HYDROGRAPHISCHE NACHRICHTEN

Journal of Applied Hydrography *03/2023* *HN 124*

Depth measurement per crowdsourcing. Damn it!

DEUTSCHE HYDROGRAPHISCHE GESELLSCHAFT e.V.
DHyG

DOI: 10.23784/HN124-03

Assessing CSB data reliability

Estimating vertical uncertainty of sample CSB data by comparing with reference multibeam data

An article by IDRIS SALAUDEEN

Crowdsourced bathymetry (CSB) is depth information gathered by mariners who voluntarily use standard acquisition sensors on their ships during their routine operations at sea. The data collected from these sources is then made publicly available. However, due to the unregulated and unsupervised nature of the data collection, the reliability of the information obtained from crowdsourced bathymetry is in question. This highlights the importance of assessing the accuracy of the data. The study explores the use of geostatistical methods to evaluate the reliability of the depths obtained from crowdsourced sources by comparing them to trusted bathymetric data. The results indicate that the density of soundings plays a significant role in determining the reliability of the crowdsourced depths, and it is possible to calculate uncertainty estimates even when the error budgets are unknown. However, the reliability of the soundings must be measured against the acceptable limits set by the International Hydrographic Organization (IHO) as outlined in the *S-44* standards for hydrographic surveys.

crowdsourced bathymetry – CSB | Delaware Bay | BAG file | uncertainty | S-44
Crowdsourced Bathymetry – CSB | Delaware Bay | BAG-Datei | Unsicherheit | S-44

Bei der Crowdsourced Bathymetry (CSB) handelt es sich um Tiefeninformationen, welche von Seeleuten gesammelt werden, die während ihres Routinebetriebs auf See freiwillig Standardsensoren auf ihren Schiffen verwenden. Die von diesen Quellen gesammelten Daten werden dann öffentlich zugänglich gemacht. Aufgrund der ungeregelten und unbeaufsichtigten Art der Datenerfassung ist die Zuverlässigkeit der aus der Crowdsourced Bathymetry gewonnenen Informationen jedoch fraglich. Es ist daher wichtig, die Genauigkeit der Daten zu bewerten. Die Studie untersucht den Einsatz geostatistischer Methoden, um die Zuverlässigkeit der aus Crowdsourcing-Quellen gewonnenen Tiefen zu bewerten, indem sie mit zuverlässigen bathymetrischen Daten verglichen werden. Die Ergebnisse deuten darauf hin, dass die Dichte der Tiefenmessungen eine wichtige Rolle bei der Bestimmung der Verlässlichkeit der Crowdsourcing-Tiefen spielt und es möglich ist, Unsicherheitsschätzungen zu berechnen, selbst wenn die Fehlerbudgets unbekannt sind. Die Zuverlässigkeit der Tiefenmessungen muss jedoch an den von der Internationalen Hydrographischen Organisation (IHO) festgelegten akzeptablen Grenzwerten gemessen werden, wie sie im Standard *S-44* für hydrographische Vermessungen beschrieben sind.

Author
Idris Salaudeen is Technical Support Expert at SevenCs in Hamburg.

idris.salaudeen@sevencs.com

1 Introduction

1.1 Background

The oceans make up 70 percent of our planet, and there seems to be a rather large misconception that the earth and the oceans have been fully mapped – especially looking at paper maps and globes of the world which typically show all the continents and oceans. However, this is not the case from a bathymetric standpoint as these maps lack detailed information on the depths of the ocean floors, but rather provide estimated information from satellite observations of the sea surface heights (CCOM 2022).

Crowdsourced bathymetry has the potential to greatly increase the global bathymetric coverage and improve the understanding of the world's oceans. The IHO recognises the potential of crowdsourced bathymetry in augmenting existing bathymetric data and has taken steps to encourage its adoption. The CSBWG was formed to help establish standards for collecting and utilising CSB data, and to work towards overcoming any technical or logistical challenges that may arise during collection efforts. This is in line with the IHO's goal of ensuring that the world's oceans and waterways are properly mapped and charted for safe navigation and effective maritime activities (IHO HSWG 2022).

1.2 The need for crowdsourced bathymetry

Traditional hydrographic surveys are typically expensive, time consuming and logistically challenging and they can only cover a small portion of the world's oceans, leaving vast areas unmapped, especially remote areas.

