NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
UNITED STATES DEPARTMENT OF COMMERCE

Multibeam Echosounder Uncertainty for NOAA's Office of Coast Survey

NOAA Office of Coast Survey

April, 2024

MBES Uncertainty For for NOAA's Office of Coast Survey

April 2024

Brian Calder [a], Bart Buesseler [b], Glen Rice [b], Matthew Wilson [b], Harper Umfress [b]

[a] *Center for Coastal and Ocean Mapping & Joint Hydrographic Center*

[B] *NOAA's Office of Coast Survey*

Executive Summary

NOAA's Office Of Coast Survey requires consistency in metadata to enable new and innovative products and processes that are of high value to the public. Thanks to recent advances within hydrography, the current metadata requirements for the uncertainty values of Multibeam Echosounder (MBES) data within the Bathymetric Attributed Grid (BAG) file format have been found to be inadequate and in need of refinement. This paper provides the background leading up to this need, a brief overview of the intention behind uncertainty values, and finally a revised version of MBES uncertainty requirements to support current and future products. These new requirements mandate the use of Combined Uncertainty and Bathymetric Estimator (CUBE)[1] or CUBE with Hierarchical Resolution Techniques (CHRT)[2] gridding algorithms to effectively exclude spurious data from consideration, and computes a variance and resulting uncertainty from those remaining data points (or "pings") for the resulting depth estimation of any individual grid node.

Background

NOAA Information Quality Act Policy states data should be qualified such that it can be effectively used by the public[3]. In hydrographic science, bathymetric surveys are qualified by coverage, feature detection capability, and uncertainty. To ensure confidence in the operational use of the bathymetry provided by NOAA's Office of Coast Survey (Coast Survey), the reported bathymetry uncertainty needs to have a scientifically sound and operationally compatible definition.

Coast Survey's understanding of Uncertainty has progressed dramatically since uncertainty metadata was being designed in 2003. It has progressed from an isolated QC tool to a

---

[1] B. R. Calder and L. Mayer. Automatic Processing of High-Rate, High-Density Multibeam Echosounder Data, Geochem., Geophys. and Geosystems (G3), 10.1029/2002GC000486, 4(6), 2003.

[2] B. R. Calder and G. A. Rice. Computationally Efficient Variance Resolution Depth Estimation. Computers and Geosciences, 106:49-59, 2017.

[3] The entire policy can be read at the following link: https://www.noaa.gov/organization/information-technology/policy-oversight/information-quality/information-quality-guidelines

fundamental representation of data quality, used in the supersession logic for our National Bathymetry[4].

When initially defined, the MBES "*Product_Uncert*" field of the BAG specification was still in flux, causing the original developers to leave additional flexibility[5] in the definition while these discussions played out. Although there is a controlled-vocabulary list of well-known interpretations of what's in the BAG uncertainty layer, the File Specification Document (FSD) does not prescribe how those list items should be generated. It may be interpreted as a standard deviation, or a variance, and also it is not specific about the number of soundings contributing to the solution. For example, if the uncertainty metadata indicates "Standard Deviation", does that mean the standard deviation of all of the observations that are used to construct the depth estimate? Or, possibly, the standard deviation of all of the observations in the vicinity of the depth estimate (i.e., including potential noise). Additionally, we have seen interpretations of this definition using an arithmetic mean of propagated uncertainties, while others use a harmonic mean, resulting in fundamentally different results, even when determined from the same point cloud. This ambiguity in the definition of "Product_Uncert" has introduced variance into software vendor's implementations, creating subtle, or in some cases, radical, differences in how uncertainty is reported. Furthermore, as much of the software code used by vendors may be proprietary, it is exceptionally difficult to get meaningful comparisons for observed differences between data sets.

Not only has the specification been in flux, but the evolution of uncertainty calculations has led to varied levels of understanding of the topic within Coast Survey. Given the fundamental importance of uncertainty values in modern bathymetry, education is also a critical aspect that must be addressed in order to increase corporate knowledge around the topic. This document will serve to address both sides of the problem, by both providing a primer on uncertainty to increase Coast Survey's understanding of the topic, as well as identifying a clear path forward for calculating MBES uncertainty in BAG products.

