Dealing with Increasing Data Volumes and Decreasing Resources

Jan Depner, Barbara Reed[1]
Shannon Byrne, Jeffrey Parker[2]
Mark Paton, Lindsay Gee[3]
Larry A. Mayer, Colin Ware[4]

[1]Naval Oceanographic Office, Stennis Space Center, MS, USA
[2]Science Applications International Corporation, Newport, RI
[3]Interactive Visualization Systems Inc., Fredericton, NB, Canada
[4]Center for Coastal and Ocean Mapping, UNH, Durham, NH, USA

Abstract-The US Naval Oceanographic Office (NAVOCEANO) has recently updated its survey vessels and launches to include the latest generation of high-resolution multibeam and digital side-scan sonar systems, along with state-of-the art ancillary sensors. This has resulted in NAVOCEANO possessing a tremendous ocean observing and mapping capability. However, these systems produce massive amounts of data that must be validated prior to inclusion in various bathymetry, hydrography, and imagery products. It is estimated that the amount of data to be processed will increase by an overwhelming 2000 times above present data quantities.

NAVOCEANO is meeting this challenge on a number of fronts that include a series of hardware and software improvements. The key to meeting the challenge of the massive data volumes was to change the approach that required every data point to be viewed and validated. This was achieved with the replacement of the traditional line-by-line editing approach with an automated cleaning module, and an area-based editor (ABE) integrated with existing commercial off-the-shelf processing and visualization packages.

NAVOCEANO has entered into two Cooperative Research and Development Agreements (CRADAs) - one with Science Applications International Corporation (SAIC), Newport, RI, USA, and the other with Interactive Visualization Systems (IVS), Fredericton, N.B., Canada, to integrate the ABE with SAIC's SABER product and IVS's Fledermaus 3D visualization product. This paper will present an overview of the new approach and data results and metrics of the effort required to process data, including editing, quality control, and product generation for multibeam data utilizing targets from digital imagery data and automated techniques.

I. INTRODUCTION

The inclusion of high-resolution multibeam and digital side scan sonar systems, along with state-of-the-art positioning and attitude sensors and other ancillary sensors on ships and Hydrographic Survey Launches (HSLs) has provided the US Naval Oceanographic Office (NAVOCEANO) with the finest-equipped survey fleet in the world. Along with this tremendous increase in bottom-mapping capability comes a corresponding increase in the amount of data that must be validated prior to inclusion into the various shallow- and deep- water hydrography, bathymetry, and imagery products that NAVOCEANO produces. With ships operating 24 hours a day, 7 days a week, and 10 months a year or more, NAVOCEANO will soon continually collect more data than anyone else in the world. If maximum data rates are used, we could face a potential of a 22-fold increase in the amount of bathymetric data to be processed-a maximum of over 2.75 terabytes per year versus the present level of 125 gigabytes per year. This figure rises to an overwhelming 2400 times the present data quantity (roughly 300 terabytes per year) if multibeam imagery and digital side scan sonar are included. Fig. 1 depicts the expected increase in future bathymetric/hydrographic data volumes and the amount that has been collected over the years that NAVOCEANO has been collecting multibeam data. Notice the prior 30 years of multibeam data barely registers on the graph!

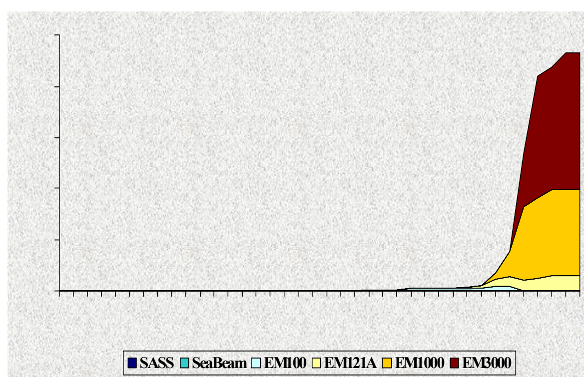

Fig. 1. Expected Bathymetry Data Volumes

II. APPROACH

In an attempt to begin dealing with the increased data volumes, NAVOCEANO embarked on a series of short-term hardware improvements and longer-term software improvements for both shipboard and in-house validation. The short-term improvements were intended to quickly improve existing processes while the longer-term solutions were being developed. The Bathy/Hydro Post-Processing system (BHPP) is a suite of in-house and contractor-developed software that consists of over 400 programs and 1 million lines of code utilized to quality control, validate, and produce bathymetric and hydrographic products. In-house software has been used for over 30 years to process both single and multibeam sonar data. With the decreasing amount of information technology personnel available to develop and maintain this in-house suite of software, alternatives to in-house software development and maintenance had to be investigated. The solution had to allow incorporation of state-of-the-art technology and reduction of the training and maintenance "tail" that accompanies maintaining software resources. Until recently, BHPP was operated primarily on Hewlett-Packard Unix workstations. After analyzing the personnel resources available for quality control and validation of the bathymetry data, an initial goal of 4:1 (collection time:processing time) for a new processing system was identified just to be able to keep up with the amount of data being collected with the new sensors. COTS products available at present provide approximately 1:1. The next goal for collection:processing ratio is 10:1 and is being approached by incorporating automated techniques (approaching artificial intelligence techniques) and upgrading hardware and network components.

Due to the tremendous increase in data we are facing, the key to making major improvements in processing throughput is to change the way we do business–***we can no longer look at every data point***! But, we must have confidence that real hazards to navigation are not invalidated in shallow water, and we must have the capability to examine every data point if necessary. The software improvements planned include three major components: (1) an Area-Based-Editor (ABE), (2) an automated data cleaning module, and (3) the incorporation of these two capabilities with COTS software packages to provide a complete integrated data processing solution. These components are designed to work in conjunction to produce the 4:1 system.

