International Hydrographic Review, Monaco, LXV(2), July 1988

PROCESSING 'LARGE' DATA SETS FROM 100% BOTTOM COVERAGE 'SHALLOW' WATER SWEEP SURVEYS

A NEW CHALLENGE FOR THE CANADIAN HYDROGRAPHIC SERVICE

by R. G. BURKE, S. FORBES and K. WHITE (*)

INTRODUCTION

The past two decades have ushered in many new and significant advances in survey technology. The work of our hydrographers can now be carried out more accurately and with greater efficiency due to improvements in positioning systems, computers and echo sounding systems. Many of these advantages have been offset by the steady and dramatic evolution that has taken place in the marine community over this period.

Basic economics dictate that ships be larger in order to make most commercial operations profitable. Operations, in turn, must be geared to maximize cargoes and minimize turn around. These same economic pressures also come to bear directly on shipowners and masters to compromise the traditional underkeel safety margins. As these safety margins shrink, the mariner is forced to rely more and more on the competence of hydrodynamic engineers in their predictions of hull behaviour under a wide variety of operating conditions and the competence of hydrographers and the accuracy of their equipment.

In order to cope effectively with these new demands, the Canadian Hydrographic Service (CHS) procured its first sweep system in 1982. The Navitronic SeaDig 201 sweep system was acquired to carry out detailed 100% bottom coverage surveys of critical areas such as harbour entrances, dock sites and dredged channels. Assembled in a transportable configuration, the system could be trucked or shipped to a remote location and be operationally deployed within a day [1]. In a demonstration of its capability, the system was deployed near Grise Fjord in the Canadian Arctic during 1984.

The success of this system provided a catalyst for the acquisition of a second and larger system. The second system, a specially designed 34.8 metre

(*) Department of Fisheries and Oceans, Canadian Hydrographic Service (Atlantic), Science Sector, Scotia-Fundy Region, Nova Scotia, Canada.

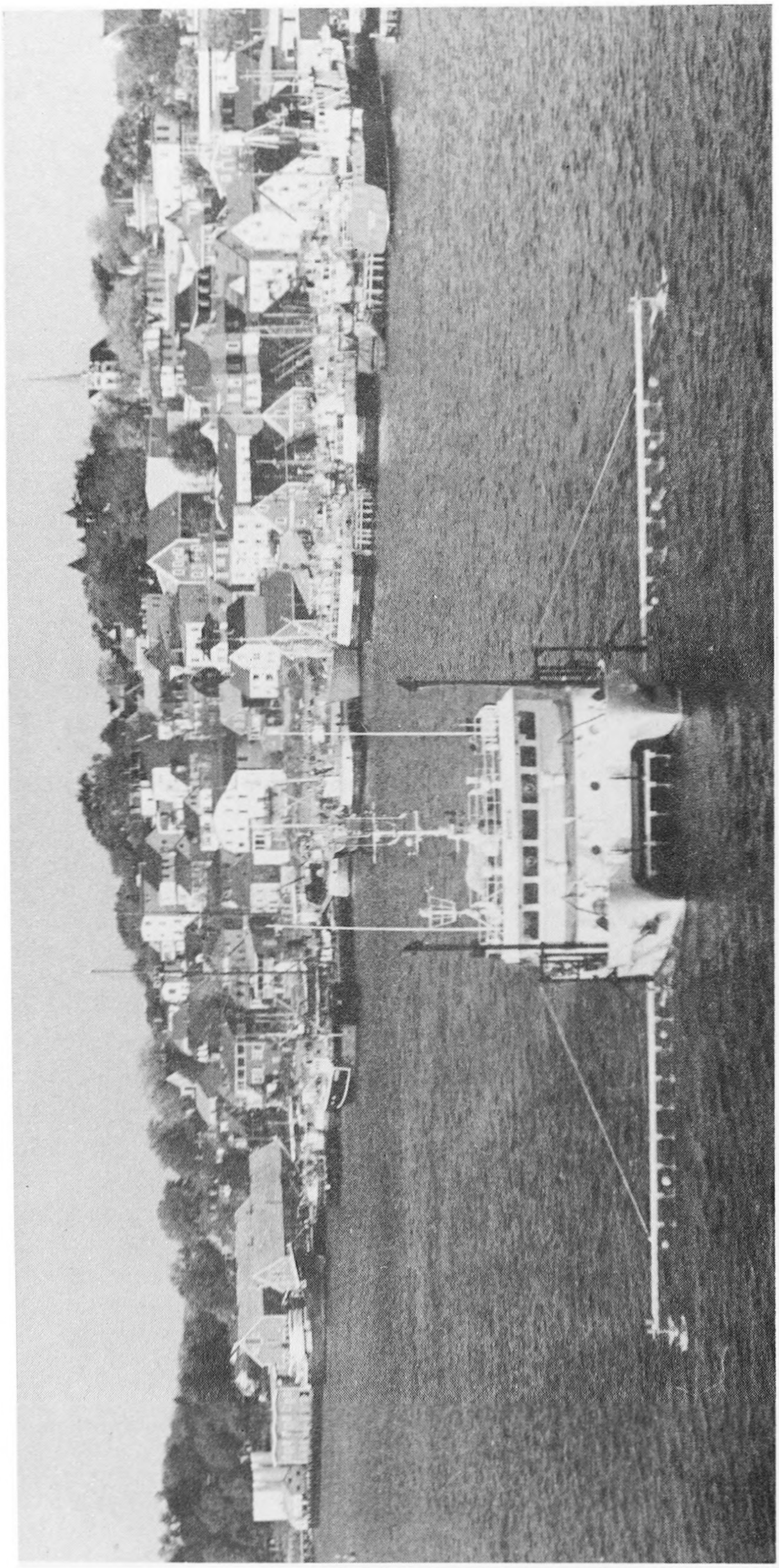
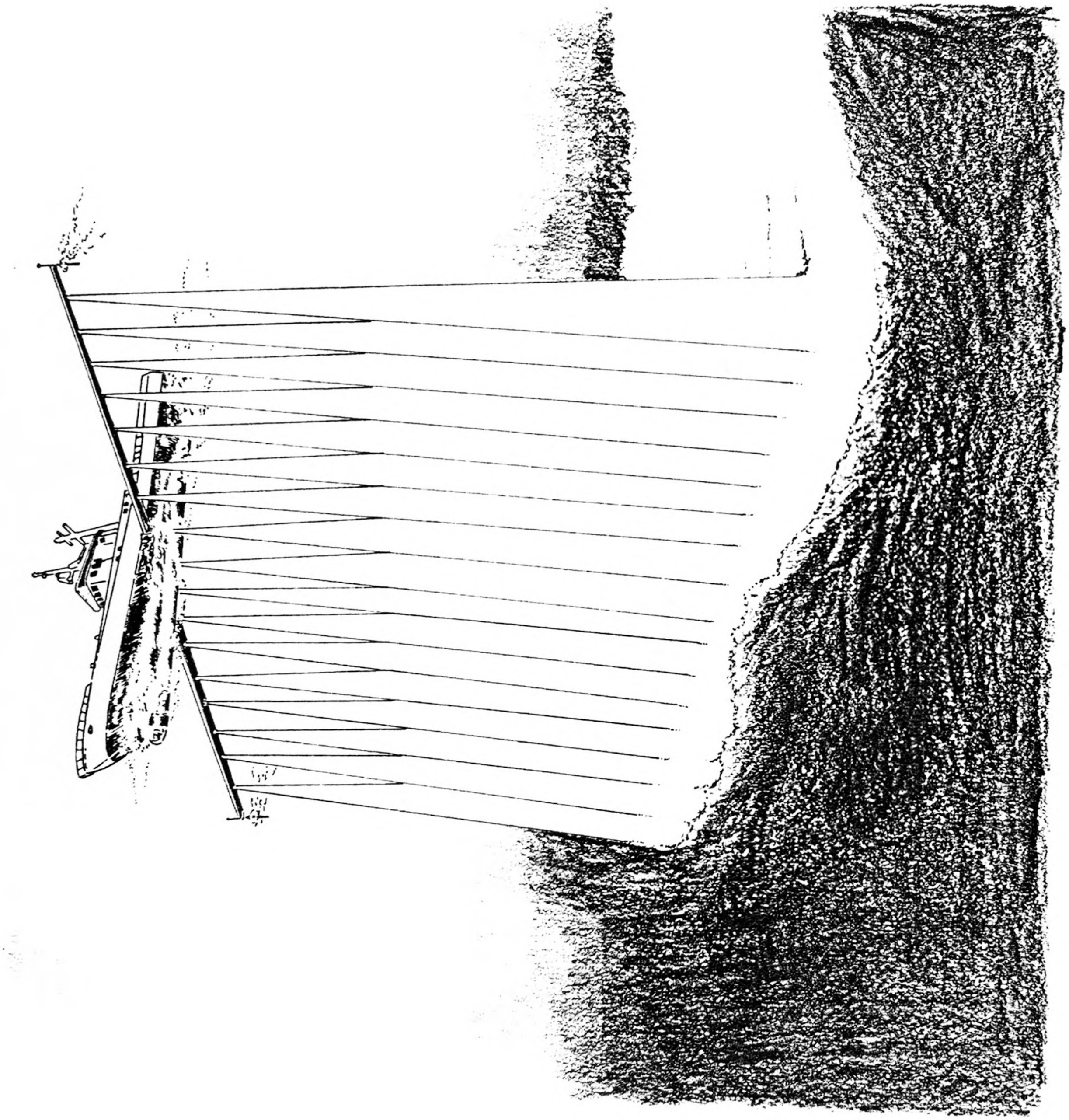


