HYDROGRAPHY FROM FISHERIES SURVEYS

Filling coverage gaps with bathymetry extracted from Simrad EK60 water column data

R. Burkhalter-Castro[1], E. Meyer[1], A. Platel[1], C. Schroeder[1],
G. Rice[2], A. Klemm[1], J. Riley[2], B. Gallagher[2]

Earth Resources Technologies[1], NOAA Hydrographic Systems and Technologies Branch[2]
rosemary.burkhalter-castro@noaa.gov , eric.n.meyer@noaa.gov , arthur.platel@noaa.gov
christian.schroeder@noaa.gov , glen.rice@noaa.gov , anthony.r.klemm@noaa.gov ,
jack.riley@noaa.gov , barry.gallagher@noaa.gov

Abstract

The collection of available fisheries acoustic data holds significant seafloor mapping potential. We estimate that in the Northeast United States, Simrad EK60 data from the NOAA Fisheries fleet could increase bathymetric coverage by as much as nine percent. This article describes initial automated processes and techniques used to: extract the seafloor from Simrad EK60 water column data collected with NOAA Fisheries vessels, obtain and apply best available information to realize the seafloor relative to chart datum, and verify the outcome with existing qualified bathymetry.

Introduction

As of January 2020, the National Oceanic and Atmospheric Administration (NOAA) has classified 54% of the United States coastal, ocean, and Great Lakes waters as unmapped (Office of Coast Survey, 2020). Worldwide, bathymetry is used for charting, modeling, and coastal and ocean zone management purposes. Global bathymetric coverage has numerous scientific and economic applications which are highlighted by the need to understand an ever-changing ocean floor. In 2011, National Centers for Environmental Information (NCEI) partnered with NOAA's National Marine Fisheries Service (NMFS) and the Joint Hydrographic Center to create an archive for water column data collected by government, academic and international vessels. Water column data is the full acoustic time series of an underwater echo sounder. This archive (*NOAA Water Column Sonar Data*) includes data from various areas around the United States Exclusive Economic Zone (EEZ) and provides RAW data files for public use (National Centers for Environmental Information, 2011). NMFS often uses the Simrad EK60 split beam echo sounder during fisheries surveys to observe biological attributes such as marine habitat and animal populations in the water column. Fortunately, this water column data often includes the seafloor and can be used to increase the overall mapped area of the EEZ (National Centers for Environmental Information, 2011).

In 2019, NOAA Office of Coast Survey (OCS) and Coastal Survey Development Lab (CSDL) took an interest in using the NMFS water column data to update coastal bathymetric charts. The National Bathymetric Source (NBS) project has begun to highlight gaps in coverage in the New England region and is motivated to fill these gaps with the best available bathymetry (Wyllie and Rice, 2020). The EK60 data from fisheries surveys was recognized as a potential source for filling these gaps thanks to previous collaboration with Pacific Marine and Environmental Laboratory (PMEL), NMFS, and Saildrone in the Bering Sea (Office of Coast Survey, 2017). Because of ties to the NBS project, Northeast United States waters are the focus region for this work.

Often, hydrographic acoustic survey data is processed in manual steps to integrate supporting information and to derive depths. While supervised processing can be effective for single cruise datasets, a procedure for automating the process for significant amounts of RAW EK60 data is needed. The NOAA Hydrographic Systems and Technologies Branch (HSTB), along with contractors from Earth Resources Technologies (ERT), developed and tested a Python script for batch processing the water column data for derived depths on chart datum. The project's ultimate goal was to develop a way for the EK60 data to be automatically synthesized into forms easily consumable by the OCS External Source Data team, thus enabling efficient access for the national bathymetry and the chart. This paper illustrates the process and describes the current results of these efforts.

The Simrad EK60 is a scientific split beam echosounder specifically intended for observing the water column with multiple frequencies. A split beam echosounder is a single beam echosounder, which is to say an acoustic transducer with a beam pattern that defines a

single primary beam, but is capable of determining the angle to a target within the beam. Given multiple frequencies, it is possible to observe and discriminate between mid-water marine organisms, such as schools of fish and plankton, or map for gas seeps and oil spills (National Centers for Environmental Information, 2011). For this study we use the EK60 as a single beam echosounder. While the majority of hydrographic surveys use a swath sonar for more efficient and complete bathymetric coverage, single beam sonars are still capable of recording bathymetry while surveying the water column (Figure 1). The majority of the data processed includes information from five transducers with discrete frequencies: 18kHz, 38 kHz, 70 kHz, 120 kHz, and 200 kHz. These discrete frequencies allow for duplicate observations of the seafloor, potentially improving the information available to a seafloor detection algorithm. The EK60 data has largely been collected in the last 20 years, making it particularly useful for updating bathymetric charts in areas where the latest survey may have been over 100 years ago. There are currently 401 cruises from 2003 to 2019 with available EK60 data, spanning a total of 1.1 million linear nautical miles of data. Of these EK60 cruises, NMFS vessels have collected 222 cruises for 811 thousand linear miles (National Centers for Environmental Information, 2011). The cumulative geographical scope of these surveys makes them particularly useful for hydrographic mapping (Figure 2).

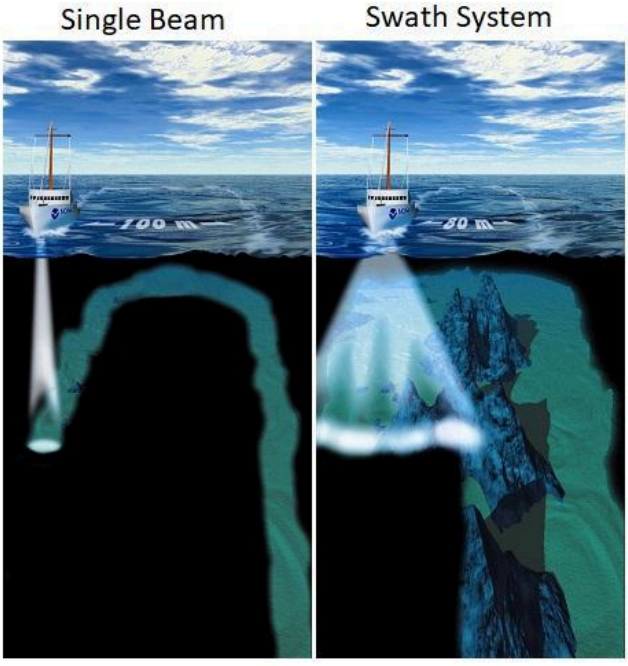

Figure 1. Single beam echo sounder compared to swath echo sounder. Both methods record the bathymetry of the seafloor. The EK60 echosounder is represented by the single beam.

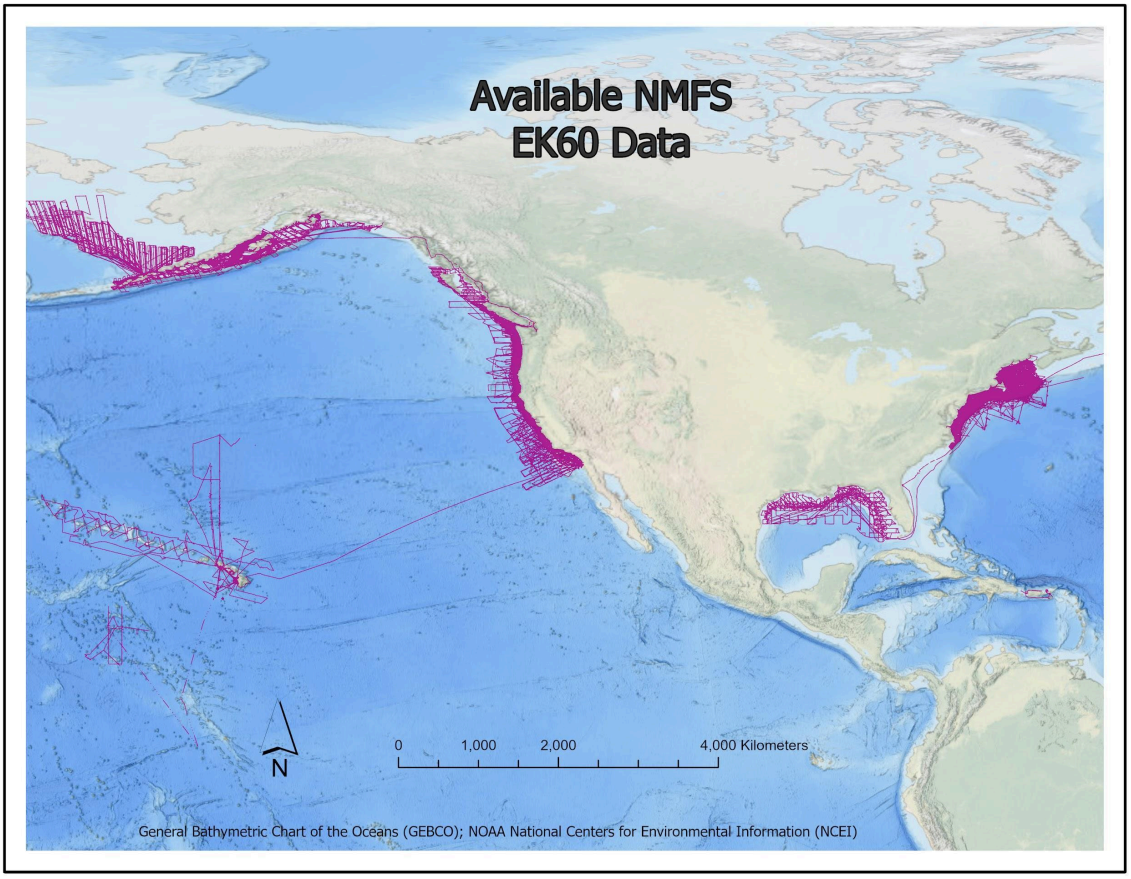

Figure 2. Available NMFS EK60 data for the entire US Exclusive Economic Zone from 2003-2019 (National Centers for Environmental Information, 2011).

As demonstrated in our test area of the Northeast United States, the expanse of the EK60 surveys has the potential to contribute to unmapped areas of the US EEZ (Figure 3).

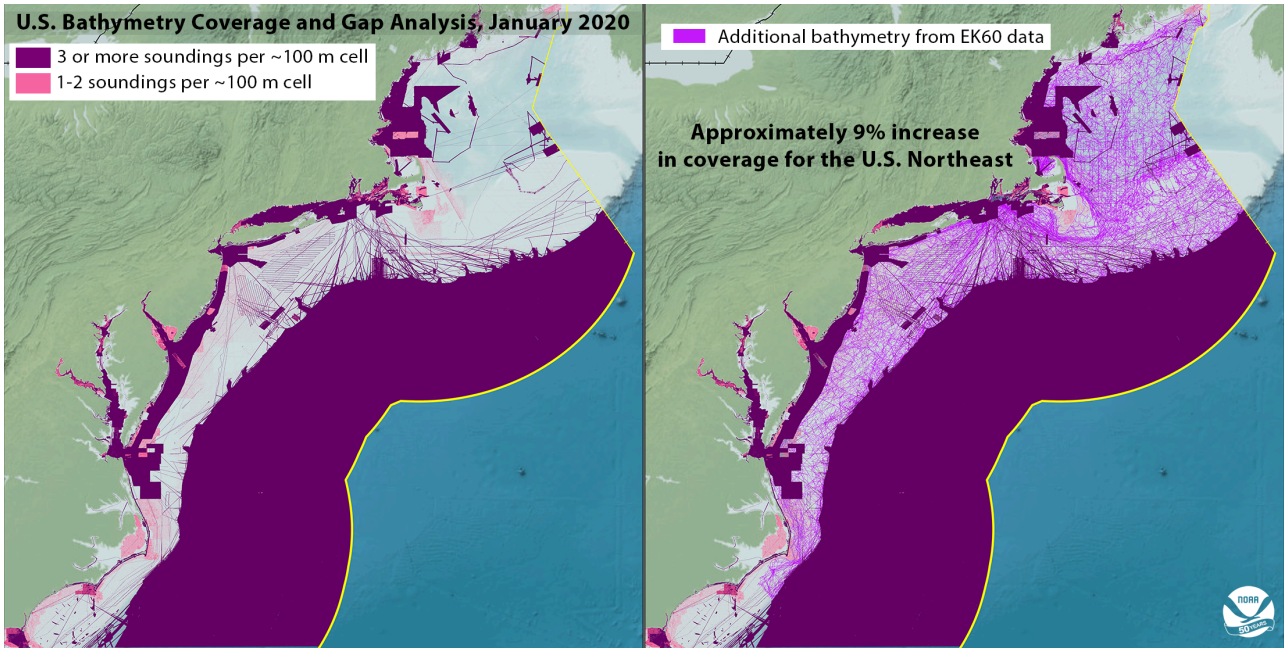

Figure 3. Bathymetric data coverage for the Northeast US EEZ (left) and additional NMFS EK60 data coverage for the same region (right). Notice how the EK60 data covers a sizable unmapped area of the United States EEZ. Figure credit: Meredith Westington, NOAA Office of Coast Survey.