Two Cooperative Research and Development Agreements (CRADAs) are now in place-one with Interactive Visualization Systems (IVS) of Fredericton, N.B. and another with Science Applications International Corporation (SAIC) of Newport, RI - that will integrate NAVOCEANO's tools with each company's commercial product(s). This will allow NAVOCEANO to migrate toward a COTS product that meets our specific requirement. Under the CRADAs, SAIC's SABER (Survey Analysis and area Based EditoR) processing and analysis product and IVS's Fledermaus 3D visualization product will both be integrated with NAVOCEANO's ABE. These products allow the scientist to view the PFM data geographically, review and assess the quality of the fliers identified by the filter, compare bathymetry with other types of data, and perform interactive editing. This approach will provide the scientist with a seamless transition between the full-resolution swath-oriented data, geographically oriented presentations of the data, and interactive 3D visualization of the data. The link between the software packages will be the PFM file structure as shown in Fig. 2 and will allow users to move easily and quickly between the visualization surface and the full-resolution data.

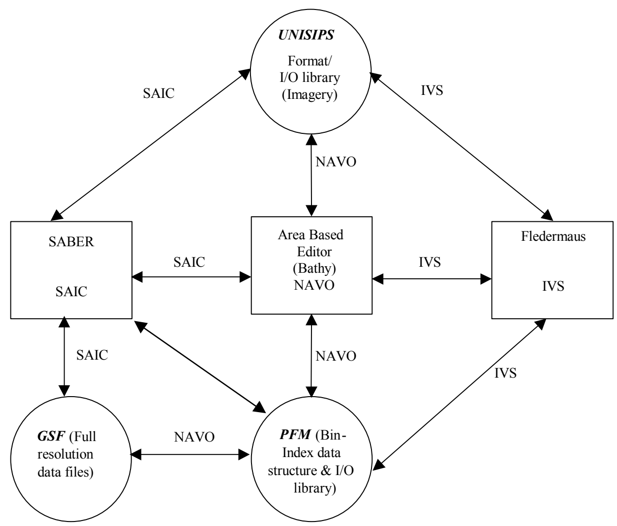

Fig. 2. CRADA Interfaces

III. FLEXIBILITY

It is important to note that PFM data "loaders" and "unloaders" are easily written for almost any native type of data format. This allows NAVOCEANO, and any other users of the software tools, to add new sensors with native data formats and not change their processing tools. In order to load and unload a data type you must be able to extract latitude, longitude, depth, and some form of status information from the data. You must also be able to directly address each individual data record in the native format file. Status flags are available in the PFM format to allow data to be marked as manually edited, filter edited, selected, suspect, contact, class 1, or class 2 (user defined parameters). Translation of these flags to and from the native format is the function of the "loaders" and "unloaders." In the simplest case, such as for ASCII XYZ data, the only status indicator available in the native format is a valid/invalid flag, so the edited flags are mapped to the valid/invalid flag. Only status information actually is written back to the native format on "unload." At present the supported data types are multibeam or single-beam data stored in Generic Sensor Format (GSF) [1], single-beam data stored in NAVOCEANO's "merge" format, SHOALS Lidar data in the Optech .out format, CARIS HDCS format, and various types of data provided in ASCII XYZ format, including the Australian Hydrographic Service Hydrographic Transfer Format (HTF). Data formats that are not presently supported by IVS and SAIC can be added by the end user since the PFM API (application programming interface) library and documentation will be available in future releases of Fledermaus and SABER. The development of a new loader/unloader module is a straightforward software task for most native data formats.

IV. DATA PROCESSING AND VALIDATION FLOW

On NAVOCEANO ships and HSLs, the Integrated Survey System (ISS-60) produces a dataset that includes vessel position, vessel orientation, multibeam and single beam bathymety, sound velocity structure, sea surface temperature, weather data, and various survey metadata. The dataset is incrementally archived from the survey control (or HSL) computer(s) to the shipboard post-processing computers on a per-watch basis for processing and analysis. Seafloor imagery data, acquired using various COTS software packages, are also archived to the shipboard post-processing computers prior to offline analysis. The imagery data are processed using the NAVOCEANO developed UNISIPS (UNIfied Sonar Image Processing System) software package. The SABER and Fledermaus software products operate directly on the outputs of ISS-60 and UNISIPS. Fig. 3 provides the layout of the dataflow pipeline for data acquired on NAVOCEANO platforms.

Interfaces are supported for data that originate from other acquisition and processing systems. Various multibeam sonar formats including Reson, Simrad, SeaBeam, L3, and XTF can be imported directly to GSF, allowing for full reprocessing if required. In addition, shoals LIDAR, HDCS, HTF, and ASCII x,y,z formats may be loaded directly into PFM.

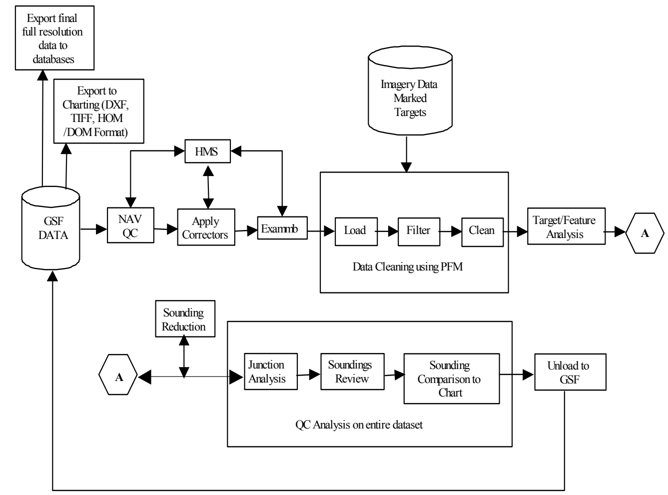

Fig. 3. NAVOCEANO Shipboard Hydrographic/ Bathymetric Processing and Analysis Pipeline.

The huge volumes of data that result from the new sensors rendered the traditional processing and validation obsolete. This does not mean that the requirements for each phase of the workflow have been replaced, but that the time taken in each phase of the work flow and the tools used must change to reflect the data volume. The traditional approach required significant time consuming manual line based processing, and basically attempted to replicate the workflow for a single beam survey with the new multibeam sonars. Data volumes have now demanded that the processing be more automated and less manually intensive and shifted the primary manual interaction to the validation and QC phases of the work flow.