FIG. 1. — The *FCG Smith*.



catamaran, was built to deploy an array of thirty-three transducers and give a swath coverage of approximately 44 metres. The *FCG Smith* was commissioned and went into service during the 1986 survey season (Fig. 1). The vessel and its equipment have been previously described in the literature [2].

In addition to the increased survey capabilities that have resulted from the acquisition of these systems, a new challenge has come forth for field hydrographers and cartographers — data sets that are orders of magnitude larger than they have traditionally dealt with in the past! The *FCG Smith* with its thirty-three echo sounders operating at their maximum rate of 20 pings per second, is capable of making 2,376,000 depth measurements per hour. A depth pre-processor, which is part of the Navitronic Sweep System, selects only a small percentage of all depth measurements for logging and subsequent post-processing. Some 300,000 to 500,000 depth measurements are logged during a typical day. Obviously, with bathymetric data sets of this magnitude it is impossible for a human to analyse, edit, and display this information in a timely fashion without sophisticated software tools and a capable computing system. This paper describes the software that has been developed by the CHS to cope with the requirements and addresses some future considerations.

Data Acquisition

The data acquisition software was purchased as part of the two turn-key sweep packages supplied by Navitronic AS of Denmark. On the *FCG Smith* system, the data acquisition software is written in HP PASCAL and runs on the HP237 computer which controls the sweep system. The package provides the following:

1. A real time operational display of all critical survey parameters.
2. Fix and line following computations with a left/right steering display. In addition, the computer is interfaced directly to the autopilot and the vessel may be steered along a survey line by either the computer or a helmsman.
3. Logging of survey parameters including depths, positions, etc., to a Tandberg TDC 3000 tape recorder.
4. A user friendly interface to the sweep system.

All post-processing software currently in use has been developed by the CHS.

Data Processing — A Brief Background

When the first sweep system became operational during 1983, the Hydrographic Acquisition and Processing System Software (HAAPS), a package developed by the CHS for processing data from conventional surveys, was used [3]. It ran on an HP1000 computer system and was designed to handle data from semi-automated chart scalers or launch data loggers. Normally, along track profiles and 'quick look' plots were produced on site, utilizing software and hardware provided by Navitronic AS. Final and more rigorous processing was carried out on the HP1000 computer system after the party had returned from the field. The HP1000 computer system lacked the speed and disc space required to cope efficiently with the volume of data. In addition, the HAAPS data structure was never designed to accommodate the multi-transducer data set. Consequently, a number of compromises had to be made in order to process the data within a reasonable amount of time. First, when the data was transferred to the HP1000 computer and formatted into 'conventional' sounding files, only three of the eighteen depths for each data set were used — the shallowest on the port, the shallowest on the starboard and the deepest sounding. Format restrictions resulted in the loss of transducer identification, which in turn severely hampered data identification and verification.