Based on the January 2020 analysis of US bathymetry coverage and gaps, as referenced in the Progress Report of Unmapped US Waters (Office of Coast Survey, 2020), Westington estimates that the EK60 data could increase bathymetric data coverage in the Northeast US EEZ by nine percent, bringing the coverage of the area from 65% to 74% minimally mapped per Seabed 2030 goals (Westington, 2020). The mapped statistic is based on the standard that a mapped area constitutes a density of at least one soundings per 100m grid cell after 1960 (Westington et al., 2019). This large reservoir of single-beam echosounder data has the potential to significantly boost US bathymetric records, especially in light of a recent analysis that estimates it would take around 177 years for a single survey platform running continuously at 7.5 knots to survey the rest of US waters to meet modern survey standards (Greenaway, Batts, and Riley, 2020). Furthermore, a majority of the Northeast US covered by the NOAA Fisheries EK60 data is within depths of 0-200m. This is considered a “high effort” zone because multibeam survey swaths are narrower at these depths than in deeper zones and would require more passes to collect complete coverage.

We implement the following steps to extract bathymetry on chart datum from the EK60 water column: seafloor detection, sound speed application, and water level, heave, and draft adjustments. The Python scripts described in this article are currently housed in Pydro, a public, Microsoft Windows-based software suite developed by HSTB. This suite contains tools, libraries, and information that support the processing and analysis of hydrographic data that aid in refining seafloor detections into depths (Gallagher et al., 2020). Since the main script exists in

the Pydro environment, it is able to access tools for estimating harmonic sound speed from historical data in the World Ocean Atlas (WOA) database (Boyer et al., 2013), and tide information from the Archiving, Validation, and Interpretation of Satellite Oceanographic (AVISO) auxiliary product: Finite Element Solution (FES) global tide model, 2014 (AVISO+ FES2014b) (NOVELTIS et al., 2014).

Ultimately, the purpose of the described workflow is to enable the automated extraction and adjustment of acoustic data from the EK60 surveys to derive useful bathymetry. As mentioned previously, the geographic expanse of the EK60 surveys makes them valuable for filling current gaps in hydrographic maps. This data also has the potential to contribute to global mapping projects such as Seabed 2030, which aims to map the global ocean by 2030 (Westington et al., 2019). This style of workflow may also have implications with other projects involving large datasets. Additionally, this project highlights the usefulness of the AVISO tide model, which can be used to process other bathymetric datasets and promote hydrographic studies in new areas.

Methods

The overall workflow for obtaining and processing EK60 data is broken into several stages. First, EK60 RAW files are queried and downloaded from the NCEI public cloud (National Centers for Environmental Information, 2017). Once the files are local to the processing resources, our code extracts the time, navigation, and water column records from the EK60 RAW files. The seafloor is then extracted from the water column data, and the time and navigation are used to query tides and sound speed to create a best estimate of depth. Seafloor extraction includes identifying prevalent features within the water column, identifying the most likely candidates, and then calculating a two-way travel time (TWTT) for the selected features. A low pass filter is applied to the seafloor to remove heave artifacts, and draft is applied as recorded in the RAW file. After processing all survey files are completed, the data are gridded at eight meter resolution for comparison to existing data. A resolution of eight meters was chosen to match the resolution of existing data at the ENC Band 3 region.

Seafloor Detections

Seafloor detections are based on image processing techniques and the assumption that the seafloor is the largest and strongest target in the water column data. This approach first extracts the water column amplitude time series for each file into a two-dimensional array with dimensions corresponding to the number of samples and the number of pings for each frequency. For each frequency, a Scipy (Virtanen et al., 2020) image edge detection method performs an initial feature detection by applying an eight standard deviation Gaussian filter to remove noise and a Sobel filter in the along-ping (vertical) direction to find significant positive gradients. We defined significant gradients as greater than one standard deviation of the Sobel gradients found

within the file, and we selected positive gradients since the seafloor should increase in amplitude from the background noise. A region of interest for the feature is then extended to include the samples after the significant positive gradient to include when the gradient reaches a minimum (maximum negative) gradient since this corresponds to a decrease in power amplitude and the backside of the feature (Figure 4).

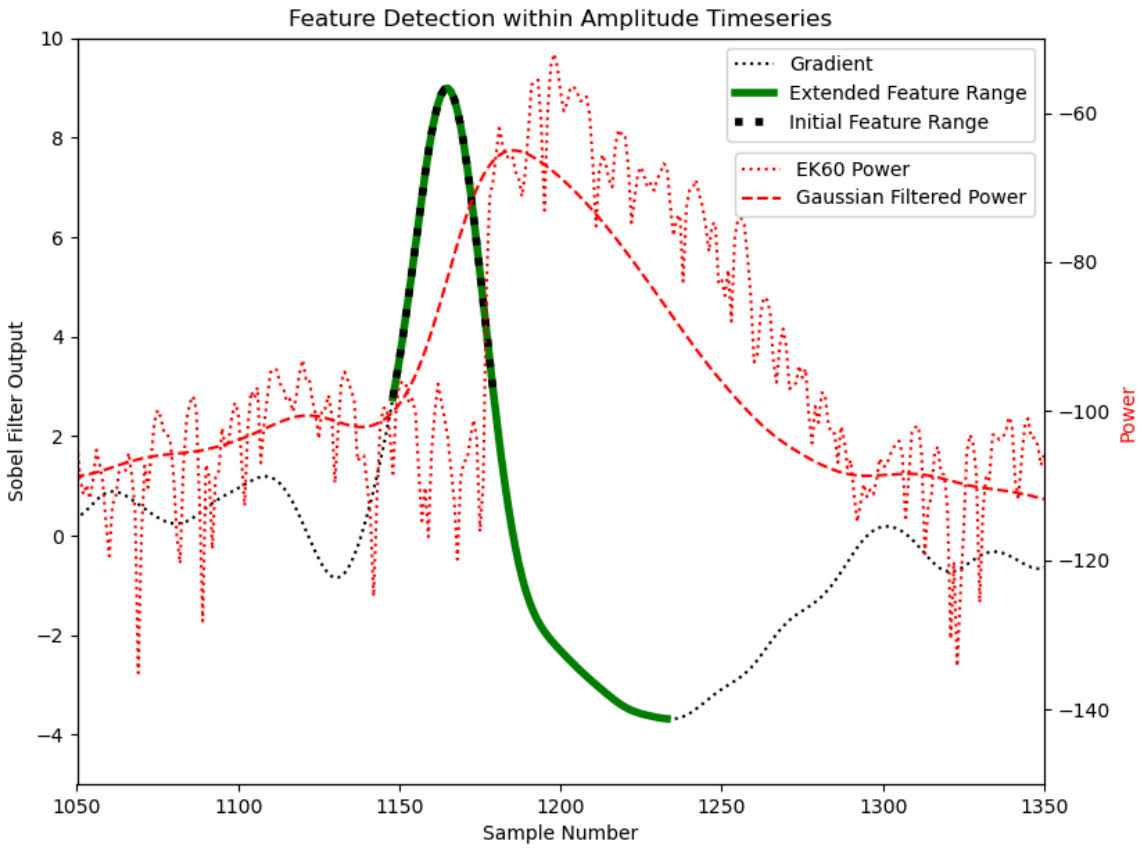

Figure 4. A subsection of the Sobel gradient and power time series in the area of a detected feature. The raw power and Gaussian filtered power illustrate change relative to the Sobel filter gradient. We demonstrate the detected feature as extended to the minimum gradient, and how this range of samples corresponds to the maximum power amplitude.

These regions of interest are then grouped by classifying across pings using OpenCV (cv2 as a python library) (Heinisuo, 2020) to create features with metrics. Each feature within the file is tracked with a start and stop ping index, the index of maximum amplitude for each ping within the feature, the average signal to noise computed between a blanking distance and the feature, and the sum of the maximum amplitudes within the features. Features are then sorted by descending sum of the maximum amplitudes for the separate features, and the maximum non-overlapping (in ping index) features are selected. Features with a signal to noise ratio less

than 30 dB are discarded. This series of steps is repeated for each transducer/frequency recorded in the RAW file.

Future iterations on detecting the seafloor will likely change how features and seafloor detections are defined. The current method begins with the feature designation at one standard deviation of the amplitude gradient above the average for the file, but the feature may preferably begin at the maximum gradient so as to better exclude water column objects near the seafloor. Also, the end of the feature is probably better defined as when the negative gradient returns to background levels, and the actual seafloor defined as a weighted average of the values within the feature. This would help smooth seafloor detections, and the actual feature range definition would have a negligible impact as the average would be computed in the linear, rather than a logarithmic, domain. Also, rather than the arbitrary 30 dB signal to noise ratio filter for a weak target currently in place, the target strength for features should be estimated and compared to established seafloor target strengths (Weber, 2020). Occasionally, areas with steep slopes appear to have overlapping features, and perhaps this can be accommodated as well.

After the TWTTs are calculated for each transducer, the data is combined into a single numpy array (van der Walt et al., 2011) indexed by time values with missing data removed. The lists are iterated through to consolidate unique time values and remove records with no values for any transducer. Currently, the lowest frequency is considered the most stable detection, but the highest frequency is assumed to have the highest fidelity on the water/seafloor interface, although also prone to misdetection on water column targets. We start by assuming the lowest frequency is correct and then progressively selecting the next highest frequency that is within two pulse lengths of the currently selected detection.

Sound Speed

Sound speed is required to convert TWTT to a distance. A common approximation of sound speed is 1500 m/s, although ocean sound speed can vary by roughly 3% from this value. The use of representative profiles derived from historical data can be used to determine a better estimate of the harmonic sound speed for a particular location and season, although the residual error induced by this estimate is a function of the variability in sound speed for the location. The sound speed is retrieved from the locally downloaded WOA13 database (Boyer et al., 2013) using the latitude, longitude, POSIX timestamp, and a rough depth measurement calculated from the low frequency TWTT (Figure 5). Because the variability in sound speed between records is usually very small, a standard harmonic sound speed value for every one hour is calculated and applied to all records within that time frame. That returned sound speed value is then used to calculate the range to the seafloor by multiplying each transducer's TWTT by the sound speed and then dividing by two. While in situ sound speed measurements are ideal, they are not readily available in the raw EK60 files. Using the modeled harmonic WOA sound speed estimates is more accurate than using a static approximation of 1500 m/s, and introduces potential error that is less than one percent of depth based on consultation with SmartMap (HydrOffice, 2020).