Initial processing tasks include graphic review of the navigation data for quality; review and/or update of data corrections to the data; and record keeping. Navigation problems identified at this stage can be addressed by editing the offending position fixes and/or reconstructing the vessel track from raw data. Once the navigation has been verified, any known corrector issues are resolved. This step is typically limited to water level correction but when necessary can include resolution of a range of potential systematic offsets. Following this, the data are “loaded” from the time/line oriented full resolution data files to the PFM data format. This process consists of binning the depth values into a grid that preserves the minimum, maximum, average, standard deviation, and the number of data points. All values that fall into each bin cell are also preserved and can be directly accessed by position.

A statistical filter can be used to automate identification of erroneous soundings that need to be invalidated from further processing. When available, targets marked from review of the side scan imagery can be used to establish a “no-filter” radius around the target location. For hydrographic purposes, these areas can then be reviewed in detail to correlate the least depth from the bathymetry with the target from the side scan. QC of the data points invalidated by the filter is supported with the visualization of both filtered and unfiltered minimum, maximum, and average surfaces. Each of the PFM bin surfaces are visualized over the area of interest and specific sub-areas may be selected to review all the contributing soundings. The validation/invalidation attributes can be controlled for each data point. Following this process, the bathymetric data are reviewed and cleaned on an area basis, replacing the traditional time or line-oriented approach. This allows the analyst to prioritize their efforts reviewing any problem areas, without having to wade through all the data, resulting in a significant reduction in the effort required for data cleaning. [2]

When the dataset has been cleaned, various analysis and validation processes are supported. An area-based sounding reduction algorithm can be run to identify the hydrographic least depths. The reduced soundings may be compared readily with current charts by superimposing these two data layers. Target positions resulting from review of the imagery data, "snippets" of the detailed target imagery, and side scan mosaics are all supported with the PFM structure. Visualization of the imagery data can add significant value during the cleaning and analysis phases of the bathymetry processing. Survey system repeatability may be assessed using statistical analysis on the area of intersection that occurs between main scheme and cross-lines. When sounding reduction and validation are complete, the "edits" stored in the PFM data structure are "unloaded" back to the full resolution data files. Hydrographic and bathymetric database product generation typically occur from the full resolution data files as the final step.

It should be noted that the while Fig. 3 shows a stepwise processing approach, the combined SAIC and IVS tools do not require that the steps be completed in a specific sequence. This is an important point, as it may be useful to assess cross-check comparisons prior to completing data cleaning. Likewise, final water level corrections may not be available until 30-60 days following the survey, so final tides may be applied after data cleaning is completed. The ability to generate 3D visualizations of the bathymetry early in the process greatly assists with data interpretation and with identification of areas that need to be investigated.

V. SABER

SAIC's SABER product is the post-processing component of SAIC's Integrated Survey System (ISS2000), which, in addition to post-processing and charting capabilities, offers mission planning, and real-time data acquisition and survey control components. SABER is designed to process hydrographic and bathymetric data in support of nautical charting surveys, seafloor characterization surveys, pipeline and cable route surveys, search and locate surveys, as well as general bathymetric and engineering surveys [3], [4], [5], [6]. SABER provides an intuitive approach capable of handling today's large seafloor datasets in a rigorous and efficient manner. The approach supports a standardized processing flow to ensure the consistency of results across a spectrum of analyst capabilities and supports all the functions depicted in Fig. 3 and described in the previous section. Full traceability of all corrections and modifications to the data are maintained. Quantitative quality assurance is provided to fully characterize processing results. All these factors support the objectives of streamlining the effort required to produce data products and minimizing survey rework when these processes are applied in the shipboard environment.

SABER operates directly on the dataset produced by the ISS-60 system, thereby ensuring the most efficient access to the data. The survey events log may be used to populate daily report templates and to summarize the progress of the survey and the processing. This processing audit and support element is referred to as the hydrographic management subsystem (HMS), shown in Fig. 3. SABER supports direct operation on the GSF files, and allows for the correction of systematic errors such as antenna and transducer lever arms, pitch, roll, and gyro biases, draft, and tide. Recent updates to support recording the full time series of motion data will allow for re-computation of the ship relative soundings from the raw travel time angle pairs for those (hopefully infrequent) occasions when this may be required. If necessary, positioning problems may be resolved by recomputing the vessel track from the raw positioning data using a forward-backward Kalman filter. Updates to the vessel position time series may be re-applied to the bathymetry data as required. All these low-level correction operations are accessible via the SABER graphical user interface.

Fig. 4 shows the main SABER screen displaying a Reson 8125 dataset from Portsmouth Harbor, New Hampshire. This 2D shaded relief visualization of the bathymetry is from the minimum filtered surface of a 0.5 meter PFM binned using a UTM projection. The survey vessel track lines are superimposed on the bathymetry. Similar presentations are available for the maximum depth surface, average depth surface, standard deviation surface, number of observations surface, and for the unfiltered versions of the minimum, maximum, and average. The ability to quickly visualize each of these surfaces allows for rapid identification of the areas that need to be investigated. The standard deviation surface visualized with the superimposed track lines can be particularly useful for locating potential problem areas as identified by patterns that are aligned parallel or perpendicular to the data collection azimuth. This surface can also help identify real seafloor features that may not be readily apparent when viewing the minimum, maximum, or average surfaces.

In Fig. 4, the far right-hand portion of the screen provides a coverage display. The area shown in black delineates the extents of the PFM file, while the rectangular outline near the bottom of the area covered locates the current area visualized in color-coded shaded relief. The rectangular area outlined near the bottom of the shaded relief display defines a sub-area that has been selected to review all contributing data points. Fig. 5 shows the SABER multi-view-editor (MVE) loaded with all soundings that were made in the selected sub-area. This interface allows the analyst to review all soundings, assess and modify any invalidations made by the automated filter, manually validate or invalidate any spurious points, as well as review sounding selections. From the MVE, any sounding can be traced back to the supporting source data file with the click of a button. Fig. 6 shows the interface that is displayed when the source data point is from a GSF file. Targets derived from the side scan imagery can be displayed and correlated to the appropriate least depths in the multibeam bathymetry. When appropriate, these least depth values can be identified as "feature" selections for the smooth sheet. The edits resulting from validity and sounding selection changes are saved as changes to Boolean flags in the PFM data structure.