An interactive editor that had been developed to work with the 'conventional' data could not be utilized with the sweep data sets due to format differences [4]. This made the source identification and checking of questionable soundings very awkward and time consuming. The production of shallow biased, overplot-removed, bathymetric plots proved to be another serious bottleneck. Disc space limited the data set to approximately 50,000 soundings. To sort and overplot this quantity of data for a plot at the scale normally used (1/2,000) required approximately 13 to 14 hours of CPU time. Typically, only 1000 to 2500 soundings could be plotted from the data set at the selected scale. Plotting required an additional hour or more.

If an error was discovered, the bad sounding would have to be located by placing the track plot over the bathymetric field sheet and examining along track profile plots and data listings to uniquely identify the data point in question. Once the offending sounding had been deleted, the whole overplot removal, plot, verification and edit cycle would repeat itself. While it was possible to process and arrive at a 'clean' data set using the HAAPS software, it was soon recognized that improved tools would be needed in the future. At this time, the CHS was in the process of phasing in new computers to replace the obsolete HP1000 systems. Consequently, no effort was made to enhance the HAAPS software package for processing sweep data. The selection of the Micro VAX II and the decision to acquire a second and larger sweep system occurred at approximately the same time.

Data Structures — An Important and Basic Design Consideration

During the initial evaluation of the new computer package for the *FCG Smith* sweep system, the following fundamental considerations were taken into account:

1. The system should be capable of processing all data on site to create a 'clean' data set suitable for incorporation into the construction of a nautical chart or a patch that could be issued via a Notice to Mariners.

2. The system should be able to cope with the volume of data gathered and have sufficient capability to keep abreast of the processing workload.

3. The system should allow the data processor to quickly and conveniently identify and trace data backward or forward through the processing system. It should also allow the processor to quickly compare the results from a number of survey passes over a critical or contentious point.

4. The system should be simple and easy to use.

Items 1 and 2 are largely dictated by the power of the computer system, whereas 3 and 4 tend to be more a function of the software tools that are developed. The selection of the Micro VAX II as the computer system resulted from a detailed evaluation of a number of systems by shipboard users of computers at the Bedford Institute of Oceanography. Figure 2 shows the current configuration of the system acquired for the *FCG Smith*.

During the early stages of the software design, it became abundantly clear that the data structure would be a critical element in the overall software package. If computer speed and storage were not a consideration, an ideal situation would be to establish a data structure that would retain all data on-line and in a coherent structure for both the 'raw' and processed data. While this concept was given serious consideration, it had to be rejected simply because of disc storage limitations.

In the end, a data structure that is referred to as a 'flat' file (FFILE) by the CHS was selected. The FFILE was designed to retain all the 'raw' data in addition to the processing status flags and corrections associated with each sounding and position. It should be noted that the 'raw' sweep data is in a somewhat compressed format. Each position has associated with it thirty-three depths, boom orientation, tide, time, data set number, etc. In addition, each sounding has a 'percentage factor' that is used to interpolate its actual position relative to the current and previous fix. The percentage factor is measured by the depth pre-processor at the time the data is logged. Associated with each depth is a status flag byte whose bits are set in accordance with the various processing and editing functions that are carried out on the data. Figure 3 shows a breakdown of the status flag.

By choosing this approach it is possible to store 'raw' and processed data within the same data structure, even though it is in a somewhat compressed fashion. For the purposes of displaying, editing and plotting the data, it is necessary to expand the file into a new and temporary work file (PFILE). The PFILE contains most of the information within the FFILE and is considerably larger as each depth is given a unique position. The data processor may optionally elect to run a pre-selection program on the FFILE to flag redundant data and reduce the size of the PFILE. The selected depths within an overplot removed PFILE may be flagged back to the FFILE. Consequently, it is not essential to save the PFILE as it may be easily recreated at any time by retrieving the appropriately flagged soundings. Should there be a need to examine all data from one or a number of 'flat' files, as may be the case in a contentious area, it is a simple matter to override the appropriate flags in order to extract the desired data.