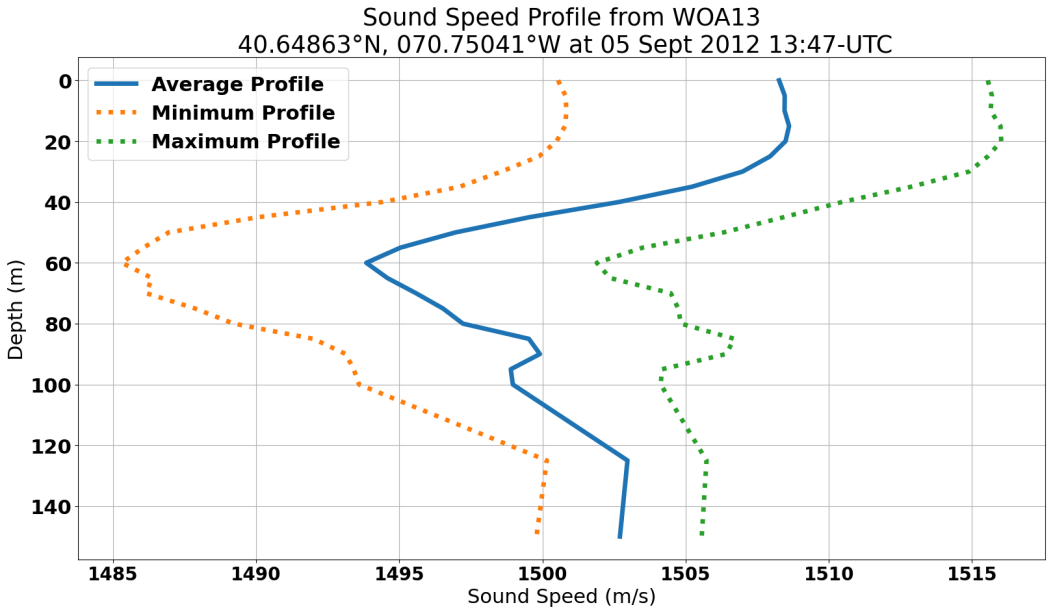

Figure 5. An example of an average sound speed profile from WOA (Boyer et al., 2013). These profiles were used to calculate a harmonic sound speed used to correct the EK60 two-way travel time.

Heave

The heave of a ship can significantly influence depth measurements, so we apply a low pass filter designed to determine heave values and remove high-value artifacts from the seafloor detections. A fifth-order Scipy Butterworth filter is applied with a ten second cutoff by resampling the seafloor at one Hz and applying the filter. Heave is then estimated by differencing the original and bandpass data and then resampling the heave to the original timestamps. This provides heave as a separate, though dependent, variable alongside the original detections such that it can be removed and replaced with a better solution if desired. The approach of applying a lowpass filter was inherited from the previous Saildrone work, although that approach did not attempt to separate the heave value. Further work for using a highpass filtered heave from the GPS GGA string will likely be explored in the future.

Tides

Correction for tides requires two separate but related components: realization of the tidal datum and a model for time-dependent water levels. We achieve both through AVISO+ FES2014b database, which includes 34 tidal elevation components defined on a 1/16° grid. The FES2014b hydrodynamic model includes assimilation of long time series satellite altimetry data from Topex/Poseidon, Jason-1 & -2, TPN-J1N, ERS-1 & 02, and Envisat, as well as the assimilation of global tidal gauge data. The FES software distribution contains a C/Python API

that may be used to calculate tidal height time series for any particular point within the global grid. FES2014 was produced by NOVELTIS, LEGOS, CLS Space Oceanography Division, and CNES. It is distributed by AVISO, with support from CNES (NOVELTIS et al., 2014).

Tide corrections in situ to the vessel tracks are computed to a mean lower low water (MLLW) datum to adjust the EK60 observed depths for charting. Interpolated tide predictions were first pre-processed to compute datums on a 1/60° grid, in line with the sampling resolution of the NOAA gravimetric xGEOID products (National Geodetic Survey, 2019). Tidal datums representative of the NOAA 19-year National Tidal Datum Epoch were resolved by averaging select semi-diurnal extrema (higher high waters, high waters, low lower waters), spanning one year of FES2014b computations at each grid node. For each EK60 sonar transmission latitude, longitude and timestamp, FES2014b-predicted tides relative to the local mean sea level (LMSL) are computed. For navigational safety, the LMSL tides are adjusted to MLLW by spatially interpolating separation values from the 1/60° datum grid. The resulting MLLW tide heights are then used to adjust the observed depths obtained from the EK60 seafloor detection processing. In addition to MLLW-LMSL, the datum grid includes the aforementioned mean high water datums and geoid height, as well as (local) mean sea surface ("ellipsoid") height, so that the EK60 data may be referenced to the NAD83 or WGS84 reference frames.

To validate AVISO+ tide predictions, comparisons were made to modeled water levels computed from NOAA CO-OPS Discrete Tide Zones at 67 locations offshore of the Northeast US coast. Differences were computed at the centroid of each tide zone, where verified water level observations from NOAA tide stations were processed using the publicly available model parameters (amplitude and time offsets) attributed to each zone. The comparison time period spanned five days from May 1st to May 5th, 2020, using six-minute tide data available through the CO-OPS tide data API (Center for Operational Oceanographic Products and Services, 2020). We subtracted the NOAA zoned tides from the AVISO+ predictions and found the average difference at the 67 locations ranged from 7cm to 51cm, with standard deviations ranging from 6cm to 17cm (Figure 6).

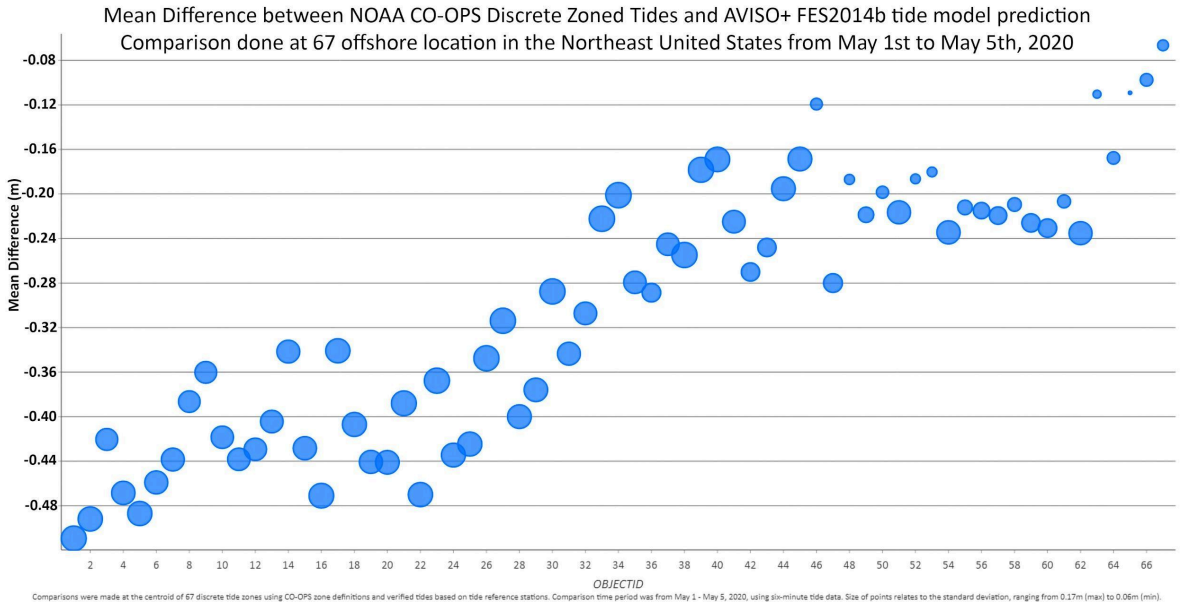

Figure 6. Mean difference between AVISO+ FES2014b tide predictions and NOAA CO-OPS discrete zoned tide observations at 67 offshore locations.

We observed some correlations between the mean difference and geographic location, especially east of Cape Cod, which raises questions concerning this approach to estimating survey time tides (Figure 7). The measured differences were mainly a consequence of variations in tidal magnitude rather than tidal phase (e.g., Figure 8). While meteorological effects could explain some varying geographic biases given the short time frame of the comparison, it is important to note that these comparisons are on datum. Both models (NOAA zoned tides and AVISO+ predictions) include a potential bias in resolving the tidal datum definition relative to the mean sea level away from water level observation points. The AVISO+ FES2014 model includes assimilation of tide gauge observations as well as long-term altimetry measurements of the ocean surface. The amplitude of the aviso model may be superior in terms of a reduced datum bias in the offshore region. More effort to compare these on-datum water level models for a larger range of times and locations would be a valuable contribution to both this project and other projects seeking a solution for a first order tidal height estimate.

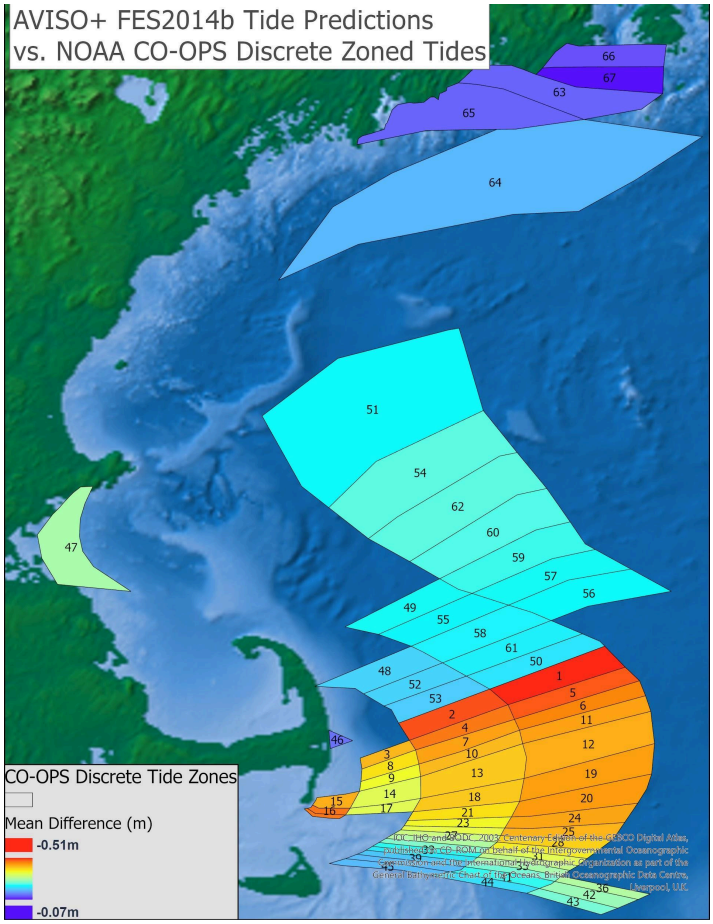

Figure 7. CO-OPS Tide Zones used for comparison to the AVISO+ FES2014b tide predictions.

Representative tide data from the region used to compare EK60 results with reference bathymetry in Rhode Island Sound is shown below (Figure 8 and Figure 9).

