection.

Interactive Navigation Cleaning

In navigation cleaning the operator selects a line to be cleaned, by clicking on it in window 2 with the mouse (Figure 2). The Kalman filter can then be applied (via a menu selection) to the selected line and the rejected positions, are highlighted on the display. An additional set of overlaying windows appear to allow interaction and editing of positions in the selected line as shown in Figure 4. The user can use the mouse to sweep out a rectangle to select one or more positions, and specify deletion with or without interpolation. If interpolation is specified, the depth values are given positions based on linear interpolation between the predecessor and successor "good" positions. For additional information on a selected line, the operator can choose to display a set of graphs which show plots of such attributes as Heading, Pitch, Roll, Heave, and Tide against time.

DETECTION AND CLEANING OF DEPTH BLUNDERS

Our strategy has been to develop algorithms capable of dividing the set of soundings into statistically defined classes according to how far they deviate from the estimated true surface. These classes can then be used either to automatically clean the data (the program automatically rejects soundings beyond a certain class level) or they can be used as part of an interactive cleaning process to enhance productivity. In the latter case the operator only needs to inspect those areas where the flagged soundings are highlighted graphically, and if the operator agrees with the set of algorithmically rejected soundings in a particular region, then all of the soundings can be rejected in a single operation. Our data cleaning algorithms rely on the generation of regularly gridded digital terrain maps and these can only be efficiently implemented by creating spatial subsets of the survey data.

Spatial Subset Creation

For computational reasons, as well as for ease of use, a survey region is interactively divided into a number of spatial subsets as illustrated in Figure 5. The operator can select one of a set of preselected sizes from a menu (e.g. 500 metres, 1km, 2km) or can set an arbitrary size. The selected subsets are displayed as shown in Figure 5.

Criteria for Outlier Detection

To determine whether or not a given sounding is anomalous in depth we use a method which is a generalization of the bin statistics described in VARMA et al (1989). In essence this method computes both a running weighted average μ and a running weighted standard deviation σ over the region of the survey (WARE,

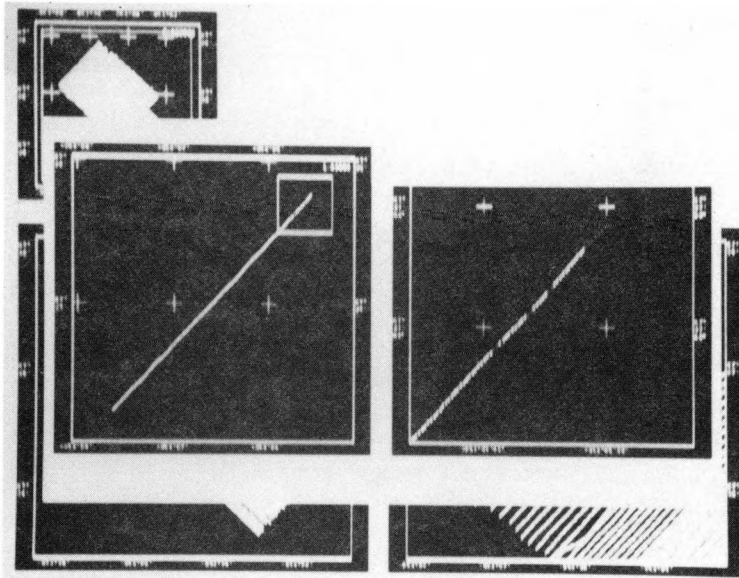


FIG. 4.- The set of windows which appear when the user wishes to interact with a line for the purpose of cleaning the navigation.