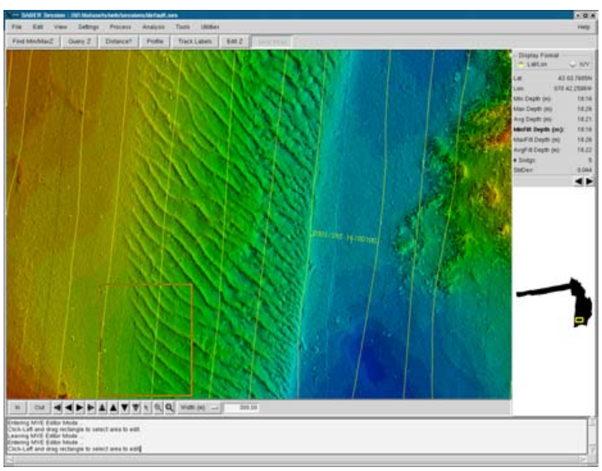

Fig. 4. SABER display showing Dual-Head Reson 8125 dataset from Portsmouth Harbor, New Hampshire.

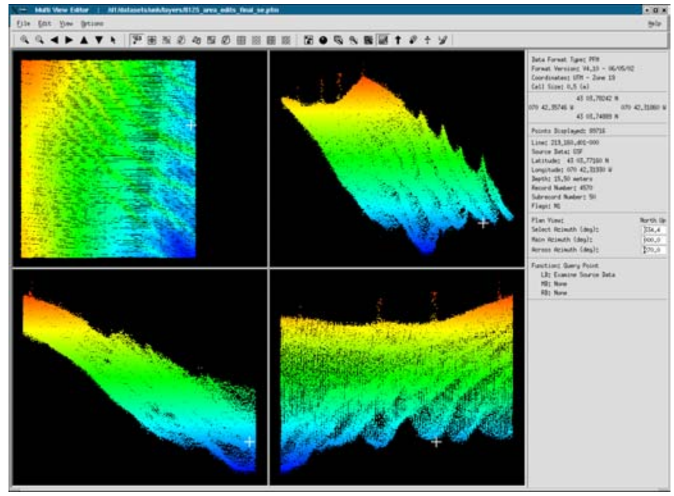

Fig. 5. SABER Multi-View-Editor showing sand waves and a series of four lobster traps.

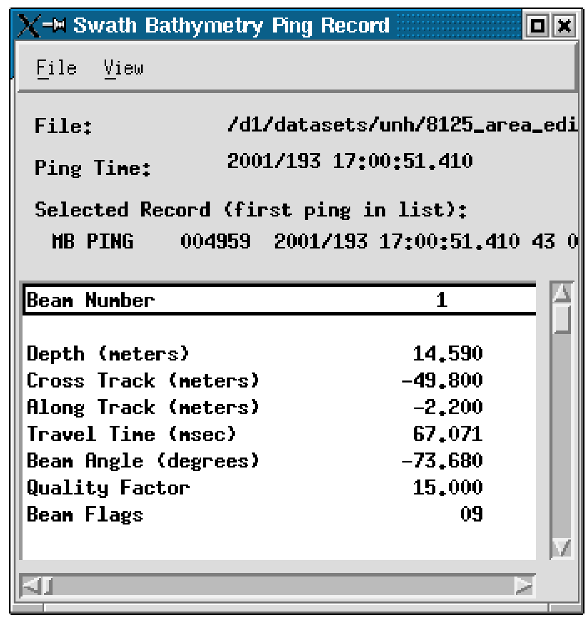

Fig. 6. SABER exammb program launched from MVE to support data point traceability back to the source data file.

The sounding reduction algorithm uses an area-based approach to reduce the full dataset to a subset that preserves the significant seafloor features using a shoal-biased approach. Soundings that have been selected are highlighted in the MVE display, allowing for detailed review and QC against all neighboring data points. Selected soundings may be displayed on the SABER screen in text form superimposed on the current chart for further review and QC. Fig. 7 provides an example of this display.

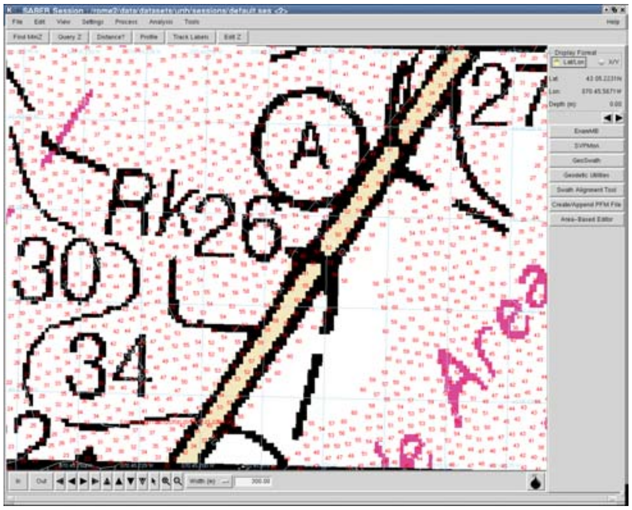

Fig. 7. Reduced Soundings Compared with the Chart.

SABER provides several quantitative QC utilities. Once such utility, referred to as **accutest**, allows for the evaluation of individual beams of a multibeam echo sounder for compliance with hydrographic accuracy standards. Because it is not possible to have a reference with true depths and positions in a marine operations area, a reference surface is generated from the near nadir beams of the multibeam sounder. Survey lines for evaluation are run across the reference area so that all beams may be compared to the surface. **Accutest** provides a tabular summary of data sorted by the beam number of the multibeam echo sounder. This summary includes the total number of comparisons, the number of comparisons meeting a difference criteria set by the analyst, the mean absolute difference, the mean positive difference, and the mean negative difference along with the number of comparisons for each category. In addition to providing an evaluation of each beam, these results also allow the hydrographer to evaluate the adequacy of all corrections applied to the data. This tool has proven useful for evaluating newly installed or upgraded systems prior to starting survey operations and for assessing the suitability of all corrections that have been applied to the data during survey operations.