Crowdsourced bathymetry offers a solution to these challenges by leveraging the power of community involvement and affordable modern technologies such as single-beam echo sounders, GPS systems and data loggers. The concept is simple: anyone with a platform equipped with these technologies can collect bathymetric data and contribute to the mapping of the world's oceans. This data can then be combined with other data sources to create comprehensive bathymetric maps (Jencks et al. 2021).

1.3 CSB error components and data reliability

The reliability of CSB data is typically a subject of concern because the data is often acquired during routine voyage operations using low-cost sounding equipment and data loggers. Added to this underlying concern is the lack of metadata information that is useful to determine the various error sources and estimate the total error propagated on CSB measurements. These include instrument errors, vessel draft errors, speed of sound errors, motion and attitude errors, and tides and water levels errors. All these errors shelter several component errors that constitute the total propagated uncertainty (TPU) upon crowdsourced depth measurement which affects the overall quality and reliability of CSB data (Fig. 1).

1.4 Goal

The goal of this study is to analyse a sample CSB data set and determine its reliability for use in hydrographic applications. This study aims to estimate the uncertainty of the CSB data comparing the collected depth measurements with those obtained from a standard hydrographic multibeam survey. The goal is to determine the reliability of the sample data and draw conclusions on its usage and applications in hydrographic contexts.

2 Data set and area of study

The CSB and MBES data sets analysed in this study were freely obtained from the IHO DCDB website. About 4 GB of data was downloaded in total from the website, as separate compressed folders for each data set (CSB and MBES) in ZIP format.

Both CSB and MBES data sets are from the Delaware Bay area in the east coast of the United States (Fig. 2). The area is a shallow water area with an average depth of around 17 metres. The area also serves as an anchorage area for vessels and leads to the inner areas of the Delaware river, where the vessels can enter the Chesapeake & Delaware Canal (C&D Canal). The canal connects Delaware area to the Chesapeake area of Maryland, United States.

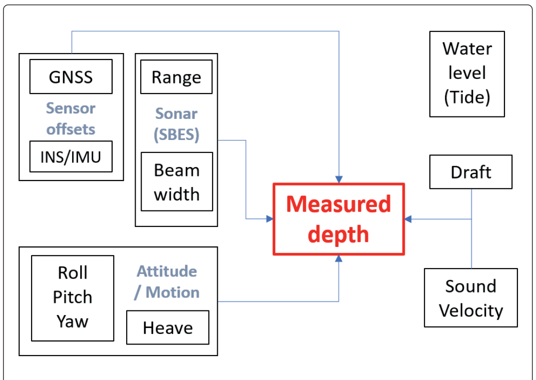

Fig. 1: Illustration of sources of errors propagated on measured depths

2.1 CSB data

The CSB data sets were obtained as lines in GeoJSON and XYZ formats and compressed into a ZIP folder. The collective line data is primarily constituted of over 500,000 individual points that have been acquired with the Rose Point Coastal Explorer software. The software passively logs depth and position data that is being collected by the echo sounder and navigation sensors onboard the acquisition vessel.

2.2 MBES data

The MBES data were obtained as Bathymetric Attributed Grid (BAG) files. The BAG format is a form of bathymetric raster data typically having two bands – the elevation (depth) band, and the uncertainty band. It is noteworthy that BAG files are not necessarily raw files but have already undergone advanced processing with all necessary corrections already applied. In fact, BAG files are finished products of any acquired survey, as they