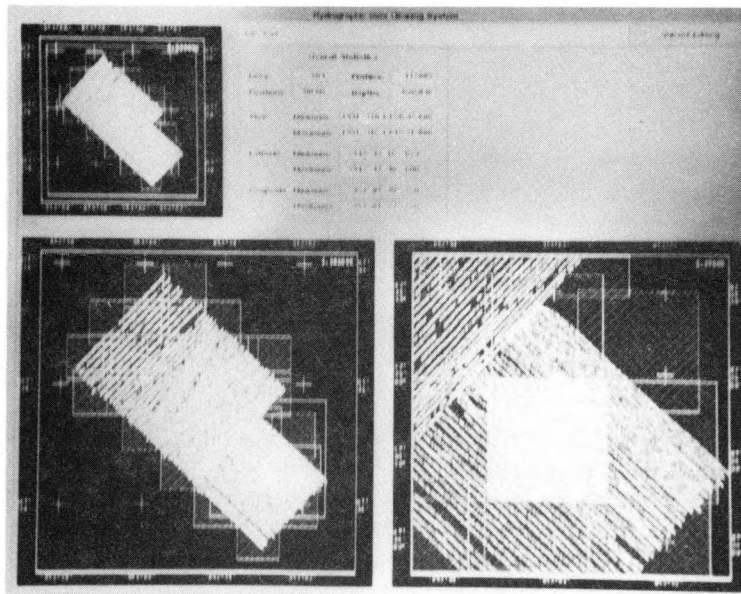


FIG. 5.- Spatial subset creation. The user can interactively tile the survey area which square subsets. The scale can either be present or defined arbitrarily by the operator.

KNIGHT and WELLS, 1991). The statistical classification of soundings is subsequently based on a linear combination of μ and σ .

Our method involves the generation of a circular weight field w_i around each sounding d_i (see Fig. 6). At the center of the field the point is given of a weight of 1.0 and this weight declines linearly to zero at the circumference, which is radius r away from d_i . For the Simrad EM100 system, the weight field diameter varies as a function of depth, reflecting the fan out of the beams. For the Navitronics system the beams are vertical and a constant diameter weight field is used.

The mean depth μ_p at position p on the horizontal plane given a set of n soundings is computed using

$$\mu_p = \frac{\sum_{i=1}^n d_i w_i}{\sum_{i=1}^n w_i} \quad (4)$$

where

$$w_i = \begin{cases} 1 - \text{distance}_i/r \\ 0 \text{ otherwise} \end{cases}$$

and $\text{distance}_i < r$ is the horizontal distance between d_i and p .

The vertical standard deviation σ_p for point p on the plane is given by

$$\sigma_p = \sqrt{\frac{\sum_{i=1}^n w_i (d_i - \mu_p)^2}{\sum_{i=1}^n w_i}} \quad (5)$$

The above formulae provide a way of computing a continuous estimate of the mean and standard deviation of a surface derived from randomly scattered discrete data points. Each sounding has an influence only within a fixed radius r and the influence of the sounding decreases linearly to r . Like binning this technique relies on a reasonably constant density of soundings. Fortunately, a constant data density is also a goal of the data acquisition procedures.

In order to use μ_p and σ_p in data cleaning we first compute their values on a regular (500x500) grid over the selected region of the survey. Computed in this way μ is a digital terrain model, while σ is a digital discrete sampling of the standard deviation surface. The method for this computation involves accumulating sums of weighted depths and weighted sums of squared depths. These are computed for each cell in a large two dimensional arrays which correspond to the area of interest. An estimate of the surface can subsequently be derived in a

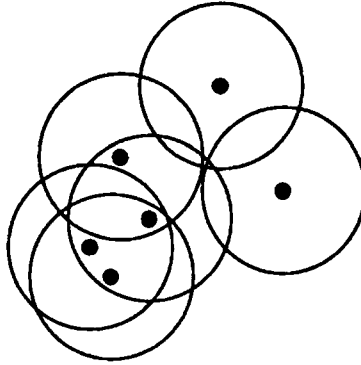


FIG. 6.- A plan view of a set of soundings each with a circular region of influence.

straightforward way as can the standard deviation of each point on the surface (see WARE, KNIGHT and WELLS (1991) for details).

Statistical Classification of Depth Anomalies

We pre-flag the data with eight levels corresponding to differences from the estimated surface. We have found that different linear combinations of μ and σ are appropriate to different data acquisition systems. For the Navitronics system the levels are defined by

$$\text{level}_{jp} = \mu_p + C_3\sigma_p - j(C_1\sigma_p + C_2) \quad j = 0..7 \quad (6)$$

where μ_p and σ_p denote the estimated means and standard deviations at point p in the survey. The reason that a positive constant C_3 is useful is illustrated in Figure 7. This shows some hypothetical data containing two outliers. The outliers have pulled up the estimated mean and they have locally increased the standard deviation. By adding $C_3\sigma$ we get an estimated surface which is closer to the "true" surface and this becomes a better baseline for detecting outliers.