A Brief Uncertainty Primer

Models of Uncertainty

All measurements of physical quantities come with some level of variability. For example, if you weigh 100 bags of sugar that are labeled as 500g, although it would be reasonable to expect that they would all be *approximately* 500g, it would be unusual for them all to be *exactly* 500g. That is, you would (and should) expect each bag to have a weight (or mass) in the range of, maybe 480g to 520g, although the average weight computed by the usual arithmetic mean (the sum of weights divided by the number of samples, in this case 100) should ideally be arbitrarily close to 500g - so long as the manufacturer is honest, and the scales are fair.

---

[4] For more info on the National Bathymetric Source see https://nauticalcharts.noaa.gov/learn/nbs.html
[5] Current definition: *"NOAA standard product uncertainty V1.0 (a blend of CUBE uncertainty and other measures)"* See Table 1 in the BAG spec at https://bag.readthedocs.io/en/master/fsd/FSD-BAGStructure.html#uncertainty-grid

Although simple, this example highlights most of the key elements in specifying and interpreting uncertainty. That is:

(a) Everything is uncertain. Every measurement - without exception - has some level of uncertainty. It can be reduced to some extent by appropriate methods, but cannot be avoided, or completely removed: at best, we need to estimate how big it is, and any features it might have (e.g, biases, distribution), and provide this information unambiguously to the end user of the data.

(b) Every measurement is different. Even if you have a standardized process, each time you use it, you are going to get a variation. This is true whether you are filling sugar into bags or measuring depth (even nominally the same depth). Uncertainty is a measure of the level of the expected variability.

(c) Error and Uncertainty are not the same thing. While you cannot say beforehand what the difference is going to be between the target (e.g., a 500g bag) and the reality (i.e., what the bag weighs), you can estimate what you think it is likely to be, either by looking at lots of measurements, or through your previous experience. The former (actual difference) is the error, the latter the uncertainty. Usually, since you don't know the exact value of the thing that you are trying to measure (e.g., the depth), you cannot actually find the error: only estimate the uncertainty.

(d) There are two different types of uncertainty that we might have to consider: variations of the bag's weight about 500g that are from variations in the loading machine or sugar (e.g., if the sugar is clumped and a chunk falls in at the last minute), and systematic shifts in the weight of the bag (e.g., if the manufacturer intentionally or accidentally put an extra weight on the measurement platform so it always under-fills the bags, of if there is a fault in the user's scale that always causes the bags to read heavy). These two different types of uncertainty are known by different names depending on context, but the most common pairs are[6]:

- Variance (random) and Bias (shift)
- Stochastic (random) and Deterministic (shift)
- Precision (random) and Accuracy (shift)

In general, random effects in bathymetry come from variations in the measurements (e.g., electrical or acoustic noise), and shifts come from uncorrected offsets (e.g., a bad patch test) but can come from shifts in environmental factors, particularly sound speed. It is usually assumed that we can (and do) resolve most biases before collecting (or at least processing) our data. This is not always possible, however, or practical[7].

---

[6] The precise meanings can change a little depending on whether you are talking about the data or the process that generates the data, but these distinctions are more than required for this assessment.

[7] For example, we might technically be able to resolve a sound speed mis-observation by taking more observations, but choose not to so as to avoid stopping the survey frequently

(e) Uncertainty is not, usually, a blunder. For example, if the sugar bag has a manufacturing defect that means that half of the sugar falls out before it gets to the store, or if the bag doesn't open fully and half of the sugar falls on the floor when it is being filled, this is different from the natural variation in the weights of the correctly filled, correctly made bags. In a bathymetric context, we might make a distinction between the "fluffy" variation of depths detected on a flat surface, and an "outlier" from an acoustic return off the hull of the survey platform, or bubble sweep-down passing the transducer. This does not imply that an uncertainty cannot be estimated for all measurements - a fundamental assumption in estimating uncertainty is that we don't have any blunders, and so having an uncertainty that does not match what the data does is a very useful way of detecting blunders (and is used in CUBE, CHRT, and many other processing schemes).