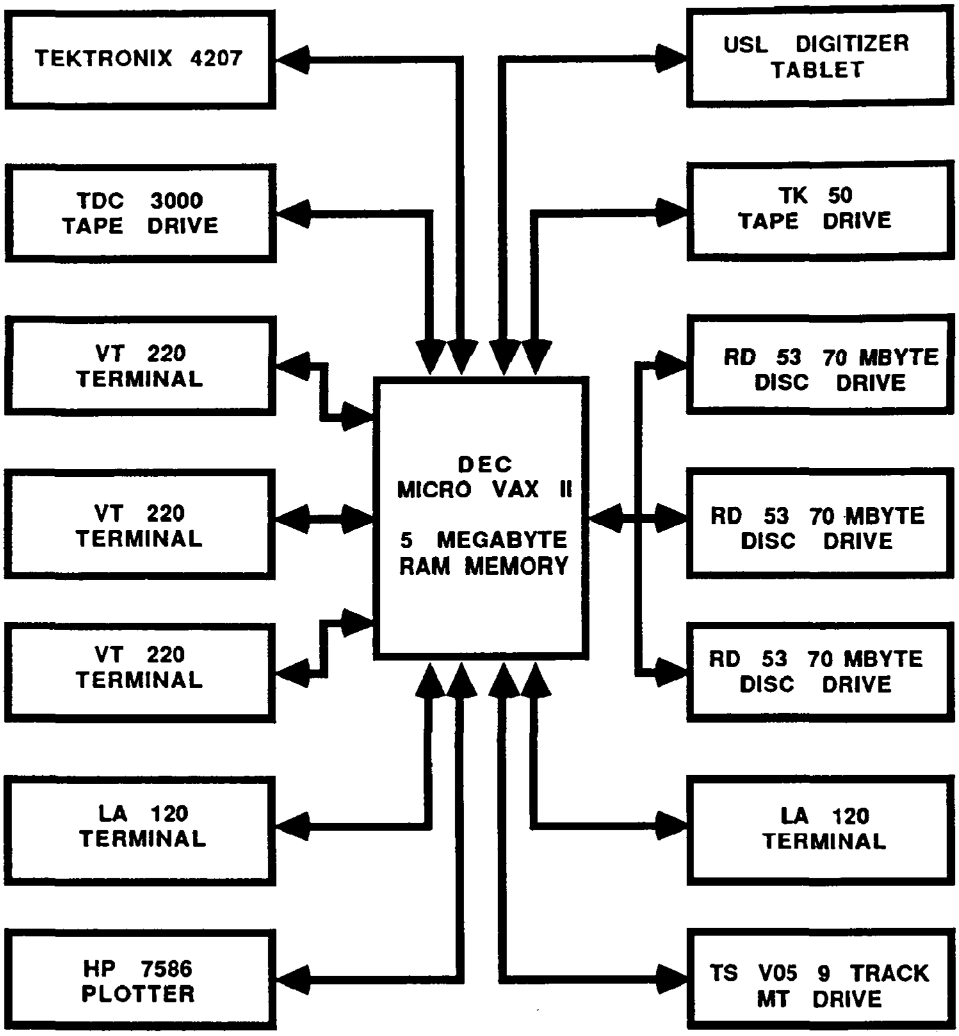


FIG. 2. — Simplified block diagram of the computer system.

Processing Software

A suite of programs has been written covering all stages of the processing, from the initial transfer of the 'raw' survey data logged on the Tandberg TDC 3000 tape recorder to the Micro VAX II computer to the final transfer of the processed data into the NTX format. This latter step enables the data to be used by the CHS Computer Assisted Chart Production system known as CARIS II [5].

The main routines that have been written for data processing are as listed:

1. TANDBERG — transfers logged sweep data to the Micro VAX II disc. The transfer process creates a new disc file known as a 'flat' file (FFILE).

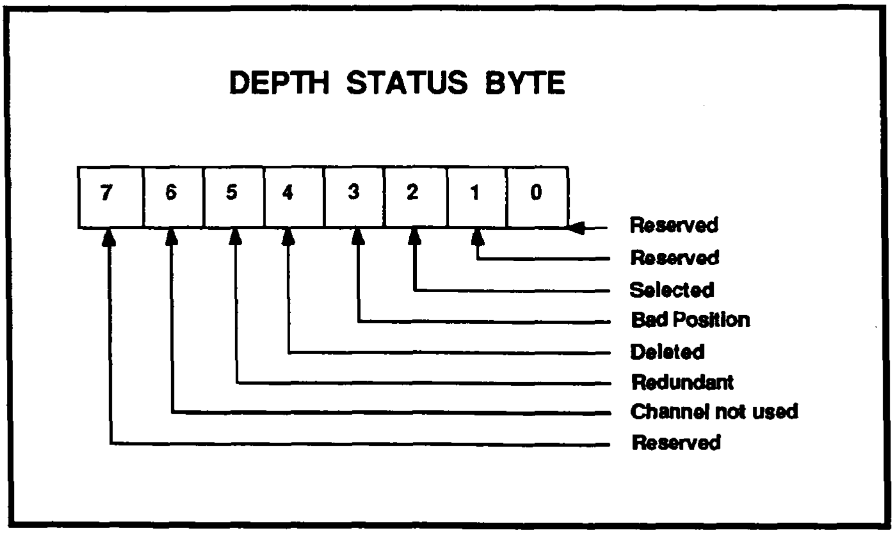


FIG. 3. — Depth status byte.

The FFILE contains all the logged data in addition to a status byte for each depth and fields for additional data and tide corrections.

2. RAWLIST — lists an entire FFILE or any portion of it.
3. RAWMERGE — concatenates FFILES.
4. TIGEN and COTDE — generates, applies, and edits tide corrections to the FFILE.
5. RAWSCAN — scans the FFILE and carries out the following functions:

   5.1 Reports data gaps.

   5.2 Performs critical depth selection. A 3-dimensional search is carried out on the depth data set. Any depths that do not contribute any additional depth information for processing are flagged as redundant.

   5.3 Identifies and reports questionable depths.

   5.4 Flags data in the FFILE that are to be rejected due to bad positions. This function is not carried out on the first pass. Another routine, PCREAT, creates a filter file which is subsequently utilized by RAWSCAN to identify bad positions.

   5.5 Performs user defined edits based on times, limiting depths and transducer number.

6. PCREAT — reads the FFILE and selects depths as determined by the various status bits, computes a unique position for each selected depth and writes the data to another file (PFILE) with a format suitable for plotting bathymetry, the vessel's track, or a swath plot to show the area covered by the sweep survey. The program also uses a position filter to flag bad navigation and a position filter file is created for RAWSCAN to update the 'flat' file.

7. FDATA — utilizes the PFILE to suppress soundings that will overplot at a given scale. The selection algorithm is shoal biased and the shallowest soundings are flagged for plotting. It is also possible to select and examine a deep bias subset of the same data or merge a number of PFILES into one file.
8. CSORT — sorts the file (PFILE) on the basis of user-selected depth intervals. Each depth interval may be represented by a unique colour. This program is required because the HP 7586B is a carousel pen plotter and frequent pen changes result in grossly exaggerated plot times.
9. PMANUAL — provides for the manual entry of buoys, control points, rocks or depths.
10. DAPLT and COPYPLOT — creates and plots vessel survey tracks, selected depths, swath coverage, buoys, rocks and control points.
11. PLOCATE — locates, using a plotted field sheet of selected soundings and a digitizer tablet, any questionable sounding requiring further verification or checking. All depth data at that location from each survey pass can be quickly located and displayed for editing. This program requires that all the FFILES for the data in question be on line as well as the merged PFILES with overplot suppression carried out.
12. RAWUPDAT — maps back to the FFILE the selection and processing status of all corresponding soundings in the PFILE.