*Preflagging is done for all soundings d_i
if $(d_i < \text{level}_{ji})$ then flag at level j*

The purpose of constants C_1 , C_2 and C_3 is used to establish a baseline approximation to the ocean floor given by $\mu - C_3\sigma$. C_1 and C_2 provide the threshold steps above the baseline. C_2 is a constant in meters, whereas C_1 is a proportion of the standard deviation. C_1 allows for larger steps to exist where the data is noisy (or the seabed is rough) than where it is clean (or smooth). We have established that $C_3 = 1.0$, $C_2 = 0.05$ and $C_1 = 0.1$ provides for an 8 step sequence which covers the range of conditions we have found in a data sample from a 1989 survey at Louisbourg, Cape Breton. We do not as yet know the general validity of these parameters.

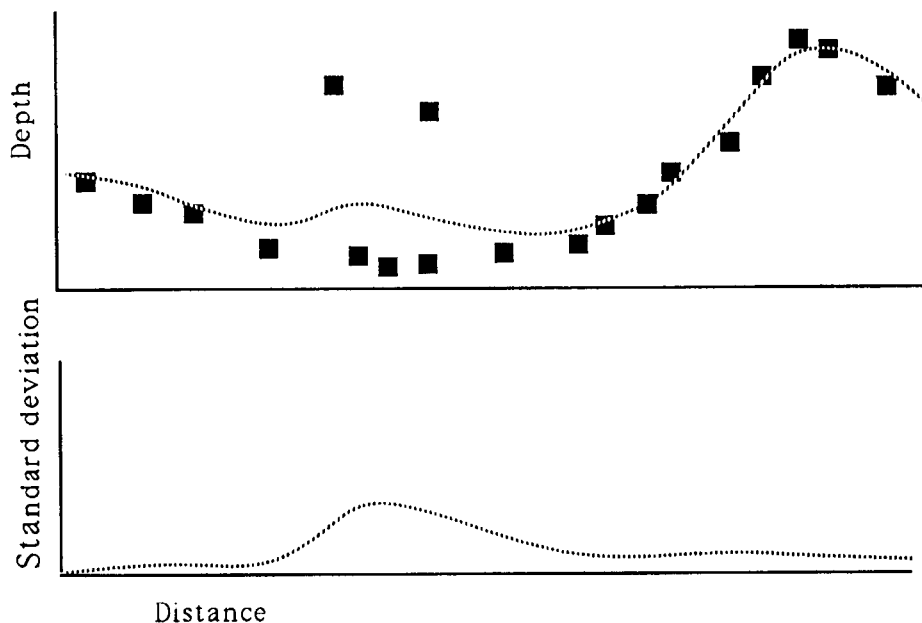


FIG. 7.- The dotted line in the upper graph represents μ while the dotted line in the lower graph represents σ . Note that the depth axis is inverted. That is, higher values (greater depths) appear towards the bottom of the graph.

Evaluation of Algorithms

In order to obtain a meaningful evaluation of the algorithm approach we implemented the data thinning and shoal biasing method as described in our introduction. Once this was implemented we could make meaningful comparisons between algorithm flagged data and data flagged by a hydrographer in order to determine the number and sizes of the discrepancies.

Our algorithm assessment method was as follows:

Step 1. Apply shoal biased thinning (ORAAS, 1975) to the data ignoring all soundings with the navigation error bit set (bit 3) and the bad sounding bit (bit 4, set by the hydrographer). This file contains approximately 10% of the original soundings and we call it FILE A.

Step 2. Apply shoal biased thinning to the data ignoring all soundings with the navigation error bit set (bit 3 set by the hydrographer) and the algorithm flagged bit set (bit 1 set by the algorithm). This file also contains approximately 10% of the original soundings and we call it FILE B.

Step 3. For each sounding in FILE A determine the nearest neighbor in FILE B and determine the depth between them. Note that there is an asymmetry here in that FILE A is given a primary status as a reference. However, our results suggest that it makes little difference to reverse the roles of A and B.

Step 4. Compute a histogram of the depth differences between FILE A and FILE B.

What we are doing here is essentially comparing the field sheet resulting from one processing method with the field sheet resulting from a different processing method. This method can be used not only to evaluate an algorithm against a human operator, it also can be used to tune the algorithm to match the performance of a particular operator, and to compare one operator against another to see if consistent procedures are being followed. It can also be used for training purposes.