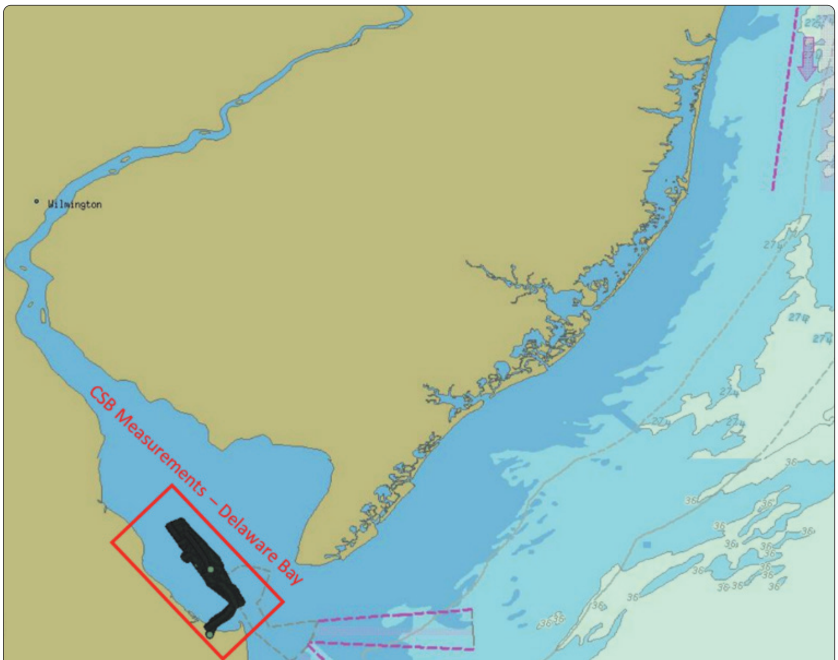

Fig. 2: View of CSB data set in the area of study

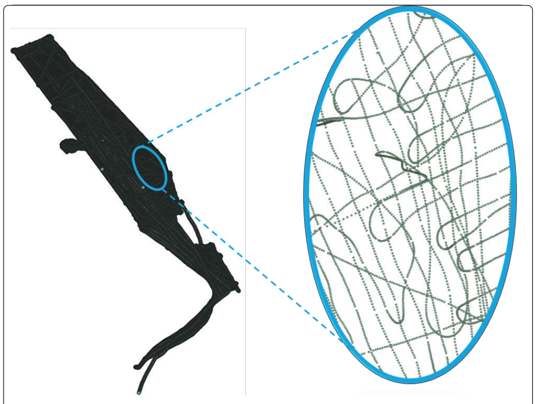

Fig. 3: Graphic of the CSB data sets in Delaware Bay depicting the individual points

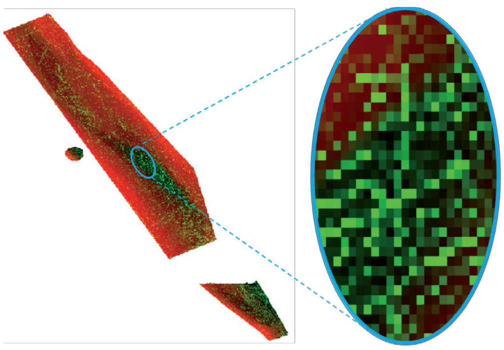

Fig. 4: Graphic of the MBES data sets in Delaware Bay depicting the individual cells (pixels)

would have been fully processed and attached with an uncertainty layer.

The MBES data were acquired by the US National Oceanographic and Atmospheric Administration (NOAA) hydrographic survey vessels equipped with standard sensors for hydrographic data acquisition.

2.3 Metadata

The metadata files for the CSB lines were available as JSON files in the compressed ZIP folders for each CSB data set. Individual CSB lines also contained metadata information embedded into the GeoJSON files. The metadata for the MBES grids were available as separate files that could be downloaded from the Survey Report Page of the data of interest on the NOAA NCEI website. The MBES metadata are primarily available in XML format that was already integrated into the BAG file.

An overview of metadata information provided includes:

- information about the acquisition platform,
- information about the data provider,
- the Coordinate Reference System (CRS) information.

3 Research problem

Estimation of uncertainties in any scientific measurement is to quantify the errors propagated within the measurement and provide a sense of quality of the observed measurements. It also allows for attributing of a certain level of reliability to the observed measurements.

However, the issue with attempting to estimate uncertainties and quantify the errors propagated within physical measurements such as depths is that the true depth value is usually unknown, and it is unlikely that any measurement would precisely equal to the true value.

Since the true depth value is usually unknown and it is unlikely that any measurement would precisely equal to the true value, the IHO has addressed this issue by employing the use of »Uncertainty« and »Confidence Levels« to quantify errors and determine the accuracy of depth measurements (Sanders 2011). This rationale will guide the methods of this study.

4 Methods

This study attempts to estimate the propagated uncertainties in crowdsourced bathymetric data, and since the CSB data were acquired using a single-beam echo sounder, the study adopts the »comparison with multibeam« method of measuring uncertainty. A well-established method for analysing single-beam measurements that is recognised by the International Hydrographic Organization (IHO) and has been acknowledged in IHO Publication *B-11* and IHO Publication *S-44*.