On completion of sounding reduction and QC analysis, the edits, which have so far been saved to the PFM data structure, are "unloaded" to the native data format. In the case of the GSF data files, primary product generation occurs from the updated GSF data files. This includes the exporting of selected soundings and soundings marked as features to COTS Electronic Navigational Chart (ENC) and Digital Nautical Chart (DNC) production packages. For certain data products, the cleaned and validated full resolution native data files are input into various databases.

VI. FLEDERMAUS

The massive amounts of data that NAVOCEANO vessels are producing with the new systems present many challenges: the establishment of the quality of the data acquired and then the interaction, integration, interpretation, and presentation of the data. If properly handled, the inherent density of the data available also presents tremendous opportunities. Integrating inter-active 3D visualization into the overall processing system allows the operators to take full advantage of this data density and, in doing so, allows them to interact with, explore, and analyze, complex multidimensional data. When properly geo-referenced and treated, these complex data sets can be presented in a natural and intuitive manner that allows the simple integration of multiple components from various sensors without compromise to the quantitative aspects of the data .[7]

The human visual system has an enormous capacity for receiving and interpreting data quickly and efficiently and therefore must be an integral part of any effort to understand complex data. The key is to be able to present the data in as intuitive a fashion as possible. The more intuitive the presentation, the more rapidly data are interpreted and the more new information can be extracted from that data. These elements have been incorporated in the Fledermaus interactive 3D software application [8]. The software was specifically designed to facilitate the interpretation and analysis of very large, complex, multi-component spatial data sets. Fig. 8 illustrates the simple integration of multiple components without compromise to the quantitative aspects of the data.

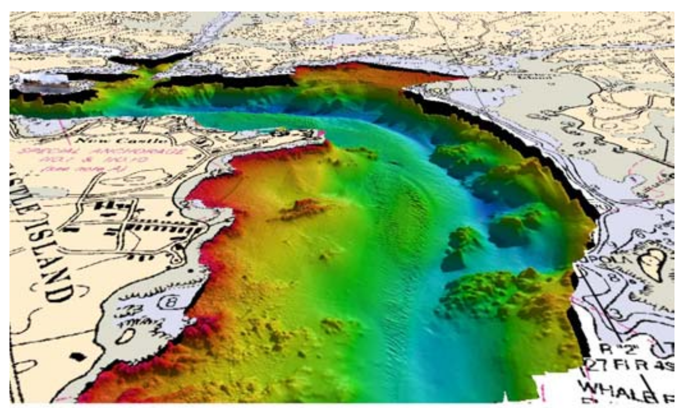

Fig. 8. Portsmouth – chart and combined surveys.

The traditional use of 3D data visualization was in the latter stages of processing when the majority of the processing had been completed. This is under-utilizing the power of 3D data visualization and not improving the efficiency of the processing pipeline. The use of visualization of the data on an area basis, instead of the normal line by line process, in the initial stages of processing immediately provides the general character and structure of the sea floor that has been mapped. It also intuitively shows the areas that will be more difficult to process and any areas of importance. In this fashion, the 3D scene focuses the user to the critical areas and improves the efficiency of the data processing and analysis. An important element to integrate the 3D area based view was to provide a data structure (PFM) that allowed direct access back to the underlying full resolution data from the information presented in the 3D scene.

The PFM structure has been incorporated in the Fledermaus software as an object, thus allowing any PFM object to be viewed and processed in an intuitive 3D environment, along with any other associated data object. These data objects include such things as a surface from previous surveys, a geo-referenced image of the chart, map or aerial photograph or satellite image, or imported boundary line objects such as a coastline or contour.

The PFM object is the most complex of Fledermaus objects and also includes a 3D editor for viewing and editing of the individual data points. It is important to note that although the 3D editor provides an intuitive interface for the data points, Fig. 9, the aim of the revised work flow is to automate the processing as much as possible and therefore minimize the actual amount of editing in 3D.

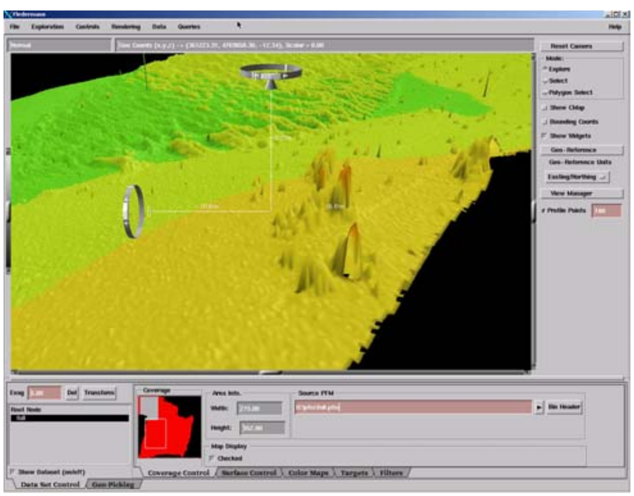

Fig. 9. A 3D view of the shallow binned surface of the Reson 8125 data in the PFM structure with the coverage map lower left.

The PFM object is accessed through the overall coverage map and from here, the user may load a region into the main 3D visualization window. Variations in the accuracy of a survey, from either systematic or random errors, are defined clearly. Outliers are readily apparent, and survey coverage can be reviewed easily. The intuitive presentation is also supported with the ability to dynamically change the surface between shallow, average, and deep, either edited or including all deleted data. Edited and selected soundings can be flagged and the surface colored by a variety of attributes for each bin in addition to depth – density, standard deviation, edited, and class or confidence values. A 3D view with the surface colored by data density is shown in Fig 10.