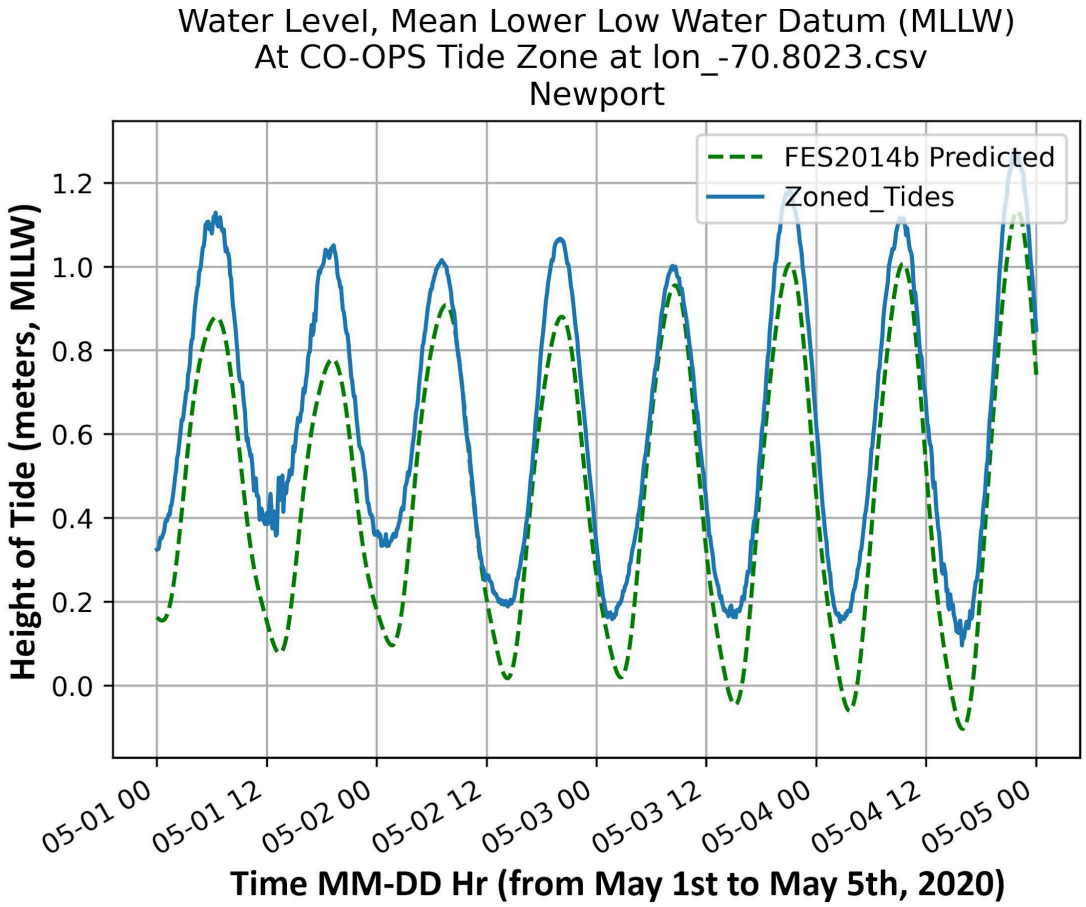

Figure 8. Comparison of AVISO+ FES2014b predicted tide values and verified tides in Rhode Island Sound.

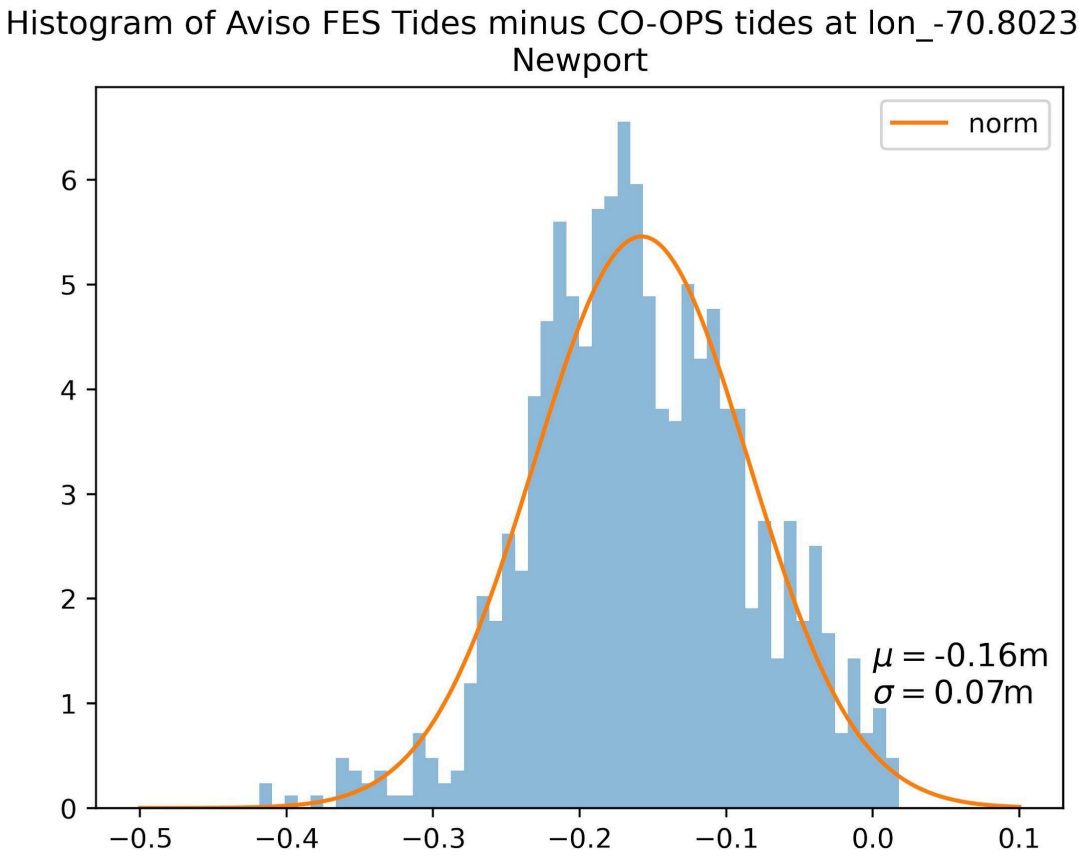

Figure 9. Histogram showing the distribution of differences between AVISO+ FES2014b predictions and NOAA CO-OPS discrete zoned verified tide-corrected values in Rhode Island Sound.

We recognize this is a restricted geographical and temporal analysis and does not represent the quality of the tide prediction everywhere or with time dependent events. Local meteorological effects are not well accounted for with this approach, but this can also be true of zoned tide models that propagate shore based measurements over large distances. This comparison demonstrates the AVISO model can be used to improve a final depth estimate to a similar order as traditional tide models and is thus a reasonable solution to apply as a first order solution for offshore fisheries data.

Transducer Draft

Another significant piece of information when deriving depths is accounting for transducer draft. Contemporary NOAA Fisheries Survey Vessels (FSVs) have the EK60 mounted on a drop keel and thus can significantly change the draft of the transducers even during a cruise. Nominally these drafts range from six meters when the drop keel is retracted to the hull and nine

meters when fully deployed. We elected to use the draft value provided in the EK60 datagram since other manually logged information was not readily available or necessarily more reliable. If the draft value in the EK60 file is found to be zero, which is assumed to be incorrect, the process defaults to a maximum draft value of nine meters because these vessels are more likely to operate with the drop keel deployed (Walsh, 2020).

The final step of the primary script is to save the georeferenced seafloor detections with meaningful processing information to enable downstream review. The output of the process includes an ascii text file with columns for time, longitude, latitude, sound speed, heave, tide, used draft, depth from each transducer, and the selected detection. Also, a log file is created for each RAW file in the directory with any errors. If the default draft value is applied because a zero draft was found in the file, it is logged here. An overview graph consisting of two subplots, one showing scatterplots for each transducer and the best detection, and the second showing the heave, sound speed, tide, and draft, are also created. This image helps visualize each file's depth measurements at each frequency and the factors that influence them. Finally, an image with the water column data with the detections for each frequency is also produced to support a quality review of the seafloor detection.

Additional scripts were used to download RAW data from the NCEI database as well as to extract navigation values from text files for analysis. One particularly useful script we wrote was gridData.py, a script written to grid point data and export rasters with a given resolution. This code grids the processed output of the seafloor detection and georeferencing script in order to streamline the comparison process to existing NOAA surveys. For each survey, all files from the previous phase of the process are concatenated into a single geodataframe and gridded. The gridded data are broken into tiled GeoTIFFs that have eight meter node spacing with separate layers for the mean, minimum, and maximum values for the detections within the cell footprint.

The resulting statistics are based on the difference between the gridded EK60 seafloor detections and the qualified bathymetry. That is to say that the bathymetry values from collocated NOAA surveys were subtracted from the EK60 bathymetric values to determine overall differences between measurements.

Results

Once the previously described methods were tested sufficiently to justify work on a larger scale, the results of the seafloor detection algorithm were analyzed for reliability and accuracy against EK60 data collected from 19 NMFS cruises. In general, these methods show promising results when compared to contemporary NOAA hydrographic surveys from the same regions.

With the successful execution of the script each RAW file is represented by multiple processed results, such as an overview plot and several time-series images (Figures 10 and 11). The following figures are examples of results from a survey conducted by NOAAS *Henry B.*

Bigelow, which provided the basis for the testing data used in this study (National Centers for Environmental Information, 2011).

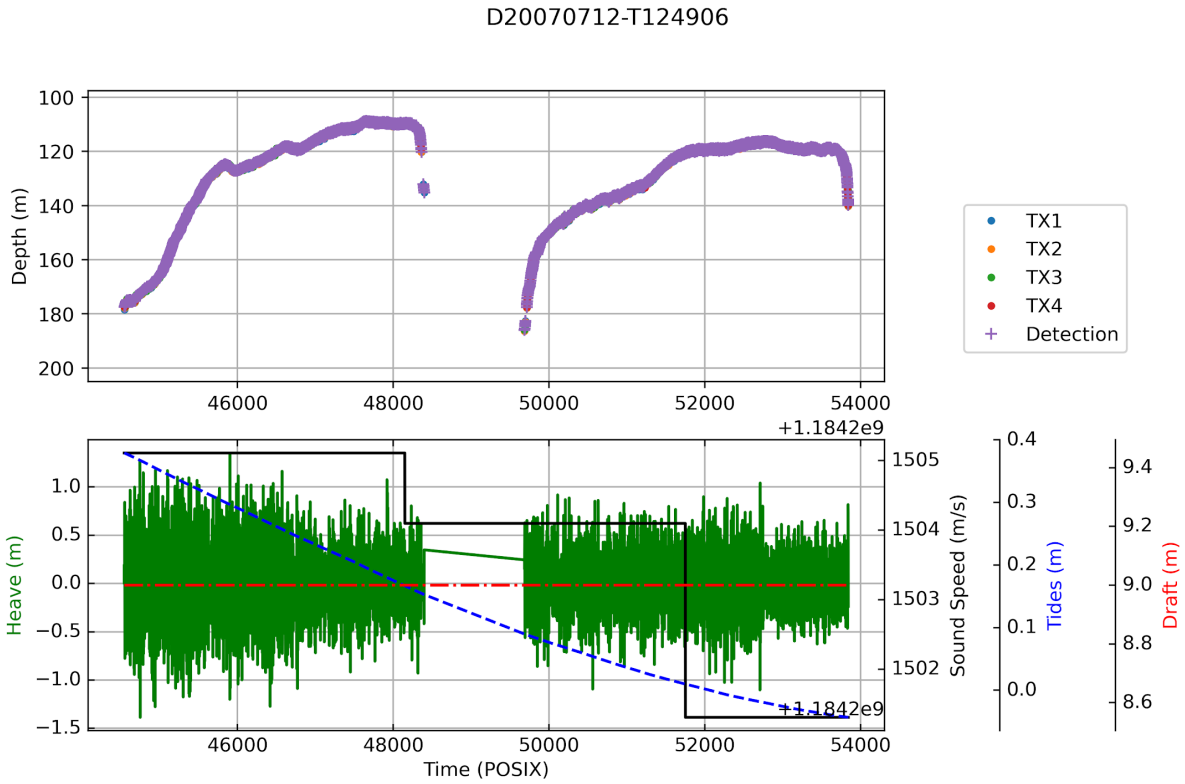

Figure 10. Example of overview plot generated by the EK60 script for RAW file D20070712-T124906 of the NOAAS *Henry B. Bigelow*. The top graph shows seafloor detections for all transducers, which are all colocated in this case, while the bottom shows values for tides, sound speed, heave, and draft.

The time-series image (Figure 11) visualizes the water column for a single RAW file. The red line indicating the seafloor shows a sudden change in depth, most likely from the survey ship passing over a steep drop off. A second return for the seafloor is also noticeable underneath the red line in the top image. Time-series such as this one showcase visible marine features of both the water column and the ocean floor to aid in the processing of bathymetric data.