The method involves comparing the single-beam CSB measurements with the available reference data from multibeam echo sounder (MBES) surveys, to assess the reliability of the CSB measurements. The MBES data are obtained from a hydrographic office (NOAA) and are considered to be the standard and authoritative data for comparison. The deviations between the CSB and MBES depths are used to estimate the uncertainty of the measured CSB depths.

Therefore, geostatistical methods were adopted in this study to assess the quality of the CSB data. The method involves calculating the error distribution of the depth values based on Total Vertical Uncertainty (TVU) calculations specified by the IHO *S-44* standards.

The maximum allowable TVU is calculated using this equation:

$$TVU_{max} = \pm\sqrt{a^2 + (b \cdot d)^2}$$

Where:

- a represents that portion of the uncertainty that does not vary with depth,
- b is a coefficient which represents that portion of the uncertainty that varies with depth,
- d is the depth,
- (b · d) represents that portion of the uncertainty that varies with depth.

4.1 Workflow concept

A simple processing workflow that follows the method stated above to determine the reliability of the CSB data was designed. The workflow process overlays the CSB point depths on the MBES gird and compares both measurements in areas where corresponding depth positions.

4.2 Tide correction and datum conversion

Correcting for tides and converting the CSB data to the same vertical reference as the MBES data is an important step in ensuring an accurate comparison of the two data sets. The tidal corrections help to eliminate vertical offsets that exist between the two data set reference systems. This involves converting the CSB data from its original ellipsoidal vertical reference (WGS84) to the same vertical reference as the MBES data, which in this case is the Chart Datum.

The MBES data has already been corrected for tides and referenced to the Mean Lower Low Water (MLLW), which is the Chart Datum used for US charts produced by NOAA. To correct the CSB data, the VDatum software was used. VDatum is a vertical datum transformation tool developed by NOAA and is widely used in hydrographic surveys in the US. It can transform vertical data between different vertical and horizontal datums, using actual tide data and tidal predictions to produce accurate tide and water level corrections to depth data.

By using VDatum, the CSB data was converted to the same vertical reference as the MBES data, allowing for a more accurate comparison between the two data sets.

4.3 Processing

After the corrections were made, the workflow process was implemented using the FME software, which was chosen due to its ability to handle large processing of many depth points. FME is a licensed software and may not be freely accessible. However, QGIS software could also be used as an alternative tool to implement this process, but it may be very slow and may sometimes become unresponsive depending on the size of the data set.

The output from FME was a CSV file containing CSB and MBES depths of corresponding positions, i.e. X, Y and both Z_{CSB} and Z_{MBES}. The CSV file was later imported into other hydrographic processing software like Qimera for further processing as well as QGIS software and Microsoft Excel for statistical analysis and graphical visualisations.

5 Results and discussion

This section includes some graphics for visual comparisons between the processed CSB data and the reference MBES data. It proceeds to outline the results of further statistical analysis and uncertainty estimations carried out on the data sets based on the methodology defined for the study and in line with IHO recommendations.

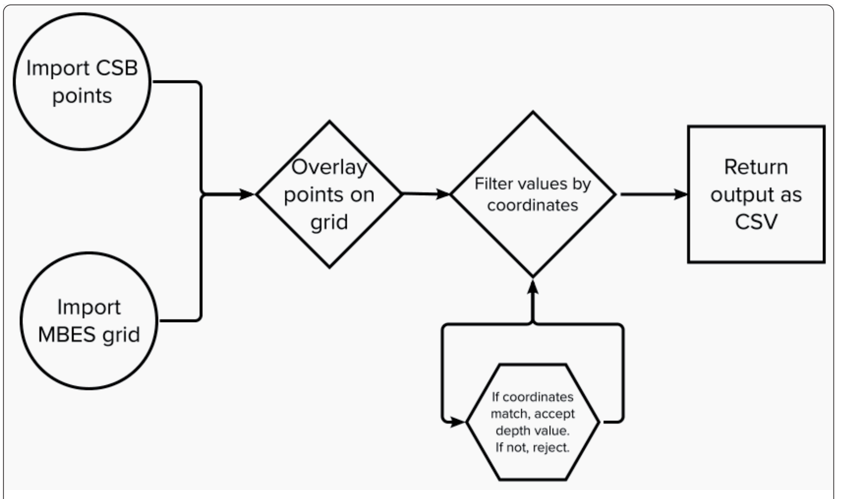

Fig. 5: Simple illustration of the data processing workflow

Firstly, the CSB data was converted from points to a gridded bathymetric surface in Qimera to allow for visual comparison of the CSB and MBES data set and to see how the generated CSB grid will compare visually to the MBES grid.