Results from Algorithm Evaluation

We have based our analysis on a comparison between operator cleaned data and algorithm cleaned data from a portion of a 1988 CHS Caraquet survey performed using a Navitronics system. The section we examined contained approximately 200,000 soundings. Using this data we tuned parameters C_1 and C_2 of equation 1 so that the algorithm gave results which were as close as possible to those of the operator. The results we show were computed with the parameters. We have established that and $C_1 = 0.1$, $C_2 = 0.05$, $C_3 = 1.0$ and $j = 2$ for equation 3. This simplifies to

$$\text{if } d_i < \mu_p + 0.8\sigma_p - 0.1 \text{ flag sounding as bad}$$

Depth differences derived using the assessment method described above are shown in Figure 8 in the form of a histogram. It plots the results of the combined data for two 250 m² samples of the Caraquet survey, each of which contains approximately 18,000 soundings in the original data and approximately 1,400 soundings when the shoal biasing and thinning algorithm was applied.

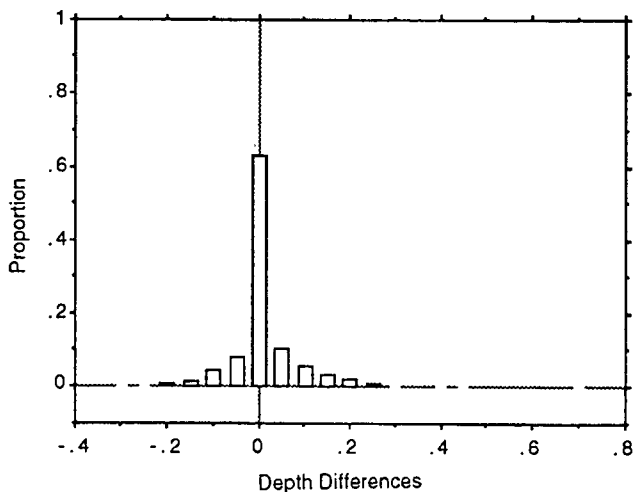


FIG. 8.- The distribution of depth differences (in metres) between nearest neighbours in the algorithm flagged data set and the operator flagged data set.

Some Summary Statistics

Out of a total of 2,836 soundings in the operator flagged *thinned* sample (two field sheets) 31 algorithm flagged soundings differ by more than 0.2 metres from the operator flagged sample. 204 algorithm flagged soundings differ by more than 0.1 metre. To put it another way, 93% of soundings are within 0.1 metre and 99% of the soundings are within 0.2 metre of the nearest neighbour in the field sheet based on operator flagging.

The above differences can be either negative (algorithm is conservative) or positive (algorithm is less conservative). The values of the latter types are of special concern because of the potential risks to navigation. Looking at these separately, in the same sample regions, the very worst discrepancies are 0.7 meter (1 sounding), 0.45 metre (1 sounding), 0.40 metre (2 soundings) and 0.3 metre (6 soundings).

The horizontal distances between nearest neighbours are plotted as a histogram in Figure 9.

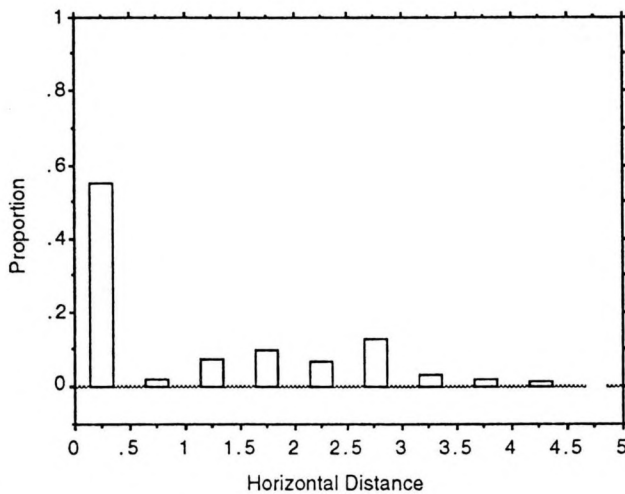


FIG. 9.- The distribution of horizontal distances (in metres) between nearest neighbours in the algorithm flagged data set and the operator flagged data set.