A simplified block diagram of the processing is shown in Figure 4. RAWUPDAT is not shown on the diagram as it can normally be run at a number of different stages in the processing.

Data Processing

The first task that must be performed before any data processing can take place is the transfer of data to the Micro VAX II (Fig. 4). Once this is accomplished, tide corrections are added to the FFILE. The normal practice on larger projects is to 'pre-process' each day of data to 'clean' and edit it prior to merging with the files from the previous days. RAWSCAN is used to flag redundant data, report data gaps and edit bad or unwanted data. It should be noted that data is never discarded. It is flagged and may be recalled at a later date by instructing the various programs to override the appropriate status bits in the depth status byte.

Normally, the 'pre-processing' is done overnight in a batch mode with the aid of procedure files. The process is started at the end of the survey day and a track plot, swath plot and overplot-removed bathymetric plot are available the next morning (Fig. 5, 6 and 7). The first look 'pre-processing' takes 1 to 5 hours.

Once the 'quick look' has been completed, the editing and verification normally require considerably more time and go through several iterations. The amount of effort required is dependent upon factors such as bottom complexity, frequency of false echoes, quality of positioning, etc., and varies widely from data

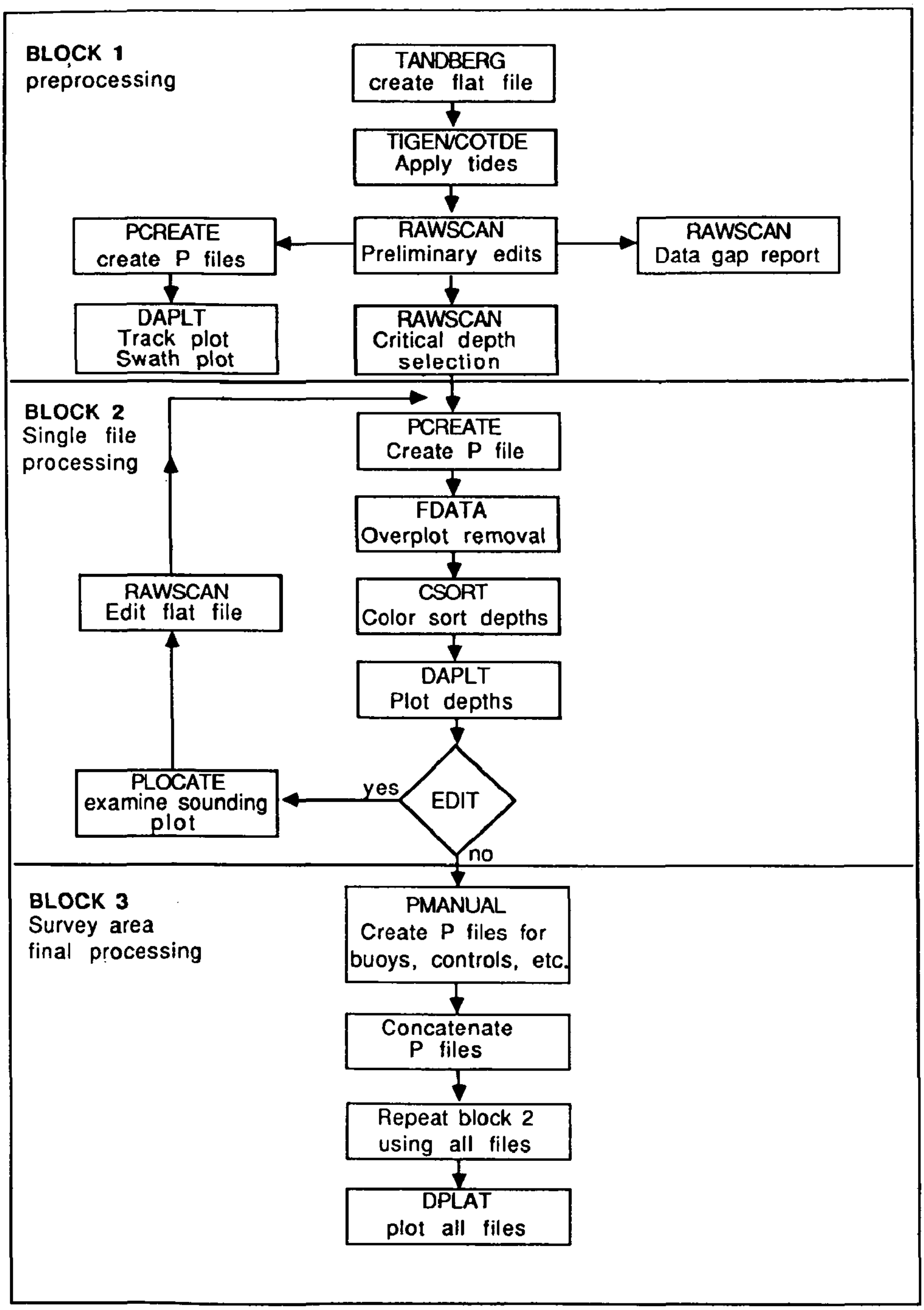


FIG. 4. — Simplified block diagram of the data processing.

set to data set. At this stage in the data processing, one can examine the echograms generated at the time the data was logged by the Navitronic Sweep