A cell size of at 30 m was defined for creating the CSB grid surface, which is the same resolution as the reference MBES grid. This is to ensure a balanced visual comparison of both grids and see how well the CSB data represents the seafloor (Fig. 6).

Based on the images provided, it seems that the grid surfaces of the CSB and MBES are quite similar. However, there are still some holes visible on the interpolated CSB grid surface. Further interpolation could be done to cover these holes and eliminate any gaps, but this may increase the risk of interpolation errors. Despite the holes, the CSB surface still looks very similar to the reference MBES surface. If

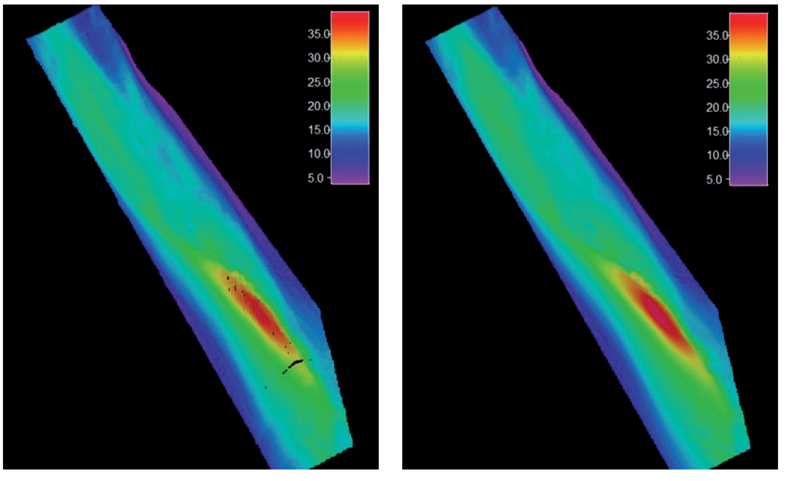

Fig. 6: Images of both CSB (left) and MBES grid surfaces (right) at 30 metre resolution for visual comparison

Details	CSB	MBES
Dimensions	512 rows × 384 columns	512 rows × 512 columns
Cell size	30 m	30 m
X range	478,440.00 to 489,960.00 m	476,335.73 to 491,864.35 m
Y range	4,304,430.00 to 4,319,790.00 m	4,304,002.28 to 4,319,530.90 m
Z range	0.5 to 40 m	3.56 to 39.06 m
Coordinate system	NAD83 / UTM zone 18N 2	NAD83 / UTM zone 18N 2
Mean depth	17.67 m	17.91 m
Standard deviation	5.76 m	5.78 m
Surface area (2D)	50,055,300 m^2	49,889,074.725 m^2
Number of soundings	383566	N/A

Table 1: Statistics summary table for CSB and MBES surfaces in Delaware Bay

the holes are ignored, it is difficult to distinguish between the two surfaces. Moreover, a statistical summary of the surface information for both CSB and MBES grids further proves the similarities between the surfaces and are outlined in Table 1.

To better illustrate the agreement between the surfaces, a cross-section analysis was conducted by selecting an area of interest and plotting the cross-section profiles of both data sets on a graph to display any differences in conformity. This approach provides a clear visualisation of how the cross-section profiles compare between the two data sets (Fig. 7).