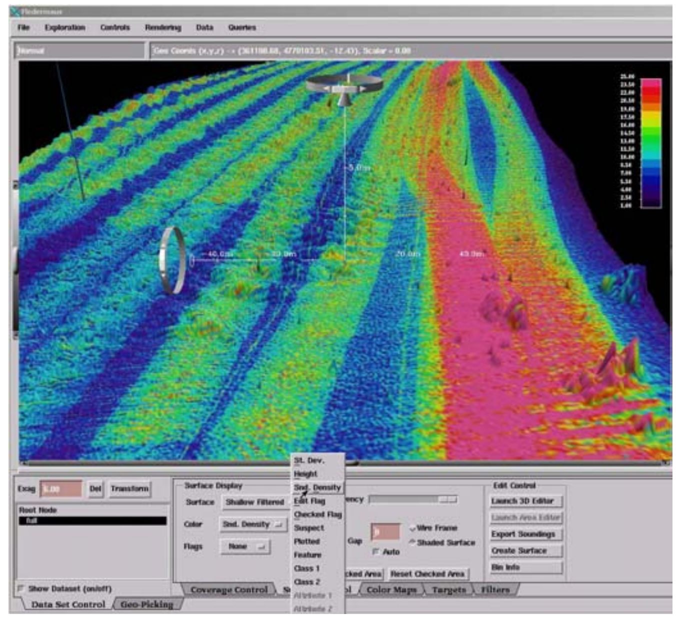

Fig 10. Another view of the same Reson 8125 data with the surface colored by the data density

When outliers or features are detected that require further analysis, a sub-area is selected, and the individual data points presented in the 3D editor. See Fig. 11 and 12. The 3D editor has a similar interactive exploration interface as the main visualization window and allows the soundings to be colored by depth or attributes – file, line, ping, and beam. Deleted sounding can be shown or hidden in the display, as can the various binned surfaces. The editor also allows selection of soundings by various criteria and deletion or selection as significant features. Further details of the selected soundings can also be obtained by drilling down further to the native files formats of the original data.

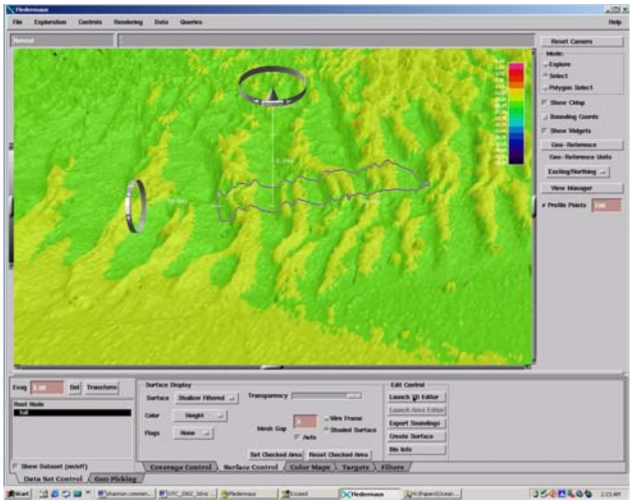

Fig. 11. Selection of area in main Fledermaus window for load of soundings in the 3D Editor.

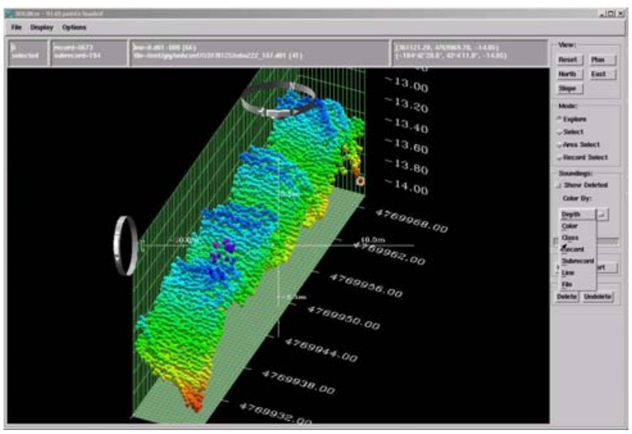

Fig. 12. Fledermaus 3D Editor with small section of sandwaves loaded and colored by depth.

The PFM structure includes support of target files for features that have been identified during data acquisition. Fledermaus displays these as point objects in the 3D scene and allows concurrent display of any side scan sonar image snippets to aid in the verification and feature classification, as shown in Fig. 13.

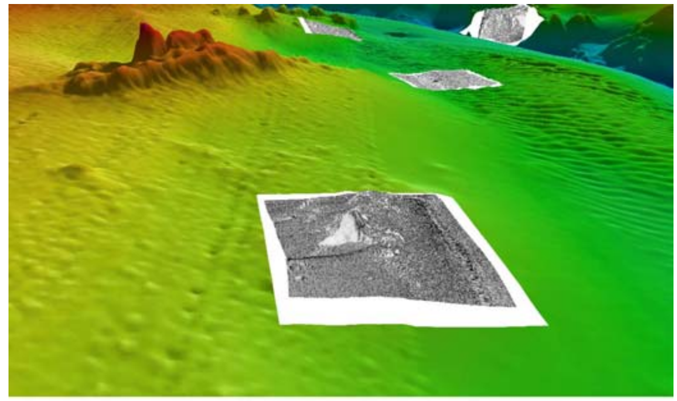

Fig. 13. A scene of Portsmouth Harbor, New Hampshire. The main DEM is colored by depth from a multibeam sonar survey. Draped on the surface are a number of side scan sonar images.

The challenge of the ever-increasing data volumes makes it essential to become more efficient. 3D visualization is a significant element of meeting this challenge, and when integrated in the overall process, can produce value in areas such as efficiency, accuracy, completeness, integration, and communication. Variations in the accuracy of a survey, from either systematic or random errors, are clearly defined. This provides an assessment of whether the survey meets requirements. Early detection of any out-of-limits condition not only will improve the accuracy of the survey but also will provide greater efficiency by limiting the requirement for re-surveying. Visualization provides the complete picture of all the data gathered during the survey or available from other sources, and it is inevitable that without the complete picture, maximum value will not be obtained from data, and errors will be made in the processing.