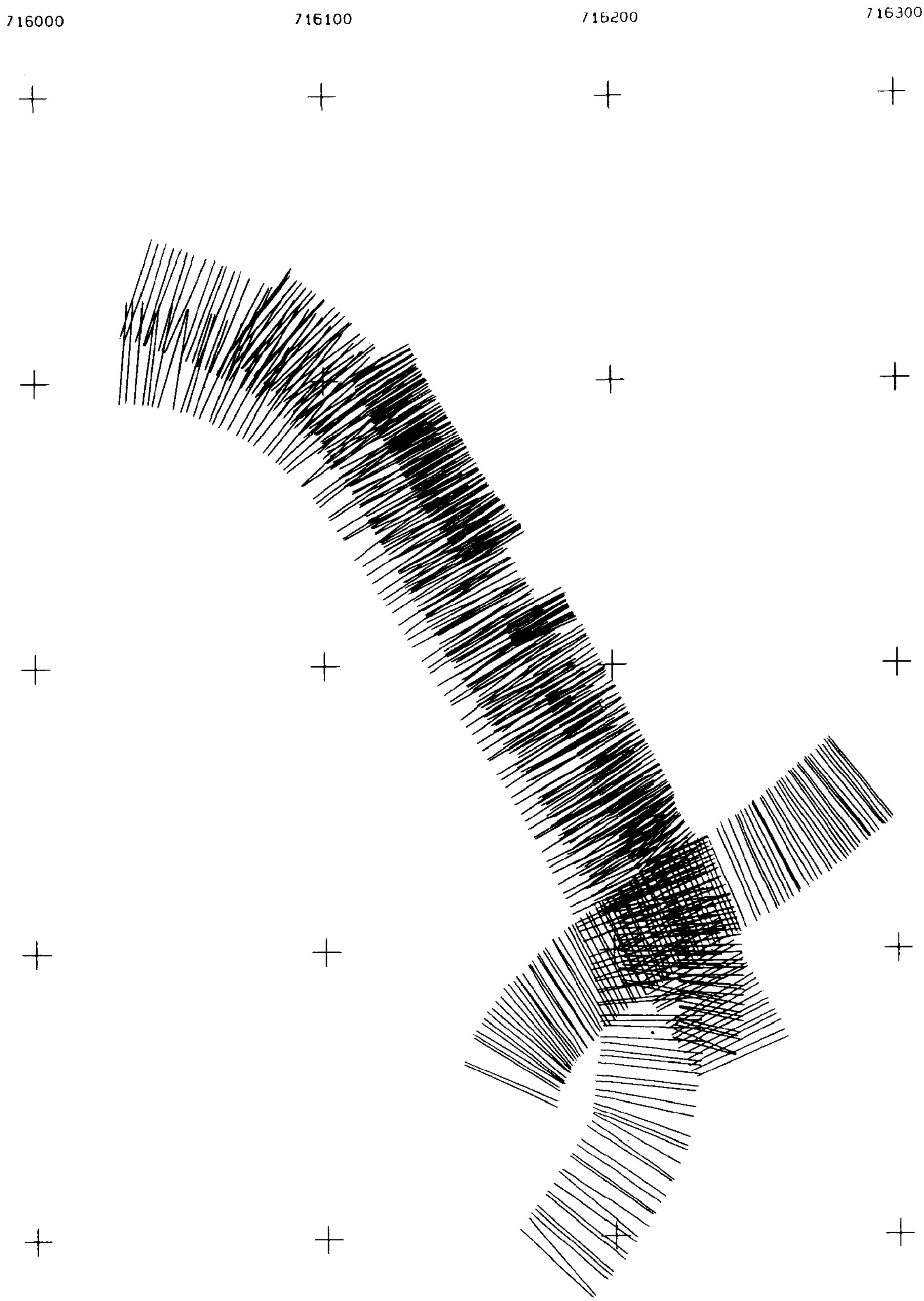


FIG. 5. — Swath plot.