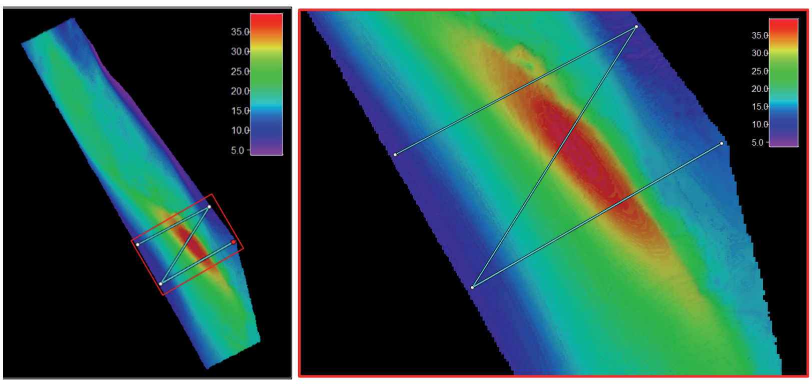

Fig. 7: Cross-section line drawn on CSB and MBES gridded surfaces

The cross-section profiles for the CSB and MBES measurements indicate a high degree of conformity, with very little difference between the two profiles (Fig. 8). The difference profile appears almost flat and straight, with only a few minor variations in the form of zigzags, and the lines depicting the profiles are almost indistinguishable, except for their different colours.

A histogram of the depth soundings was plotted to show the distribution within the CSB and MBES measurements (Fig. 9). The distribution plots are helpful to depict consistency between a measured variable and the true value of the variable.

As revealed by the distribution plots, the CSB measurements display a similar distribution to the MBES measurements, with the majority of the soundings falling within the same range. There is no significant displacement between the two data sets, indicating the absence of systematic bias. The consistency between the two data sets is a result of the sounding density of the CSB data. The CSB measurements can be used as a reliable alternative to map the gaps in the Delaware Bay area. However, it is important to note that this conclusion is based on the specific data set and methodology used in this investigation and may not necessarily hold for other data sets or surveying methods.

The next step is to assess the performance of the CSB measurements against the International Hydrographic Organization (IHO) *S-44* minimum standards for hydrographic surveys. The evaluation is critical to determine the reliability of the uncertainty assessments carried out in this study for the investigated CSB measurements.

An error distribution was then plotted from the depth difference between the CSB and MBES measurements (Fig. 10). The error distribution plots are graphical representations that display the distribution of random errors between the observed variables. In this case, the variables are the CSB measurements and the reference MBES measurements. The plots are useful in quantifying the errors in the investigated observations by using upper and lower error limits that have been calculated at a 95 per cent confidence level, as recommended by the IHO.

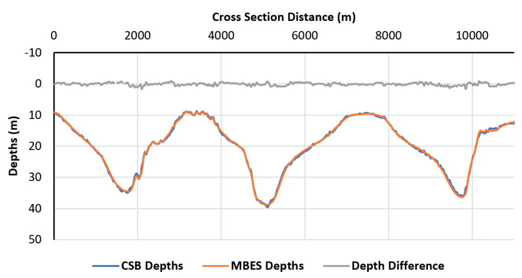

Fig. 8: Cross section showing conformity and the difference between CSB and MBES surface profiles

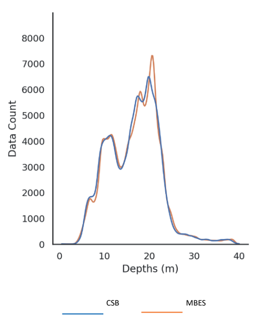

Fig. 9: Soundings distribution graphs of CSB and MBES depths

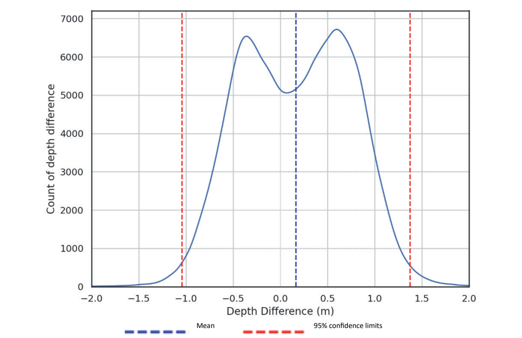

Fig. 10: Error distribution graph of the depth bias between CSB and MBES measurements

The result of the error distribution plot was ascertained further using the Cross Check tool in the Qimera software to generate a Cross Check Report which is presented in Fig. 11.

The Qimera Cross Check tool performs a statistical analysis of beam footprint values referenced to a selected Dynamic or Static Surface (QPS Qimera 2022). In this case, the tool performs a statistical analysis on the CSB point depths referenced to MBES grid surface. The Cross Check report provides values such as standard deviation and mean error necessary to estimate the uncertainty and determine the reliability of the CSB data. It goes further to test the reliability of the CSB data against IHO survey orders and finds that the estimated uncertainty for the CSB data is accepted within the IHO Survey Order 2 limits.