VII. METRICS

In support of the Second International Conference on High-Resolution Surveys in Shallow Water, and working in cooperation with Reson Inc. and the University of New Hampshire's Center for Coastal and Ocean Mapping (CCOM), SAIC conducted a survey of the common data set test area in Portsmouth Harbor. This survey was conducted over a five-day period in early June 2001 using a Reson 8125 dual head multibeam sonar and a POS/MV. The 8125 Portsmouth Harbor data was selected to assess the overall reduction in effort that can be achieved by using the area-based processing approach versus the line oriented processing approach. An analysis was conducted to quantify the resources required to process and analyze this high density shallow-water bathymetric dataset using SAIC's SABER software product. The analysis included a demonstration of the improvement in processing effort that results from applying an area-based filtering/review approach to data cleaning versus a traditional line-by-line edit/review approach. In addition to the effort comparison, this analysis included a comparison of several of the multibeam datasets acquired over the test area in Portsmouth, NH including Reson 8125, Reson 8101, and Simrad EM3000D datasets.

Table 1 provides an overview of the area covered and the relative size of the bathymetric datasets for the 8125, 8101, and the EM3000D.

Table 1. Summary of survey metrics for the 8125, 8101, and EM3000D datasets.

	Reson 8125 (Dual)	**Reson 8101**	**Simrad EM3000 (Dual)**
Area Covered	2.5 Km^2	4.0 Km^2	1.2 Km^2
Depth Range	2.0 – 25 m	1.0 – 25 m	2.0 – 23 m
Time On-Line	10.6 hours	20.9 hours	2.7 hours
Elapsed Survey Time	4 days	7 days	1 day
Total Pings	433,415	564,702	39,891
Total Soundings	104,019,600	57,034,902	10,132,314
Total Bathy Files	117 files	379	43 files
Size of Bathymetric Dataset	1.3 gigabytes	716 Megabytes	161 Megabytes

Table 2 provides a summary of the effort required to complete each step in the processing. Steps that required continual operator interaction are marked as such. The dual Pentium III 550 MHz computers used for this analysis were running RedHat Linux version 7.1. The line-oriented data-cleaning step required nearly 40 hours to complete. For the high density 8125 dataset, the effort for this step alone exceeded the time required to conduct the survey operations (4 days x 8 hours/day). With the area-based approach the interactive portion of data-cleaning required four hours to complete, clearly providing a significant improvement. These figures result in an acquisition to processing ratio of approximately 1:1.3 (32:42) for the area-based approach, and 1:2.4 (32:77) for the line-oriented approach. Processing is defined to include the effort associated with the generation of a clean, corrected, and validated dataset, but does not include the effort associated with product generation. It should be noted that the data rates from the dual head 8125 configuration are approximately twice as large as the data rates for the highest data volume system on which the 4:1 goal was established.

Table 2. Summary of the effort required to complete processing

Processing Step	**Common Steps**	**Area-Based Specific Steps**	**Line-Oriented Specific Steps**
Convert XTF -> GSF	2 hours (non-interactive)		
Establish tide correctors by zone	2 hours (interactive)		
Apply preliminary tides	2 hours (non-interactive)		
Grid and initial QC	8 hours (interactive)		
Load PFM data files		4 hours (non-interactive)	NA
Initial editing		4 hours (interactive)	39 hours (interactive)
Unload PFM		5 hours (non-interactive)	NA
Apply verified tides (zones already defined)	2 hours (non-interactive)		
Review 3D visualization	4 hours (interactive)		
Junction Analysis and review	8 hours (interactive)		
Update Edits/ Correctors	8 hours (interactive)		
Sounding Selection	2 hours (non-interactive)		
Comparison to chart	8 hours (interactive)		
Total Interactive Effort	38 hours	42 hours	77 hours

As expected, a significant improvement in the total effort required to clean and validate the 8125 dataset was observed with the area-based filter and editor versus a traditional line-oriented approach. This was due to the success of the automatic filtering techniques and the area-based approach to interactive data cleaning. For the 8125 dataset, a ten-fold decrease in the effort required for the interactive editing step was observed with the area-based approach. With a beam width of 0.5 degrees by 1.0 degrees, the 8125 produces a very dense dataset, and in the depth range of this survey, allows for multiple observations on each seafloor feature that might prove to be significant for hydrography. For this particular dataset, the input of targets detected from the side-scan sonar to the automated filter did not prove to be necessary. Of those reviewed, all targets detected on the side-scan sonar had multiple coherent bathymetric observations from the 8125 and therefore were not identified as suspect by the automated filter. The integration of targets from a side-scan sonar is generally considered an important input to the automated filter in order to ensure that known targets/features are not identified as suspect by the filter. If collection of side-scan data is not practical, the automated filtering step may be bypassed, with the advantages of the area-based interactive cleaning still fully supported.

The line-by-line approach duplicates effort where data are over-sampled, or over-lapping. The area-based approach does not result in this duplication of effort where data are over-sampled, and the area-based approach provides the analyst with a visualization of the surrounding data to assist with data cleaning decision making. The latter point was demonstrated to be significant with the observation that during the line-oriented data cleaning, lobster traps were interpreted to be spurious data and were invalidated, while these features were recognizable as lobster traps in the area-based environment.

A more realistic view of the timesavings from the area-based approach to editing the data can be achieved by comparing just the interactive portions of the editing process. Non-interactive processes such as filtering, loading, unloading, and sounding selection can be made much faster just by using faster hardware. An increase in load speed by a factor of 2 for a 1.3 billion sounding data set was achieved by using dual 866 MHz processors and RAID-5 IDE storage instead of the dual 450MHz and Ultra-2 SCSI storage. In addition, since an operator is not required to monitor these types of processes, this time can be used by the operator to process other data sets. The interactive processing time can be reduced by hardware and software upgrades but not on the same order as non-interactive processes. Comparing the interactive processing times of the old method to the new method, it can be seen that the time reduction was actually on the order of 10:1.

It is now critical that the remaining interactive processing steps are examined and additional efforts made to further reduce operator interaction.