This plot shows that about 55% of the horizontal distances were less than 0.5 metre. And the remainder were distributed in a region with a radius of less than 5 metres.

What these results tell us is that our present algorithms can come close to simulating the performance of an operator, but that there are significant and possible hazardous discrepancies between the results obtained by the algorithm and those obtained by the operator.

Interactive Algorithm-Assisted Editing

The basic screen layout for interactive screening is the same as that for navigation cleaning. The same three levels of zoom are used as can be seen in Figure 10. However, in this case, windows 2 and 3 show individual soundings rather than the positions of the ship's track. Also, when in editing mode the user can create an elongated selection box in window 2 (as illustrated), and the soundings in that window are shown in window 3 viewed in elevation, that is, the y axis shows depth).

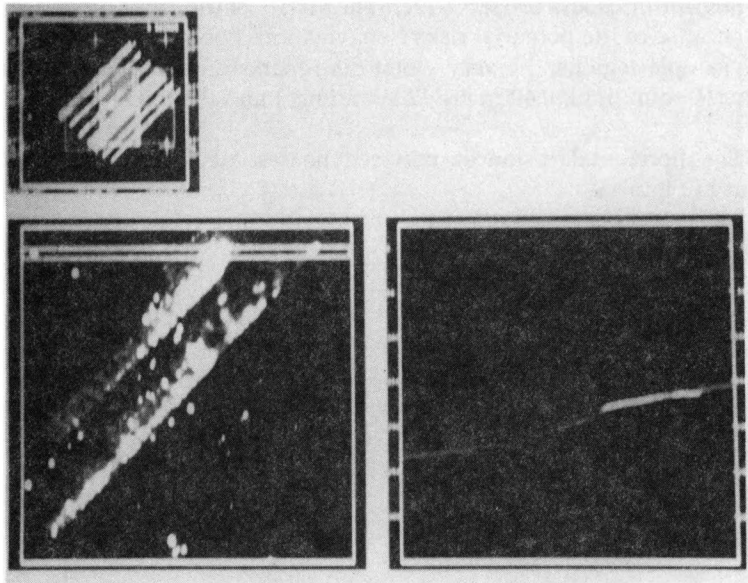


FIG. 10.- Screen for interactive depth editing. The soundings in the elongated box shown in window 1 are displayed in window 2. The background colour for window 1 denotes the standard deviation. Regions tending towards white have a locally high standard deviation.

The overall process of data cleaning can be described by the following three steps.

Step 1. Use the Spatial Subset Creation facility to partition the area of the survey into square "field sheets" or boxes. These have an arbitrary orientation with respect to the initial survey. Call these boxes $b_1, b_2, b_3, \dots, b_n$.

For each box b_i the hydrographer does the following:

Step 2 Compute μ and σ and Preflag outliers

Step 3 Systematically scan the data to either confirm the points set by the algorithm or not according to judgement and expertise.

Additional visualization options include the ability to make a spatial coverage plot which shows the coverage of an area by the multibeam acoustics (see WARE et al, 1991). Coverage is colour coded depending on whether a region was sampled by zero, one, two or more lines. This allows operators to visually determine the degree of overlap between survey swaths, and therefore whether survey specifications, of either 100% coverage, 200% coverage, or more, have been met.

Depth standard deviation (according to the algorithm in WARE, KNIGHT and WELLS, 1991) can also be displayed as a colour coded map. In this case the operator can see areas where the depth standard deviation values are high, and the operator will typically only wish to do a detailed examination of those regions.

The results of the algorithm-based outlier detection are made visually available by colour coding outliers. The operator is then given the choice of confirming the algorithm selections over a selected area, or making a decision on his or her own judgement.

Data Structures to Assist Spatial Searches

The purpose of spatial indexing in this system is to allow efficient access to profiles falling within or overlapping with a query rectangle drawn out by the user. In order to facilitate this search each profile is associated with a bounding rectangle which is defined by the minimum and maximum coordinates of soundings in the profile. Each daily file is also associated with a bounding rectangle. The spatial indexing makes use of a Morton code also known within the Canadian Hydrographic Service as an HHCode (VARMA et al, 1990). SAMET (1990) discusses the general principles of spatial searching using Morton codes. The spatial index itself is stored as a special file with the set of files containing line information (Figure 3).