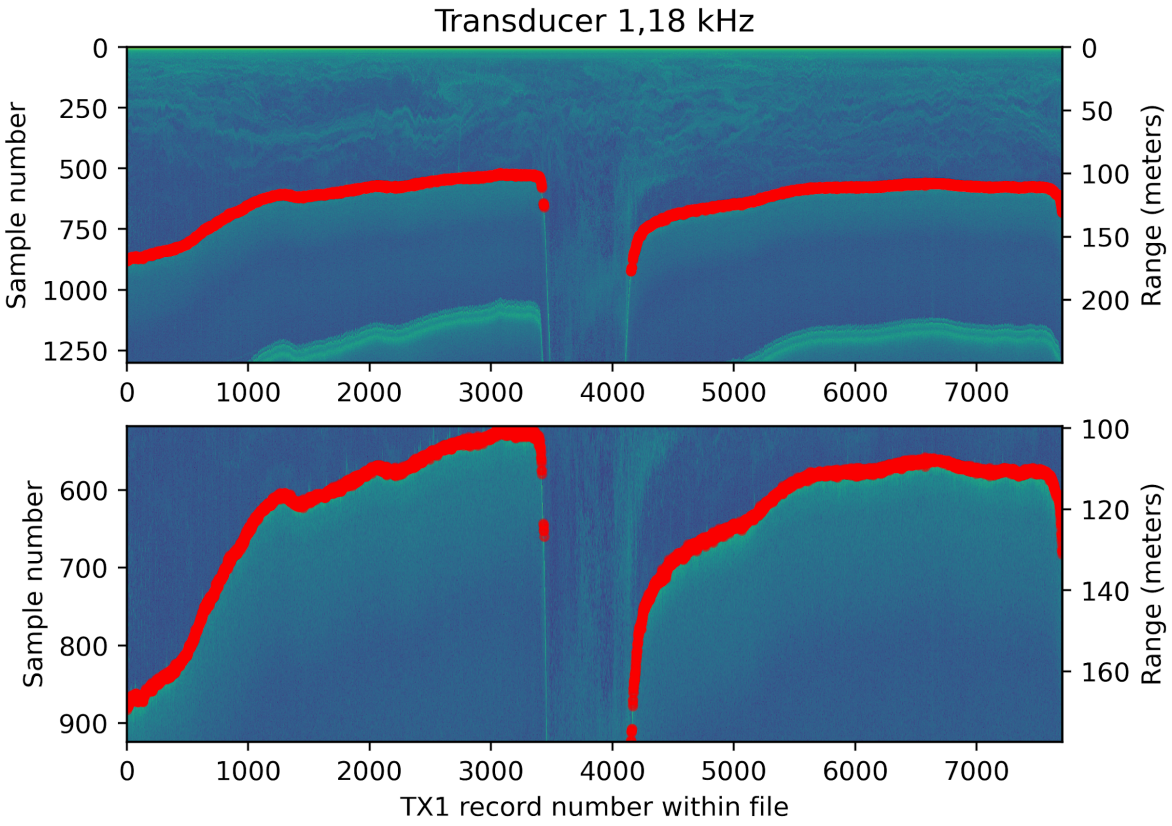

Figure 11. Example of time series image for one transducer generated by the EK60 script for RAW file D20070712-T124906 of the NOAAS *Henry B. Bigelow*. For each RAW file, an image like this one was produced for each transducer, providing a visual for the water column and seafloor at that location.

Most importantly, we needed a method for testing the accuracy of the EK60 results compared to existing bathymetry surveys. As mentioned in the methods section, a gridding script was used to test the accuracy of the processed seafloor detections. Existing multibeam surveys were downloaded from the *NOAA Bathymetry Viewer* (National Centers for Environmental Information, 2015) as a Bathymetric Attributed Grid (BAG) and compared to the final product of the derived and gridded EK60 depths. All 19 NMFS cruises were compiled and compared to a combined BAG file which held current bathymetric data (2007 - 2015) for the corresponding locations (Figure 12). The combined BAG file was compiled as an eight meter resolution raster so as to match the resulting EK60 grid resolution. The RAW data processed in this study took a cumulative 632 days (2007 - 2019) to collect by fisheries vessels over 87,134 linear nautical miles.

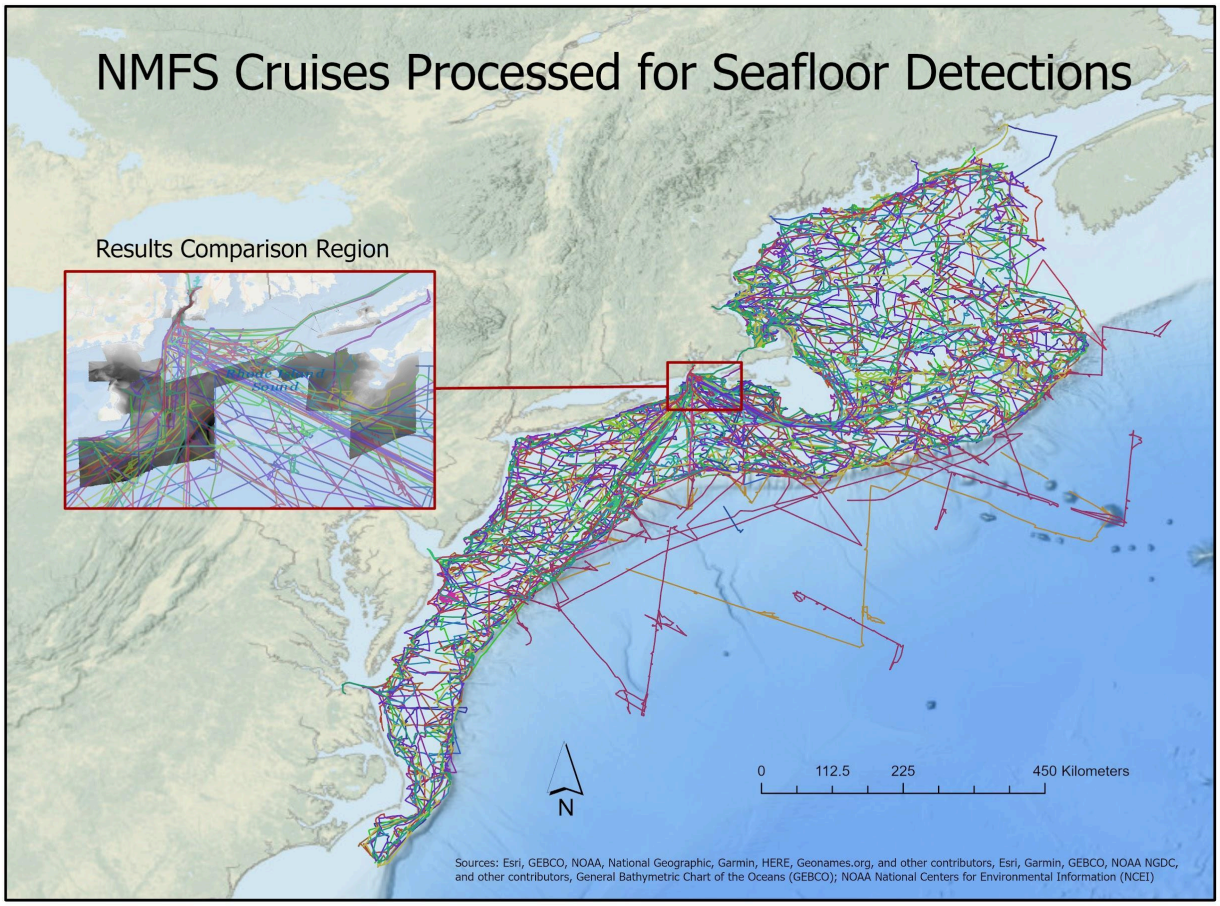

Figure 12. Cruise tracklines for the 19 NMFS cruises processed for seafloor detections and the results comparison region with survey BAG file used for analysis (National Centers for Environmental Information, 2015; National Centers for Environmental Information, 2011).

The reference BAG file survey values were subtracted from the eight meter gridded EK60 data using the raster calculator geoprocessing tool in ArcGIS Pro to create difference grids. The eight meter difference grids were then used to generate points for each grid using the raster to point conversion tool in ArcGIS Pro. Once each cruise had difference-points for each eight meter grid, the point feature classes were merged using the ArcGIS merge tool to generate collective results for all 18 of the *Henry B. Bigelow* cruises analyzed, as well as a separate result for the *Pisces 1301* cruise. The following statistics resulted from the analysis.

The EK60-generated seafloor detections from the *Henry B. Bigelow* have a mean difference of 0.68 meters deeper than the seafloor detections of reference surveys (Figure 13). After removing outliers that were greater than three standard deviations from the mean, the analysis yields a mean difference of 0.438 meters deeper than the seafloor detections of reference surveys (Figure 14). 92.2% of all EK60 results are within two meters of BAG survey values. Also, 83.9% of EK60 seafloor detections have a deeper bias than reference survey values, while 16.1% have a shallower bias. 3.5% of the EK60 results had difference values between 2.5 and

4.5 meters deeper than the BAG survey values and have a mean difference of 3.27 meters deeper than the reference survey values (Figure 15). Furthermore, 3.2% of results have differences of 4 meters or greater from reference BAG values. That being said, 82.4% of all detections were within one meter of absolute difference from the BAG survey (Figure 16).

We found that sections of cruise HB1304 account for 63.5% of the grid cells with difference values between 2.5 and 4.5 meters deeper than the BAG survey values. While the section of HB1304 values average 3.3 meters deeper than the survey values, the profile of the seafloor detections aligns closely with the profile of the reference survey seafloor detections (Figure 17). For this particular cruise section, the draft value in the EK60 files was missing and replaced by the default, centerboard-down value, of nine meters during processing. The three meter offset between seafloor detections and reference survey values is thought to indicate that the centerboard was up for this section of HB1304 with an actual draft of six meters.

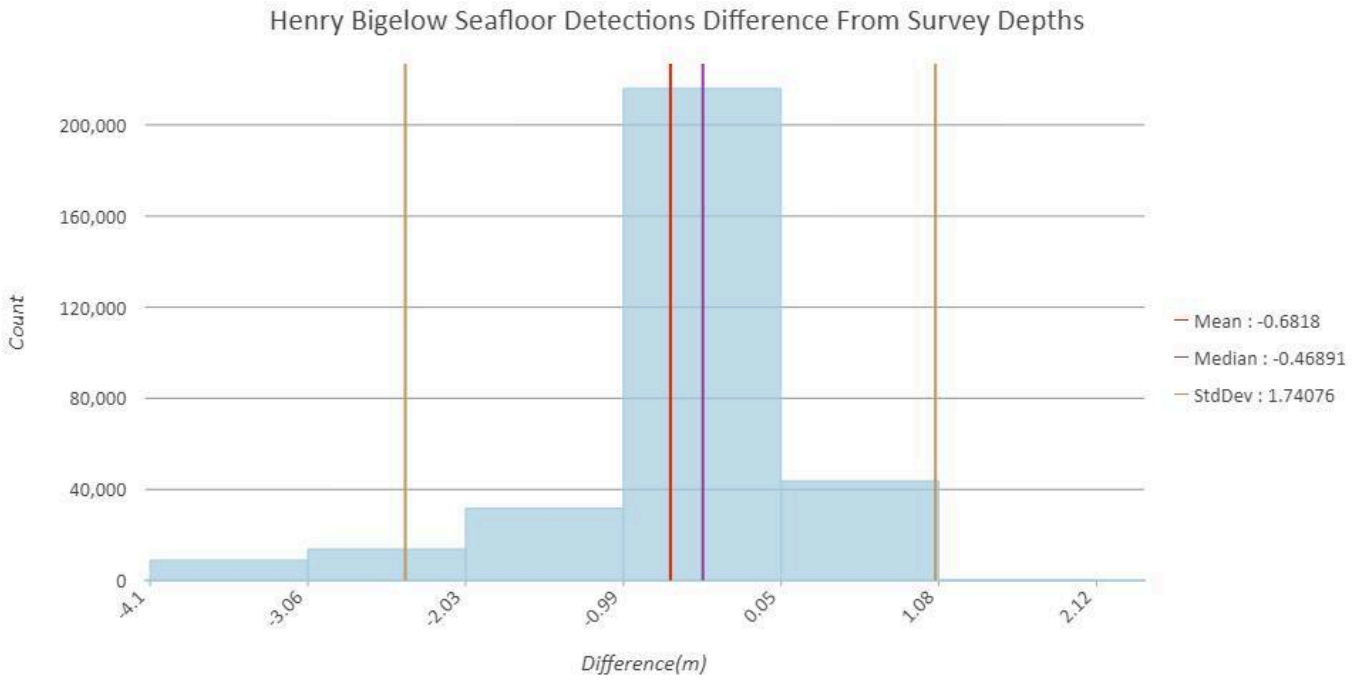

Figure 13. Distribution of difference values between *Henry B. Bigelow* seafloor detections and BAG survey depths.