A scatter plot of the CSB soundings against the depth difference was plotted within IHO Survey Order 2 limits to verify if 95 percent of the soundings truly fall within acceptable error limits of IHO Order 2 survey as calculated in Qimera.

The plots show the accepted and rejected depth soundings at 95 % confidence level and further depicts the distribution of the soundings and sounding errors with respect to the IHO *S-44* acceptable limits for Order 2 surveys (Fig. 12). It depicts majority of the accepted soundings falling within the IHO Order 2 limits. Hence, the CSB measurements taken in the Delaware Bay area conform with IHO Order 2 standards and would be accepted as an IHO Order 2 survey.

To calculate the vertical uncertainty, the IHO recommends that uncertainties in depth measurements shall be expressed using confidence levels (IHO 2020). According to the IHO *S-44* publication, »the 95 % confidence level for 1D quantities (e.g. depth) is defined as 1.96 × standard deviation«.

Therefore, the uncertainty estimates for the CSB measurements acquired within this study shall be empirically determined and calculated in accordance with the IHO *S-44* recommendations, using the standard deviation values of the concerned measurements (which is 0.597 m).

This implies that 95 % of the CSB measurements fall within an estimated ±1.2 m depth accuracy.

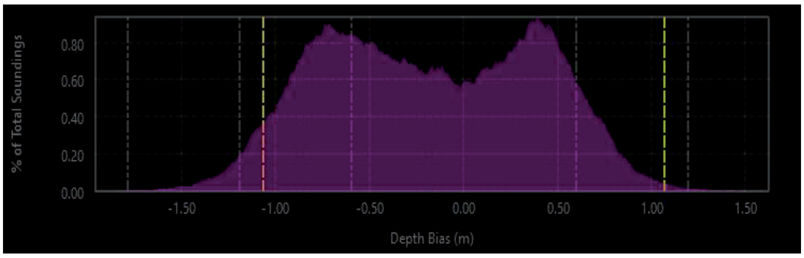

Fig. 11: Histogram plot of the error distribution from the depth bias against IHO limits

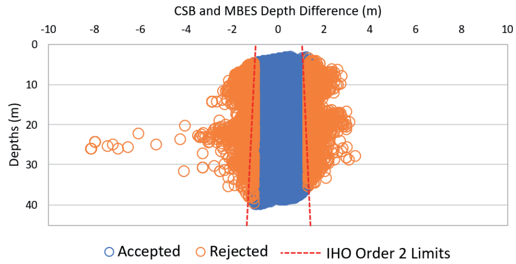

Fig. 12: IHO *S-44* plot – accepted soundings (in blue) fall within IHO Order 2 limits

Summary and conclusion

In this study, statistical methods were adopted to assess the quality of CSB data. The uncertainty was estimated from the computation of error propagated on the depth values based on the Total Vertical Uncertainty (TVU) calculations specified by the IHO in the IHO *S-44* standards.

In conclusion, this study found that uncertainty estimates for crowdsourced bathymetry (CSB) measurements can be determined through statistical computations using a reference measurement such as multibeam surveys. The reliability of CSB measurements can be further established by assessing them against appropriate IHO *S-44* survey order standards. Sounding density was found to have a clear relationship with estimated uncertainties as depicted in the sounding distribution plots. The completeness of a bathymetric data set is a function of sounding density, and the level of measurement completeness achieved for CSB measurements depends on the density of soundings within the survey area of interest. //

References

CCOM (2022): Bathy Globe Project. http://ccom.unh.edu/project/bathymetry-globe

IHO (2020): Standards for Hydrographic Surveys – S-44 Edition 6.0.0. International Hydrographic Organization, Monaco

IHO (2019) The IHO-IOC GEBCO Cook Book – S-11. International Hydrographic Organization, Monaco

Jencks, Jennifer; D. Wyatt; M. Pratellesi; P.. Wills (2021): A Commitment to crowdsourced bathymetry citizen sourced data – Help reveal the deep and share your data. International Hydrographic Review, No. 26

QPS Qimera (2022): Qimera Point to Surface Interpolation Method – Qimera Customer Support. https://jira.qps.nl/servicedesk/customer/portal/7/CQM-5556

Sanders, Pat (2011): S-44 for Dummies. https://www.yumpu.com/en/document/read/35884883/s-44-for-dummies-hypack