Data within the various line files shown in Figure 3 are already sorted by time due to the sequential mode of data collection and time stamps are stored within time index files which are associated with each of the different file types (these index files are not shown in Figure 3). The time index files allow that once a sounding (or set of soundings) has been identified by a query, all of the other attributes associated with that sounding can be accessed and displayed.

CONCLUSION

We feel that many of the strengths of the HDCS system derive from the continual advice which we have received from CHS hydrographers. This interaction has come from the regular visits of CHS hydrographers to the University of New Brunswick, and during the many field trips we have made to observe current processing systems and practices.

However, it is worth noting that hydrographers very often do not agree with one another and that their different practices are often equally valid. In these cases we have somewhat arbitrarily adopted a particular approach. Because a complex system such as HDCS necessarily enforces certain modes of operation on its users, we feel that its adoption should eventually lead to more uniform practice among hydrographers.

The HDCS has broken a data processing bottleneck which rendered several high volume bathymetry systems either inefficient or inoperative in Canada. HDCS together with existing data management and presentation routines has been

incorporated into a commercial product, the Hydrographic Information Processing System (HIPS) marketed by Universal System Ltd. of Fredericton, Canada. HIPS has been chosen as the data processing component of the Canadian Ocean Mapping System (COMS). As a result of the COMS field trials in November 1991, the Canadian Hydrographic Service has decided to install HIPS on three vessels for full production use during the 1992 field season.

ACKNOWLEDGEMENTS

The support of a National Science and Engineering Council Strategic Grant has been crucial to this work. We are grateful to Ken DOUCET, for coding preliminary data formats and navigation filters; Mike MACDONALD, Kevin MARINELLI and Martin SCHULTZ for system support; the Canadian Hydrographic Service, in particular Bob BURKE and Paul BELLEMARE for strategic support throughout the project, and CHS hydrographers: Mike LAMPLUGH, André GODIN and Mike RUXTON in the many suggestions which have enabled the cleaning system to mature.

REFERENCES

- BEIRMAN, Gerald J. (1977). *Factorization Methods for Discrete Sequential Estimation*. Academic Press, New York.
- HOUTENBOS, Ir. A.P.E.M. (1982). "Prediction, filtering and smoothing of offshore navigation data." *Forty Years of Thought*, vol. 2. Collected papers for Professor Baarda, Delft.
- ORAAS, S.R; (1975) Automated sounding selection: *International Hydrographic Review*, vol. LII(2), p. 109-119.
- SAMET, H. (1990) *Applications of Spatial Data Structures*, Addison Wesley, Reading, MA.
- SCHWARZ, K.M., M.E. CANNON, R.V.C. WONG (1989). "A comparison of GPS kinematic models for the determination of position and velocity along a trajectory" *Manuscripta Geodetica*, 14(5), pp. 345-353.
- VARMA, H.P., BOUDREAU, M., MCCONNELL, M., O'BRIEN, M. and PICOTT, A. (1989). Probability of detecting errors in dense digital bathymetric data sets by using 3d graphics combined with statistical techniques. *Lighthouse* 40, pp. 31-36.
- VARMA, H.P., BOUDREAU, H. and PRIME, W. (1990) A data structure for spatio-temporal databases. *International Hydrographic Review*, Vol. LXVII(1), pp. 79-92.
- WARE, C., FELLOWS, D. and WELLS, D. (1990) Feasibility study on the use of algorithms for automatic error detection in data from the FCG Smith, Project Report for DSS Contract FP 962-9-0339.
- WARE, C., KNIGHT, W. and WELLS, D. (1991) Memory Intensive Statistical Algorithms for Multibeam Bathymetry Data. *Computers and Geosciences*, 17(7), pp. 985-993.
- WELLS, D.E., WARE, C., NICKERSON, B.G., LEE, Y.C. and SLIPP, L. (1990) Requirements analysis and conceptual design of data cleaning tools for large bathymetric data sets, Contract report for Canadian Hydrographic Services, DSS Contract, FP 962-0-2040.
- WELLS, D.E., MAYER, L. and HUGHS-CLARKE, J. (1991) Ocean Mapping: from where to what? *CISM Journal*. AGSC. 45(4), 383-391.