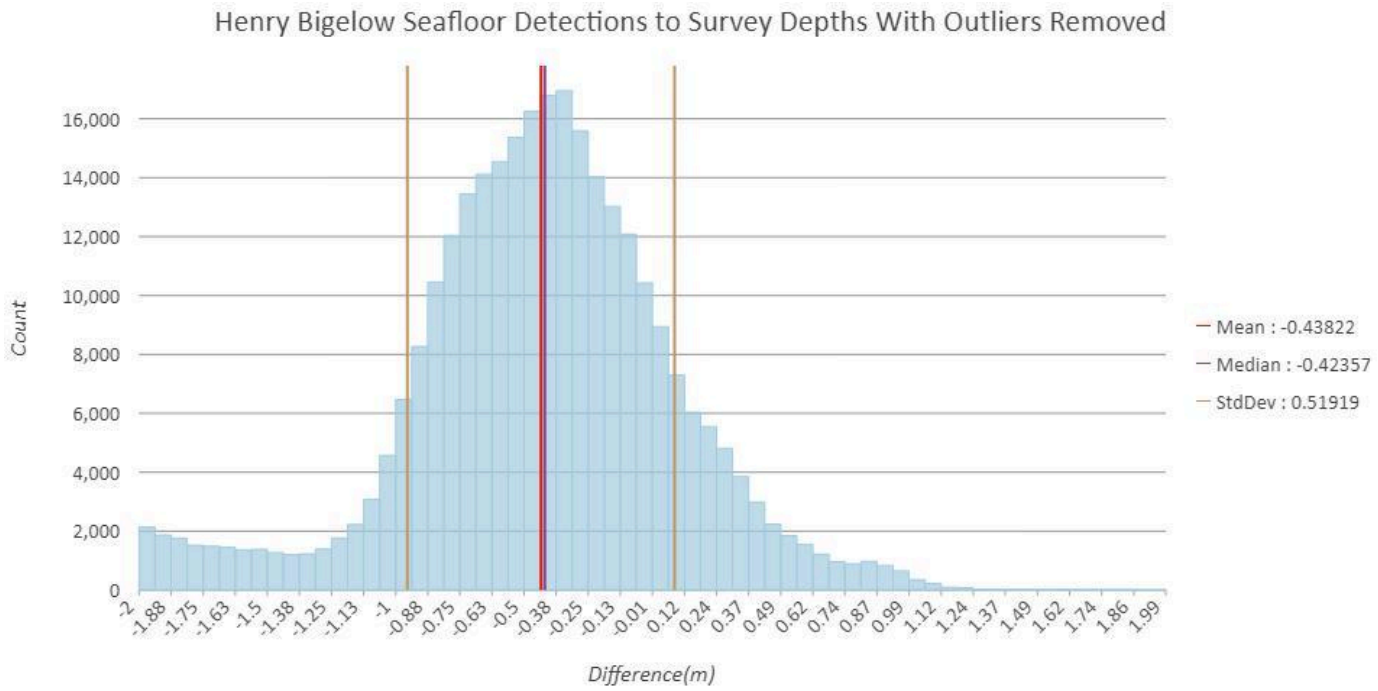

Figure 14. Distribution of difference values between *Henry B. Bigelow* seafloor detections and BAG survey depths with outliers removed (removed anything greater than two meters absolute difference).

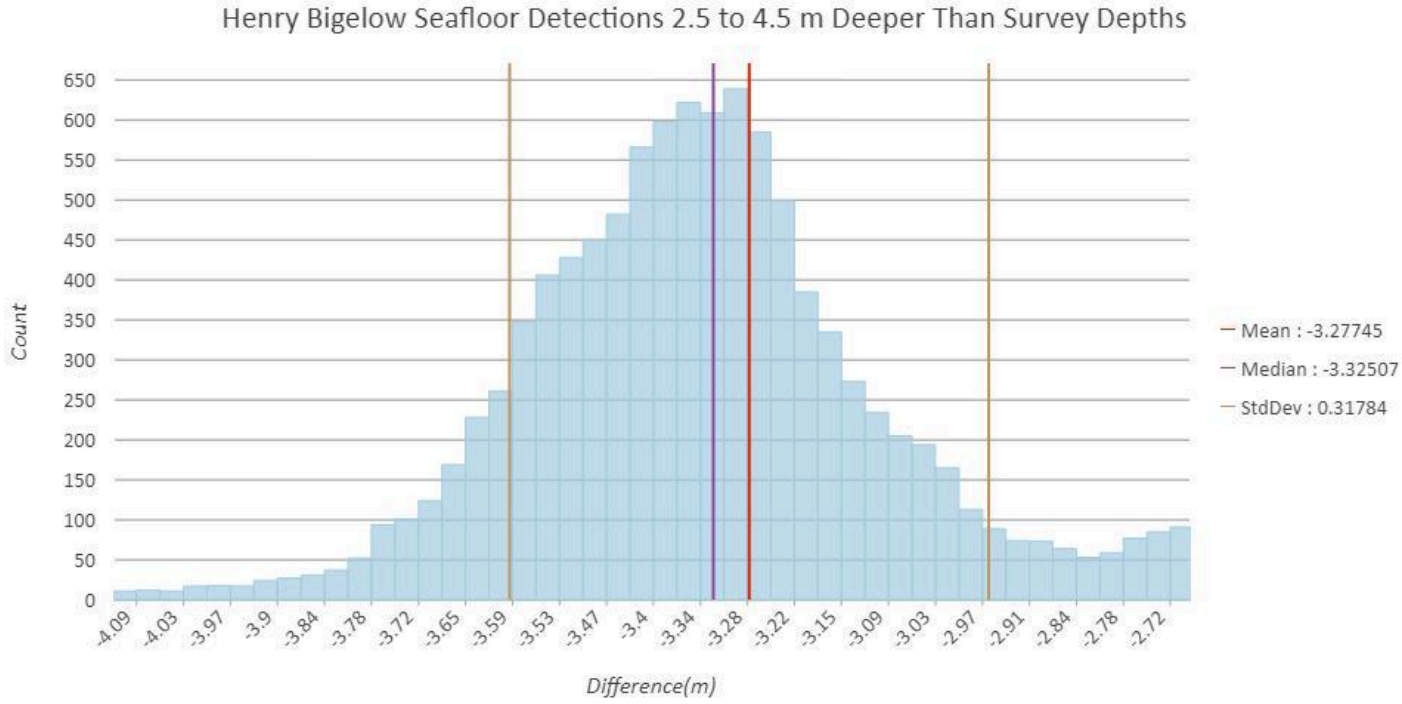

Figure 15. Distribution of *Henry B. Bigelow* seafloor detections results with difference values between 2.5 and 4.5 meters deeper than reference survey depths. The approximate three meter

bias is thought to result from FSV with six meter draft, centerboard-retracted surveys, with missing draft values.

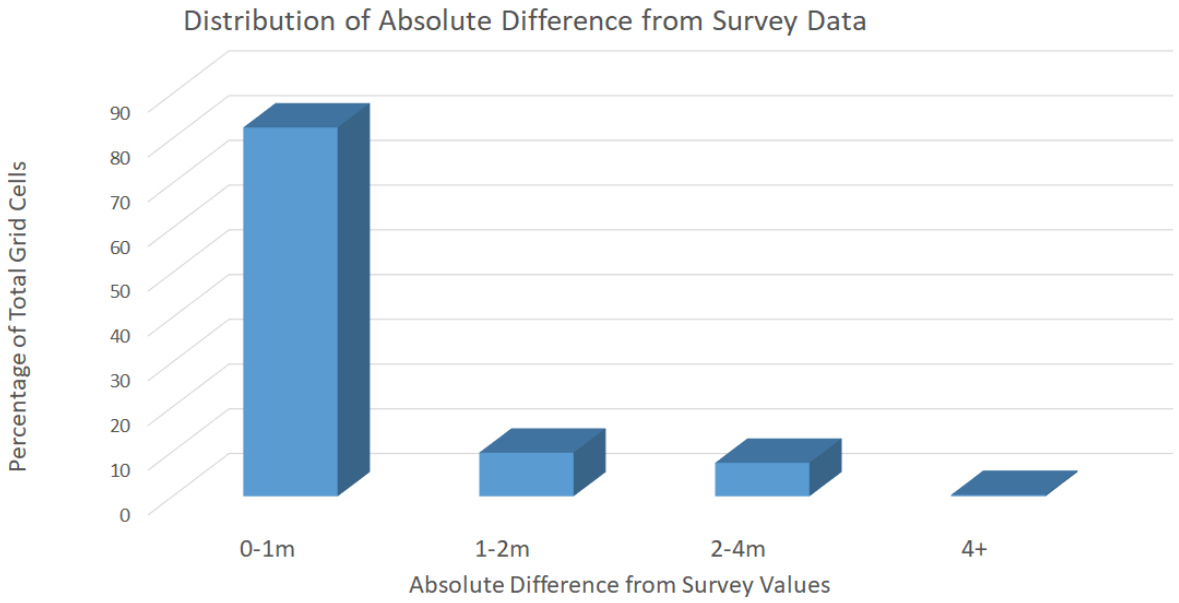

Figure 16. Distribution of absolute difference between seafloor detections and BAG survey depths.

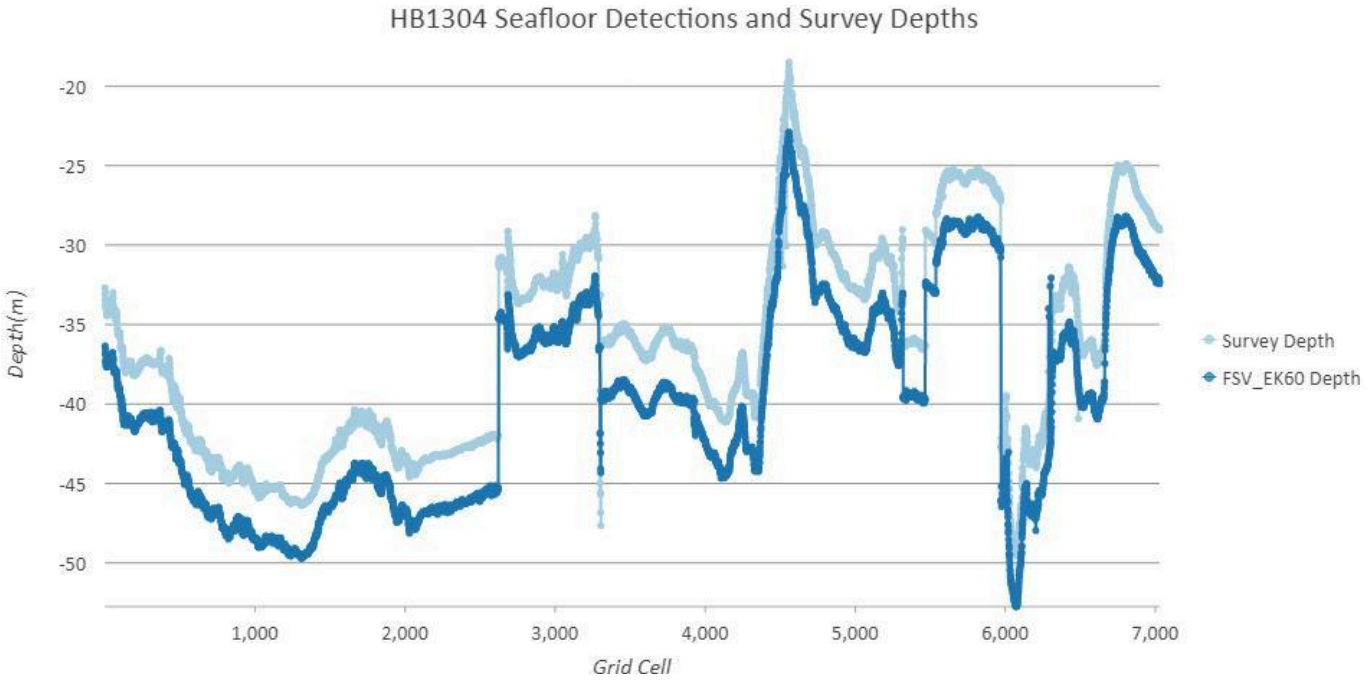

Figure 17. HB1304 grid cells with a near-constant difference between 2.5-3.5 meters, indicative of an erroneous draft corrector value.