VIII. FUTURE IMPROVEMENTS

The new PFM data structure and software tools that have emerged from the CRADAs, in concert with the network and hardware improvements at NAVOCEANO, have created a significant change in the overall processing workflow that is well-suited for the huge increase in data volumes generated by the new sensors. An important aspect of the work has been to establish the basis for a revised workflow and incorporate an automatic processing and outlier detection in place of the line-based manual process. Automated processing approaches are being combined with 3D visualization to provide an intuitive and efficient validation and QC component. SABER and Fledermaus will soon be tightly coupled so that while processing in SABER, a right mouse click will display the current data in 3D.

The efficiency and interoperability of the PFM structure allows the easy incorporation of developments such as the Combined Uncertainty and Bathymetry Estimator (CUBE) program [9] and work on the "Navigation Surface", both being developed at the Center for Coastal and Ocean Mapping/NOAA UNH Joint Hydrographic Center (CCOM/JHC) at the University of New Hampshire. [10]

Research at CCOM/JHC has shown that a significant improvement in the speed and objectivity of hydrographic data processing are possible and that the use of a Digital Terrain Model for charting has significant benefits. The CUBE approach is an attempt to deal with the majority of data processing requirements as automatically as possible. It incorporates robust statistical estimation techniques with an error model describing the MBES data being gathered, and produces as output an estimate of depth and an estimate of the uncertainty in the depth estimates, arranged in a grid over the area being surveyed. To make the estimation more robust, the model has been extended to allow multiple hypotheses about the true depth. Typically, this indicates a potential problem with the data being gathered, and a count of the number of hypotheses is used as an indicator of where to concentrate manual validation and QC. The intention is to provide a surface that gives the 'best' estimate of the true depth in any area, along with a confidence of the estimate.

This Navigation Surface process eliminates the process of shoal-biased selected soundings in fair sheet production and replaces it with a complete model of the seafloor including an uncertainty surface. The model is adjusted at critical places to exactly match the shoalest measured sounding. The result is a model on which automatic cartographic processes can be run to create a set of cartographic objects appropriate to any scale product in any measurement system. Preliminary results indicate that significant timesavings can be achieved, primarily through decreased manual cleaning and through automatic cartographic techniques. The Navigation Surface, created on the ship, is designed to be maintained and utilized through to chart production and in validation tests to date has been shown to provide a range of usable products. [10]

Additional future improvements are in the areas of automatic target detection and classification presentation, and databasing of a "smartly decimated" dataset, while maintaining data management and effective data storage of full-resolution files.

IX. CONCLUSION

NAVOCEANO has unique requirements, in part because of the sheer volume of data that must be continually processed. Obviously, the data rates are going to continue to increase, and we must continually push throughput rates and decrease the turnaround time from data collection to generation of a product. These software improvements are the first step in improving those processes, and we must include R&D efforts in such things as artificial intelligence to make the next jumps in progress. Integration of NAVOCEANO's ABE with SAIC's SABER and IVS's Fledermaus have produced a significant improvement over prior capabilities. The resulting tools will be mutually beneficial and will provide NAVOCEANO with a COTS solution that addresses the requirements for data throughput and product generation turnaround time.

Partnerships with industry–particularly the CRADAs with IVS and SAIC-have enabled NAVOCEANO to move toward COTS packages while utilizing the benefits of over 30 years of government data processing experience.

Future Improvements, including fast automated smart filtering (CUBE), Navigation Surface (to simplify and provide more robust, multi-purpose products), and incorporating ever-improving hardware and network improvements will pave the way on the path of 10:1+throughput speed.

The combination of the unique PFM data structure developed by NAVOCEANO and the area based visualization, and analysis tools, provide an efficient and accurate method for completing data cleaning, QC, and validation, while preserving the ability to access the individual soundings and native data formats.

REFERENCES

[1]. **Ferguson, J.S. and Chayes, D.A.** "Use of a Generic Sensor Format to Store Multibeam Data," *Marine Geodesy*, Volume 18, pp. 299-315, 1995.

[2]. **Byrne, J. S., Clifford, B., Simmons, W., Depner, J., Reed, B. Moestikiwati, J., Smith, G.** "Processing Data for Seafloor Mapping: Integration and Metrics" MTS Journal Vol. 35, No.4

[3]. **Evans, R.E. and Simmons, W.S.** "Outsourcing of Coastal Hydrographic Survey An Industry Perspective of a Partnership With the Government," Hydro International, March, 2001, pp 46 – 49.

[4]. **Evans, R.E., Morton, R.W., and Simmons, W.S.**, "A Dual or Single Vessel Solution to Conducting Multibeam and Sidescan Surveys for NOAA in the Gulf of Mexico: A Lessons Learned Approach," *Proceedings from U.S. Hydrographic Conference*, April, 1999.

[5]. **Parker, J.H., Miller, J.E., and Schoenherr, J.** "Collection and Processing of Multibeam and Side-Scan Data to Hydrographic Standards." *Proceedings of Oceans 96*, September, 1996.

[6]. **Miller, J.E., Ferguson, J.S., Byrne, Simmons, W.S.** "Use of an Integrated Hydrographic Survey System in Long Island Sound and Vineyard Sound," *The Hydrographic Journal,* April, 1996.

[7]. **Mayer, L.A., Paton, M., Gee, L., Gardner, J.V., Ware, C. "**Interactive 3-D Visualization: A tool for seafloor navigation, exploration and engineering" *Proceedings of Oceans 2000,* September, 2000.

[8]. **Reed B., Depner, J. Van Norden, M., Paton, M., Gee, L., Byrne, J.S., Parker, J., Smith, B.** "Innovative Partnerships for Ocean Mapping: Dealing with increasing data volumes and decreasing resources" *Proceedings from U.S. Hydrographic Conference*, May 2001.

[9]. **Calder, B. and Mayer. L.A.**, 2001, "Applications of Uncertainty Modeling to Robustness in Bathymetric Estimation," *Proceedings of the U.S. Hydrographic Conference*, May 2001

[10]. **Calder, B., and Smith, S.**, "A Comparison of the Automated Navigation Surface to Traditional Smooth Sheet Compilation," *Proceedings of the Canadian Hydrographic Conference, Toronto, Canada. 2002*