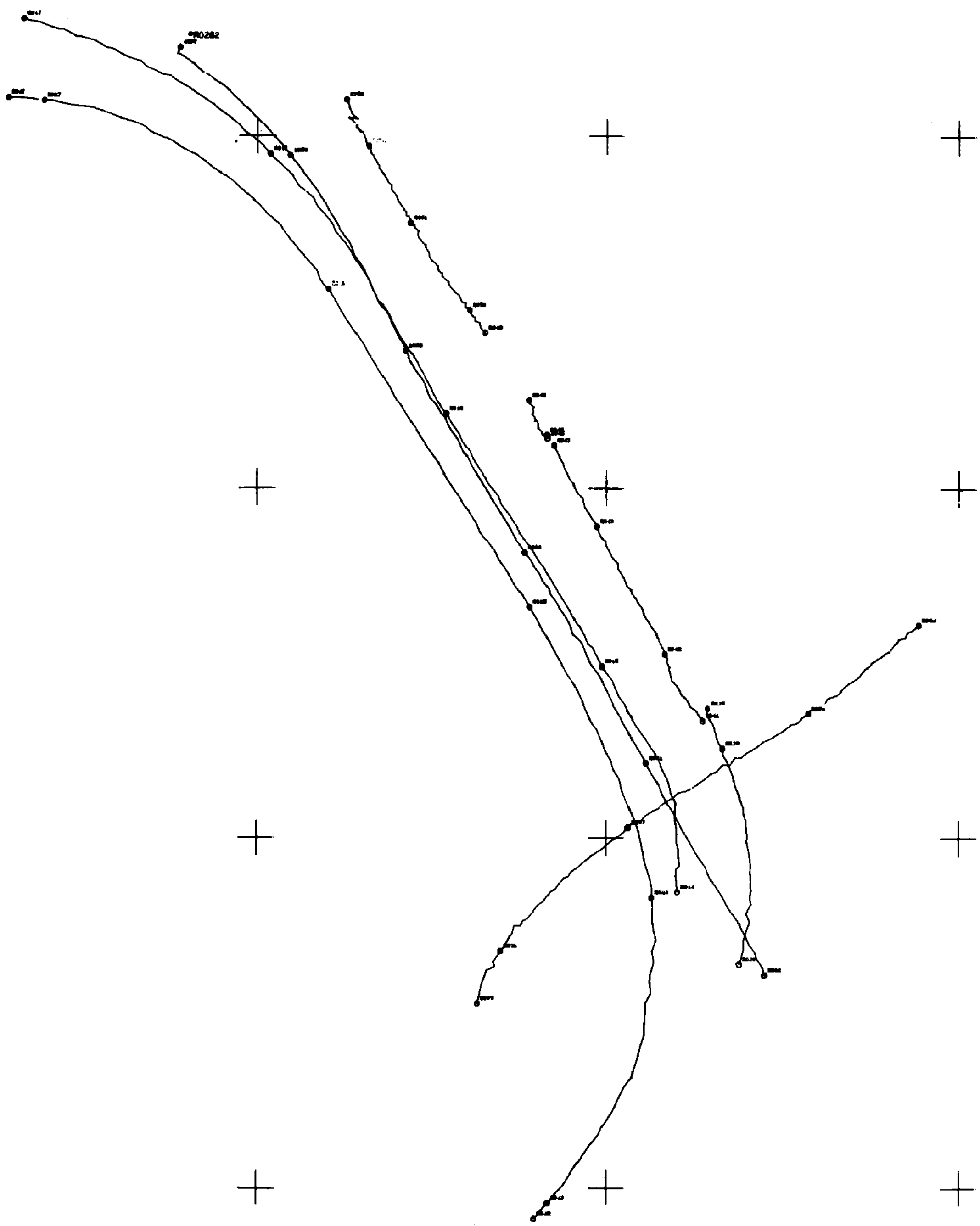


FIG. 6. — Track plot.

System as well as use PLOCATE and the digitizer tablet to point to and identify soundings in question.

The program allows the user to identify and print all the soundings within a given 'box' centered on the point being examined. Normally, the surveys are carried

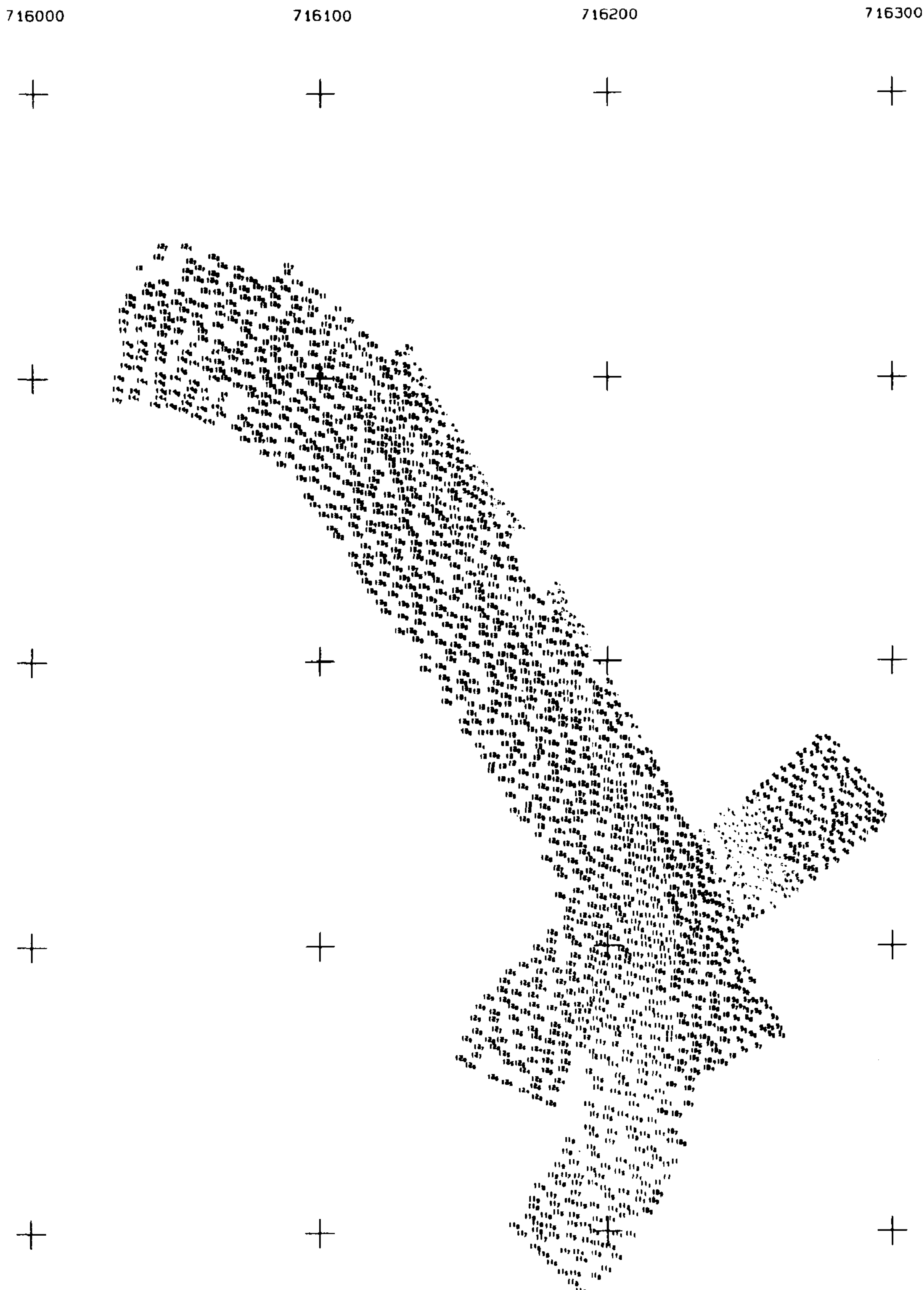


FIG. 7. — Bathymetric plot.

out with considerable overlap and there are two or more passes over every point in the survey area. The ability to find and list a pass by pass presentation of the critical depths for a given point is essential in eliminating false returns and confirming the presence of a strong and consistent target. Figure 8 shows a typical output. In addition, it is possible to list all the soundings in the box or individually list all depths for a given pass that falls within the box. PLOCATE may be used to suppress any soundings that are erroneous. After bad depths are suppressed, it is necessary to re-run the overplot-removal program (FDATA) to ensure that a new and correct depth selection is made. PMANUAL can be used to create a file containing buoys, control points, and drying rocks. A colour banded contour plot is normally generated and RAWUPDAT is used to map back the selected soundings to the FFILE, which is subsequently archived.