In addition to the results mentioned previously, the EK60 script was tested on separate surveys from NOAAS *Pisces* and NOAAS *Nancy Foster*. The resulting files indicate that the script can be transferable to other ship surveys, although some modifications for different NMEA datagrams and default draft values may be required.

The regions of the *Pisces* 1301 cruise that overlapped the reference BAG file had a mean difference of 0.62 meters deeper than BAG file values, and 85.2% of the results were within two meters of the reference survey values (Figure 18). With all outliers greater than two meters difference removed, the PC1301 cruise had a mean difference of 0.59 meters deeper than BAG file values (Figure 19). Like the *Bigelow* cruise data, the *Pisces* 1301 cruise had an increase in grid regions with difference values between 2.5 to 4.5 meters deeper than survey values. We found that these were also likely the result of missing draft data, which defaults to a nine meter value when processing, assuming that the centerboard is down. Similar to the cause of the three meter deeper bias in the *Henry B. Bigelow* results, the centerboard on the *Pisces* may have actually been in a retracted position at this point, with a draft of six meters, creating the approximate three meter discrepancy. However, unlike the *Henry B. Bigelow* processed data, the *Pisces* 1301 also had an increase in grid regions with a three meter shallower bias from the reference bathymetry, which is indicative of the centerboard being extended, but the draft value in the RAW file was erroneously recorded as retracted.

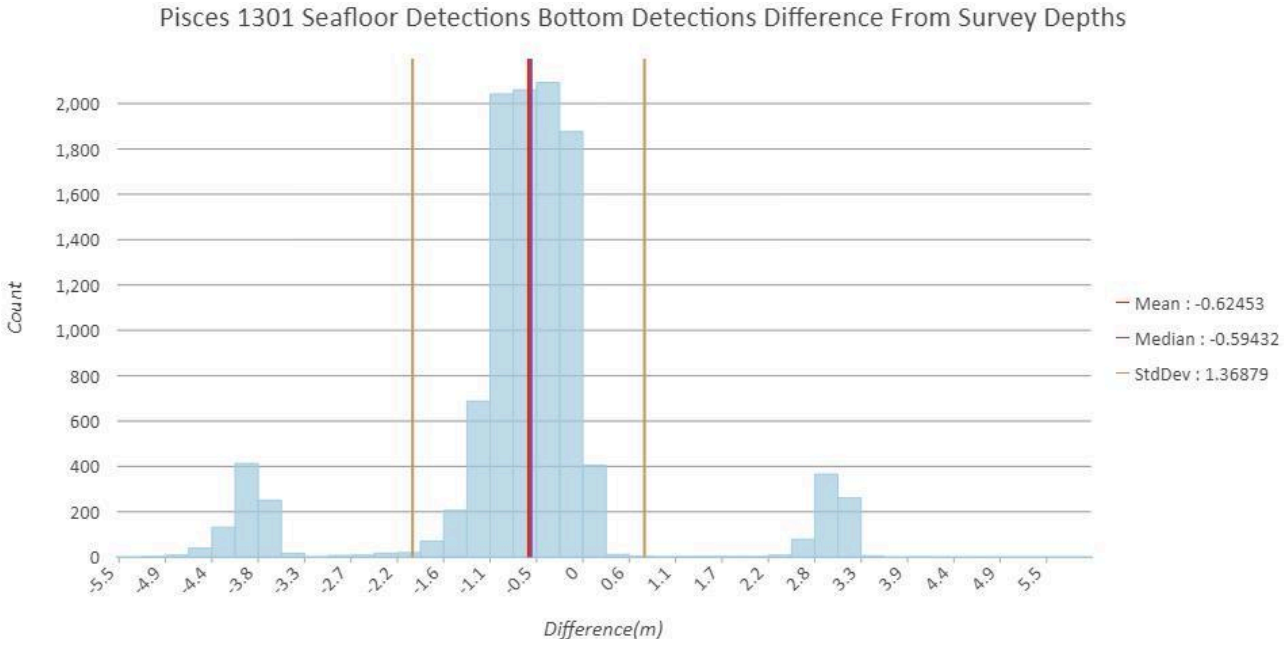

Figure 18. Distribution of Pisces 1301 difference values between seafloor detections values and BAG survey depths. The two aggregations around -3 meters and 3 meters difference are likely the result of missing or inaccurate draft measurements.

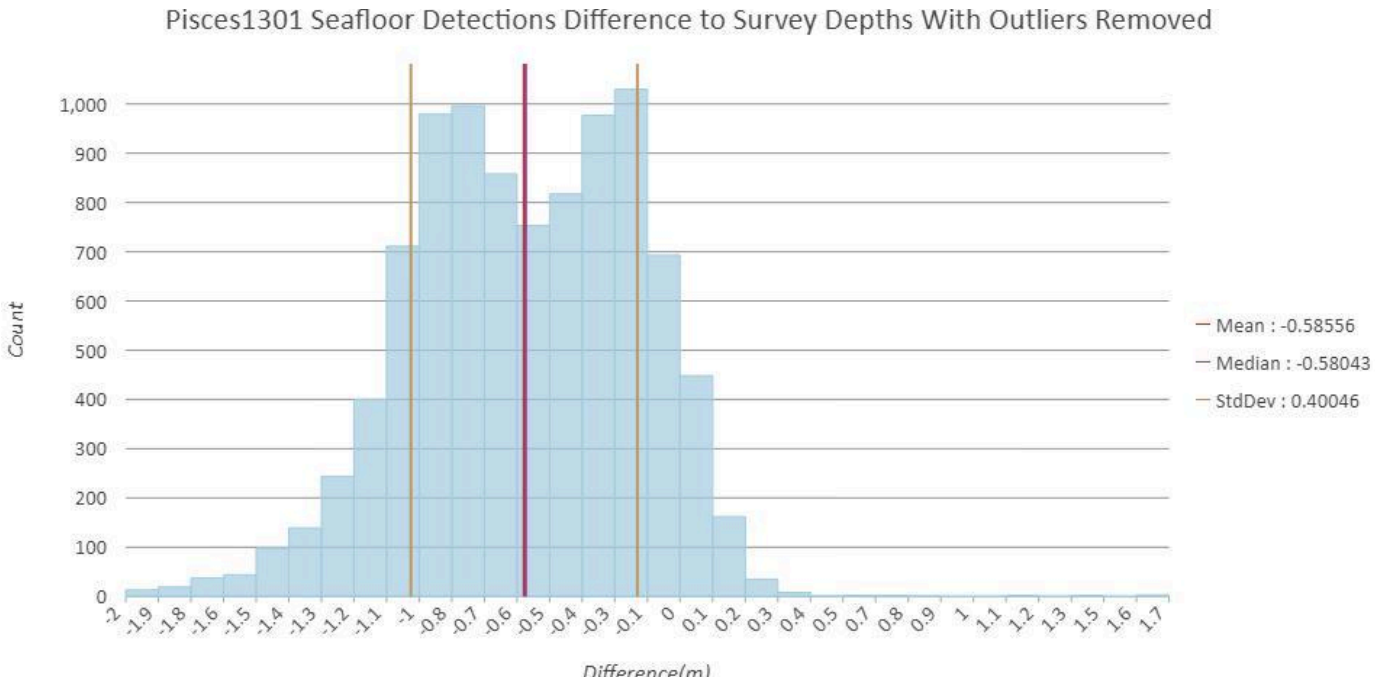

Figure 19. Distribution of Pisces1301 difference values between seafloor detections and BAG survey depths with outliers removed (removed anything greater than two meters absolute difference). The bi-Gausian distribution shown by the values in this survey is an area of further analysis we would like to explore.

Discussion

EK60 data from historical fisheries surveys are not the typical systematic and complete coverage surveys preferred by hydrographic offices, but they may be the best available data to improve bathymetric compilations in rarely surveyed regions. As mentioned previously, using EK60 for mapping purposes could increase the overall bathymetric coverage on the order of nine percent in the Northeastern United States. The average difference between the EK60 seafloor detections and the reference NOAA survey values is one meter or less (Figure 16), which is comparable to established survey and charting standards. What remains unstated is a quality metric for these data.

The various methods proposed for the reduction of EK60 water column data to bathymetry for charting bear further development to improve how they work and properly qualify their uncertainty. We discuss various justifications for why these methods are better than nothing, but significant effort remains to accomplish well qualified uncertainty estimates for each part. By way of rough and ungrounded uncertainty estimates to contextualize the resulting data relevant within typical survey metrics, we propose Table 1 for consideration.

Table 1

Large estimates for the uncertainty associated with each stage of the described processing workflow.

Source	**Estimate Type**	**Fixed (meters)**	**Depth Multiplier**
Detection	Random	2	0.01
Sound Speed	Bias	0	0.01
Tides	Bias	0.5	0
Heave	Random	1	0
Draft	Bias	0.5	0
Combined Uncertainties	$\sum_{0}^{m} Bias_m + \sqrt{\sum_{0}^{n} (Random_n)}$	3.2	0.02

The uncertainties for hydrographic surveys are sometimes described in terms of two parts describing a linear trend. The depth multiplier acts to vary the uncertainty with depth while the fixed part is not depth dependent. Adding our proposed uncertainties as a root sum square for the random components combined with estimated biases as a crude uncertainty model, we derive the fixed uncertainty to be 3.2 meters and additionally increases as two percent of depth.

In comparison to the S-44 standards set forth by the International Hydrographic Organization (IHO), the EK60 seafloor detections meet the standards for Order 2 surveys when compared to reference surveys. As stated by the IHO, these second-order surveys apply to locations where full seafloor coverage is not required and where unmarked navigational hazards are unlikely. These parameters are the least strict of all survey standards and typically only apply to surveys greater than 100 meters in depth (International Hydrographic Bureau, 2008). When evaluating the IHO S-57 zone of confidence categories (CATZOC), the EK60 data seems to fall between the parameters of B and C category surveys in terms of horizontal and vertical uncertainty, and total seafloor coverage and detection of bathymetric features. CATZOC B states that hazardous navigational features may exist but are not expected, while CATZOC C states that a full area search was not achieved and that depth anomalies may be expected. EK60 data fits squarely within the typical survey characteristics of CATZOC C, as this data is typically collected on an opportunity basis along a passage (Office of Coast Survey, 2016). Furthermore, the rough uncertainty estimates proposed in Table 1 becomes better than CATZOC C in greater than 40 meters, reinforcing the idea that these data are best considered in deep and offshore waters. Regardless, this fisheries data could substantially increase the general understanding of the seafloor beyond the boundaries of current NOAA hydrographic surveys. Additionally, it is worth noting that all EK60 data was collected during routine fisheries surveys, meaning that no extra resources were expended at sea to obtain this bathymetry.

Analyses were conducted to determine the cause of the 3.3 meter discrepancies observed between the EK60 data and the verified surveys (Figure 17). One explanation is missing draft information in the RAW file, and another is wrong draft information in the RAW file. With three possible values: six meters - when the centerboard is fully retracted and flush with the hull; nine meters - when the centerboard is fully extended; and 7.5 meters - an intermediate position of the centerboard, it would seem the intermediate position is either never used or always reported accurately since a bias of 1.5 meters is not prevalent. Our process defaults to a nine meter value when no draft information is available, so the periodic bias in either a positive or negative direction (Figure 18) would suggest that occasionally the draft value is not updated in the RAW EK60 data. A nine meter default draft value would only bias in one direction, so the other bias is likely a six meter value in the EK60 when the centerboard is fully deployed at nine meters. Unfortunately, it was clear that the default draft value is not accurate in all cases. For example, in sections of HB1304, a near-constant 3.3 meter deep bias was observed when comparing it to the reference BAG data (Figure 17), and our process reported using the default value. Given the differences between the EK60 survey HB1304 and the corresponding BAG survey, it is suggested that certain attributes such as draft be diligently recorded in EK60 files. We recommend that fisheries vessels collecting EK60 data have some automated way of recording draft over time in a machine-reading manner so that the values can be accurately maintained when performing seafloor detections.

Along with the comparisons and statistics mentioned previously, it is important to acknowledge uncertainties in the corrections used for seafloor detections in the described workflow. Further refinement to this algorithm should include uncertainty calculations for tides, harmonic sound speed, heave, draft, and other vertical and horizontal components. The largest single source of vertical uncertainty is the estimation of draft since the position of the centerboard on NOAA FSVs is not always accurately and timely recorded in a machine-readable format within the RAW file. Estimating the vertical uncertainty component from the AVISO+ tide model is an area of an ongoing study, as there is interest in applying the tide correction model to other sources of bathymetric data in US waters. Since frequent sound speed casts are not available during the fisheries cruises, this project estimated the likely sound speed profile for the survey locations and times based on past measurements available in the WOA. This estimate introduces some error, likely less than one percent of depth, but would be captured in any comparisons with reference datasets. Uncertainties for ship heave corrections have not been explored but should be included during the refinement of this algorithm.

Conclusion (Applications)

In conclusion, although these surveys are not systematic in terms of coverage, the EK60 bathymetric datasets can contribute to the overall mapped area statistic for the US EEZ. Based on the results of this paper, this data has the potential to be classified between IHO CATZOC B and C (Office of Coast Survey, 2016). The recent timeline of the EK60 surveys may offer some

updated bathymetric information in terms of accuracy and detail. This wealth of bathymetric data can be used to increase overall bathymetric coverage by up to nine percent (Figure 3) and aid initiatives such as Seabed 2030 (Westington et al., 2019). A key development in automating the processing of RAW EK60 data for multiple datasets was incorporating a global tide model (AVISO+ FES2014b) into the EK60 script. The AVISO+ tide model (NOVELTIS et al., 2014) has potential for many other applications such as crowdsourced bathymetry (CSB) projects. By automating the process of converting RAW EK60 readings into a digital elevation model (DEM) raster format readily ingested by the national bathymetry pipeline, the manual effort to use these data is significantly reduced. The workflow is designed to easily accommodate unique EK60 datasets and desired export parameters. The versatility of Python allows users to easily adjust the script to match survey specifications, making it possible for this workflow to be utilized in different projects and potentially improving other archived datasets that could benefit and improve our global bathymetric understanding. These specifications may include but are not limited to ship draft and tidal calculations. With further development, this script can iterate through downloading and processing multiple surveys worth of fisheries bathymetry data making it a useful tool for deriving information about the ocean floor from existing but previously untapped bathymetric data.

Acknowledgments

We wish to express gratitude for the thoughts and contributions of our colleagues Charles Anderson, Zach Burnett, Megan Greenaway, CDR Samuel Greenaway, LCDR Damian Manda, Meredith Westington, Sarah Wolfskehl, Katrina Wyllie, and the efforts of NMFS Survey for collecting this data and NCEI for making the data readily available on the cloud.

References

Boyer, T., Antonov, J., Baranova, O., Coleman, C., Garcia, H., Grodsky, A., Johnson, D., Locarnini, R., Mishonov, A., O'Brien, T., Paver, C., Reagan, J., Seidov, D., Smolyar, I. and Zweng, M. (2013). *World Ocean Database 2013*, NOAA Atlas NESDIS 72, Viewed 15 July 2020, Available at: http://doi.org/10.7289/V5NZ85MT

Center for Operational Oceanographic Products and Services. (2020). *CO-OPS Data Retrieval API*. Viewed 20 July 2020, Available at: https://tidesandcurrents.noaa.gov/api/

Gallagher, B., Riley, J., Zhang, C. and Younkin, E. (2020). *Introduction To Pydro - Pydro Documentation*. Office of Coast Survey, viewed 14 July 2020, Available at https://svn.pydro.noaa.gov/Docs/html/Pydro/universe_overview.html.

Greenaway, S., Batts, A., and Riley, J. (2020). “Are We Done Yet? An Empirical Estimator for Level of Effort for Seafloor Surveys - Including an Estimate for the Full Survey of U.S. Waters”, *Marine Geodesy*. 43(2), pp. 87-104, doi: 10.1080/01490419.2019.1705449.

Heinisuo, O. (2020) *Opencv-python 4.3.0.36*, MIT License, viewed 15 July 2020, Available at https://opencv-python-tutroals.readthedocs.io/en/latest/index.html

HydrOffice. (2020) *SmartMap. Smart map the world!*, viewed 21 October 2020, Available at https://www.hydroffice.org/smartmap/main

International Hydrographic Bureau. (2008). *IHO Standards for Hydrographic Survey, Special Publication 44, 5th Edition* [online], viewed 12 July 2020, Available at: https://iho.int/uploads/user/pubs/standards/s-44/S-44_5E.pdf.

National Centers for Environmental Information. (2011). *Water Column Sonar Data Collection*, doi:10.7289/V5HT2M7C, viewed 6 July 2020, Available at https://www.ngdc.noaa.gov/mgg/wcd/#:~:text=or%20Individual%20Cruises.-,Water%20Column%20Sonar%20Data,species%20identification%2C%20and%20habitat%20characterization.

National Centers for Environmental Information. (2015). *Bathymetric Data Viewer*, viewed 6 July 2020, Available at: https://maps.ngdc.noaa.gov/viewers/bathymetry/

National Centers for Environmental Information. (2018). *NOAA expands public access to big data*, Viewed 24 July 2020, Available at https://www.ncei.noaa.gov/news/noaa-expands-big-data-access

National Geodetic Survey. (2019). *Experimental GEOID Modeling (xGEOID)* [online], viewed 24 July 2020, Available at https://beta.ngs.noaa.gov/GEOID/xGEOID/

NOVELTIS, LEGOS, CLS Space Oceanography Division, CNES. (2014). *FES2014*, Distributed by AVISO+ Satellite Altimetry Data, viewed 6 July 2020, Available at: https://bitbucket.org/cnes_aviso/fes/src/master/

Office of Coast Survey. (2016). *How accurate are nautical charts?* [online], viewed 16 July 2020, Available at: https://www.nauticalcharts.noaa.gov/updates/how-accurate-are-nautical-charts/.

Office of Coast Survey. (2017). “W00427, PMEL/AFSC/NOAA Saildrone Research 2016”, *National Ocean Service*, internal report.

Office of Coast Survey. (2020). *NOAA Announces New Progress Report On Mapping U.S. Ocean, Coastal, And Great Lakes Waters*, viewed 6 July 2020, Available at: https://www.nauticalcharts.noaa.gov/updates/noaa-announces-new-progress-report-on-mapping-u-s-ocean-coastal-and-great-lakes-waters/.

Van der Walt, S., Colbert, C., and Varoquaux, G. (2011). "The NumPy Array: A Structure for Efficient Numerical Computation", C*omputing in Science & Engineering*, 13, pp. 22-30, doi:10.1109/MCSE.2011.37.

Virtanen, P., Gommers R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, I., Feng, Y., Moore, E., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", *Nature Methods* 17, pp. 261–272. doi:10.1038/s41592-019-0686-2.

Walsh, T. (2020). Personal communication.

Weber, T. (2020). Personal communication.

Westington, M. (2020). Personal communication.

Westington, M., Miller, J.J., Batts, A., and Armstrong, A. (2019). "Creating a Seafloor Mapping Plan to Fill U.S. Gaps by 2030", *OCEANS 2019 MTS/IEEE SEATTLE,* pp. 1-9, doi: 10.23919/OCEANS40490.2019.8962563.

Wyllie, K. and Rice, G. (2020). *Building the National Bathymetry*, Office of Coast Survey, viewed 28 July 2020, available at: https://www.nauticalcharts.noaa.gov/updates/building-the-national-bathymetry/.

Appendix: List of Figures and Captions

Figure 1 (half-page width). Single beam echo sounder compared to swath echo sounder. Both methods record the bathymetry of the seafloor. The EK60 echosounder is represented by the single beam.

Figure 2 (half-page width). Available NMFS EK60 data for the entire US Exclusive Economic Zone from 2003-2019 (National Centers for Environmental Information, 2011).

Figure 3 (half-page width). Bathymetric data coverage for Northeast US EEZ (left) and additional NMFS EK60 data coverage for the same region (right). Notice how the EK60 data covers a large unmapped area of the United States EEZ. Figure credit: Meredith Westington, NOAA Office of Coast Survey.

Figure 4 (half-page width). A subsection of the sobel gradient and power time series in the area of a detected feature. The raw power and gaussian filtered power illustrate change relative to the sobel filter gradient. We demonstrate the detected feature as extended to the minimum gradient, and how this range of samples corresponds to the maximum power amplitude.

Figure 5 (half-page width). An example of an average sound speed profile from WOA (Boyer et al., 2013). These profiles were used to calculate a harmonic sound speed used to correct the EK60 two-way travel time.

Figure 6 (half-page width). Mean difference between AVISO+ FES2014b tide predictions and NOAA CO-OPS discrete zoned tide observations at 67 offshore locations.

Figure 7 (half-page width). CO-OPS Tide Zones used for comparison to the AVISO+ FES2014b tide predictions.

Figure 8 (half-page width). Comparison of AVISO+ FES2014b predicted tide values and verified tides in Rhode Island Sound.

Figure 9 (half-page width). Histogram showing distribution of differences between AVISO+ FES2014b predictions and NOAA CO-OPS discrete zoned verified tide - corrected values in Rhode Island Sound.

Figure 10 (half-page width). Example of overview plot generated by the EK60 script for RAW file D20070712-T124906 of the NOAAS *Henry B. Bigelow*. The top graph shows seafloor detections for all transducers, which are all colocated in this case, while the bottom shows values for tides, sound speed, heave, and draft.

Figure 11 (half-page width). Example of time series image for one transducer generated by the EK60 script for RAW file D20070712-T124906 of the NOAAS *Henry B. Bigelow*. For each RAW file, an image like this one was produced for each transducer, providing a visual for the water column and seafloor at that location.

Figure 12 (half-page width). Cruise tracklines for the 19 NMFS cruises processed for seafloor detections and the results comparison region with survey BAG file used for analysis (National Centers for Environmental Information, 2015; National Centers for Environmental Information, 2011).

Figure 13 (half-page width). Distribution of difference values between *Henry B. Bigelow* seafloor detections and BAG survey depths.

Figure 14 (half-page width). Distribution of difference values between *Henry B. Bigelow* seafloor detections and BAG survey depths with outliers removed (removed anything greater than two meters absolute difference).

Figure 15 (half-page width). Distribution of *Henry B. Bigelow* seafloor detections results with difference values between 2.5 and 4.5 meters deeper than reference survey depths. The approximate three meter bias is thought to result from FSV with six meter draft, centerboard-retracted surveys, with missing draft values.

Figure 16 (half-page width). Distribution of absolute difference between seafloor detections and BAG survey depths.

Figure 17 (half-page width). HB1304 grid cells with a near-constant difference between 2.5-3.5 meters, indicative of an erroneous draft corrector value.

Figure 18 (half-page width). Distribution of Pisces1301 difference values between seafloor detections values and BAG survey depths. The two aggregations around -3 meters and 3 meters difference are likely the result of missing or inaccurate draft measurements.

Figure 19 (half-page width). Distribution of Pisces1301 difference values between seafloor detections and BAG survey depths with outliers removed (removed anything greater than two meters absolute difference).