Typically, at scales of 1/1,000, approximately 0.5% to 0.8% of all logged depths are selected for plotting on the field sheet. It is very difficult to give the precise time required to process a data set because each project is unique. On average, it is estimated that a typical day of survey data collection requires three to five days to process.

MONTH 1 DAY 18 YEAR 88 TIME 08:10:16

PORTAUXBASQUES_HARBOUR
SURVEY SCALE 2500
SUPPRESSION RATE 14

SEARCHING AT DIGITIZED POINT
SEARCHING FROM 79 TO 85 DM
SEARCH WINDOW 6.5 METRES NORTH BY 6.7 METRES EAST

TIME	TX	DAY	N.	E.	PLOT	NAV	SDG	DSN	TIDE	
150331	16	234	5271234	339306	PLOT	ACTUAL	79	565	11	LOOKED FOR
150331	15	234	5271233	339305	SUPPR	ACTUAL	79	565	11	SHOALEST
150326	15	234	5271231	339308	SUPPR	ACTUAL	87	563	11	DEEPEST
170459	9	233	5271230	339307	SUPPR	ACTUAL	80	652	8	SHOALEST
170457	8	233	5271232	339306	SUPPR	ACTUAL	87	651	8	DEEPEST
152637	17	234	5271234	339306	SUPPR	ACTUAL	79	953	10	SHOALEST
152635	15	234	5271231	339308	SUPPR	ACTUAL	86	952	10	DEEPEST
131603	24	236	5271230	339303	SUPPR	ACTUAL	81	995	16	SHOALEST
131606	22	236	5271232	339305	SUPPR	ACTUAL	89	956	16	DEEPEST

32 SOUNDINGS WITHIN OVERPLOT BOX

FIG. 8. — Critical depth summary from PLOCATE.

Problem Areas

Three areas of data acquisition and processing that have created the most difficulties are:

1. *Lack of Disc Space*

The Micro VAX II Computer System was acquired with the maximum amount of disc space available at the time it was purchased (210 Mbyte). The fact that this would cause some difficulties was realized; however, it was anticipated that a larger disc system would be installed at a later date. The current disc capacity limits the amount of data that can be kept on-line to about seven days of survey data. Considerable time is lost in off-line storage and retrieval (up and down loading) of files. A more suitable disc system would have a 1 to 2 gigabyte capacity. Unfortunately, fiscal restraints have prevented the system from being upgraded.

2. *Poor Positioning*

The Atlas Polar Fix positioning system has been plagued with tracking and filter problems. In addition, a positioning bias of up to 4 metres was applied to the logged data as a result of timing and data packing considerations in the logging system.

These problems created data gaps and positioning errors that made it more difficult for the data processor to identify and verify the existence of a target or false echo. Modifications have been recently made to both the Atlas Polar Fix and the logging software that should eliminate or minimize these problems.

3. *False Echoes*

Any digital echo sounder system may generate a number of questionable or false depth measurements. Incorrect depths may result from acoustical or electrical noise, fish, marine plants, suspended sediment, etc. With sweep data one normally has two or more passes over the same area. If a feature occurs consistently, then one can normally assume that a depth value is not the result of acoustic noise or fish.

Each project has its unique problems. On one survey in which the bottom was relatively flat, the sweep system would consistently pick up a number of small targets approximately 0.4 metre in height. The next day, the same thing would happen with the exception that the targets were in different locations. This was very puzzling until it was discovered that the area was notorious for lobster poaching! It is extremely difficult to carry out accurate surveys under such conditions.

The Future

The *FCG Smith* has had two very successful survey seasons. Whilst the system is considered fully operational, there is no question that improvements can be made to the software. Field staff have suggested a number of ideas where changes can make the software more user friendly or efficient. When funds become available the disc capacity will be upgraded. In addition, CARIS II [5] will be utilized for processing field data in the near future. It will expand the system's current limitation of only handling bathymetric data to include the generation and manipulation of all symbols and line work used in the generation of a field sheet or chart.

One area of interest that may hold significant potential for improving the speed and ease of data processing is digital terrain modelling. The PLOCATE outputs from various passes over a given area could be used to generate a digital terrain model. Since the data sets would be small, the data processor would be able to quickly generate and examine individual, averaged, differential, etc., models on a colour CRT. Simple three dimensional presentations should make the tedious data verification process much easier and faster. In addition, it is anticipated that the science of artificial intelligence could be applied beneficially to solving some of the challenges in processing sweep data.

Summary

The CHS has developed a software package for processing 'large' data sets from two vertical acoustic sweep systems. The package has been successfully used during the past two seasons to process an estimated 50 million soundings that have been logged over that period. Based on experience gained to date, there are a number of initiatives that will be undertaken to enhance the data processing through hardware and software modifications, as resources become available.

References

[1] BURKE, R.G. and FORBES, S. (1984): Vertical Acoustic Sweep Systems — 'A New Broom in the Closet' for the Canadian Hydrographic Service. Proceedings of HYDRO '84 (pp. 50-58), April.

[2] KERR, A.J., ADAMS, A.L. and BURKE, R.G. (1987): CSS *FCG Smith* Hydrographic Acoustic Sweep Vessel. Intern. Hydrog. Review, Vol. LXIV, No. 2 (pp. 7-14), July.

[3] BURKE, R.G., FORBES, S. and TRACEY, R. (1981): Portable Semi-Automated Chart Scalers — A Useful Tool for Acquiring Digital Depth Data. Proceedings of the 20th Annual Canadian Hydrographic Conference (pp. 98-102), April.

[4] VARMA, H.P. (1984): An Interactive Graphics Editor for Hydrography. *Lighthouse*, Edition No. 30 (pp. 20-26), November.

[5] EVANGELATOS, T.V. (1985): Upgrade of the Cartographic Workstations. *Lighthouse*, Edition No. 32 (pp. 5-9), November.