(f) Source and Product uncertainty are different. Every separate measurement has uncertainty, so any estimate made from these measurements also has uncertainty (this is the "total" in Total Propagated Uncertainty). In many cases, this process of combining measurements and tracing the associated uncertainty of the result is typically called propagation of uncertainty (or variance), and affects, for example, depth observations, where we combine an estimate of target two-way travel time, speed of sound (surface and profile), positioning, offsets, draft, and motion (and potentially other things) into an (x,y,z) declared depth. To consider depth observations as being without uncertainty (e.g., the infamous "golden sounding") is therefore woefully naive.

A special case of this which often causes confusion is where you have multiple observations of the same thing, for example the weights of different bags, and you then combine them together to give you a better estimate of the nominal weight of an ideal bag of sugar (remember: no such bag does, or can, exist). This is typically done by averaging the measurements, often using just the simple arithmetic mean. You can apply the normal uncertainty propagation method to this equation, but intuitively you can see that if the weights of the different bags of sugar are randomly distributed about the ideal weight (some higher, some lower), you should have as many above the ideal weight as below, and therefore the average should tend towards the ideal weight as these random variations cancel out in the sum[8].

But what is the uncertainty of the mean? Certainly, the estimated value should be closer to the ideal value since the over and under estimates are partially canceling out, so the uncertainty should be smaller. Under some basic assumptions, the normal assumption is that the variance should be inversely proportional to the number of samples used. Since this value says something about how much error there is likely to be in the estimate, it is usually called the standard error. This is often where the confusion sets in: the measurements all have the same uncertainty, say $s\ (g^2)$, but the arithmetic mean has uncertainty $s' = \frac{\sigma^2}{\#\ of\ samples}\ (g^2)$. Because these all have the same units, it can be hard to keep things clear, but one useful mental model is that the individual uncertainty is what you would feel if you took one bag of sugar in each hand and compared them (more or less - there is some technical detail here too), and the

---

[8] There are a number of technical details to worry about here, but they can be ignored at a first pass.

standard error is the difference you would see from a total of 50 kg if you were to put all 100 bags on the scale at the same time. In the bathymetric world, the "fuzziness" of your data in a point cloud is what the individual uncertainty estimates; the standard error is a metric of how well you can use these measurements to establish an estimate of what the underlying depth really is.

Application of Uncertainty in Hydrography

In the practice of hydrography, we usually think of uncertainty in two stages: the uncertainty of individual measurements of depth (source uncertainty), and the uncertainty of the estimated true depth at any given point (product uncertainty). The former uncertainty is typically called "Total Propagated Uncertainty" (TPU), or equivalently the pair of Total Vertical and Total Horizontal Uncertainty (TVU/THU), and is a predictive model, for the most part, of the likely variability of a depth observation given its component parts. The latter is an estimate of the uncertainty associated with an estimated depth, for example as part of a BAG file. From the preceding section it should be clear that these are quite different, but for the purposes of this document it is assumed that the TPU is provided with the input dataset, or is computed routinely as soon as the data enters the processing system[9]. We can therefore focus on the uncertainty of the estimated depth, how to interpret it, how to report it, and how to compute it.

This sequence of questions might seem backwards, but since the function of the uncertainty is to communicate something about the data to the end user, starting with the interpretation guides the rest of the effort. What is it, then, that the user wants to see in the output model, e.g., in the National Bathymetric Source (NBS)? The simplest solution, when using statistical processing such as CUBE or CHRT is just to report the standard error of the estimator, which is the natural output. This value reduces with the number of observations considered, however, and therefore can become very small in an area of dense data. While this is technically correct[10], evidence has shown that most users do not find this a useful summary of uncertainty, since it does not jibe with the evidence of a certain level of "fuzziness" of the source data that is evident in various processing tool point cloud displays.

A suitable alternative to this model is to compute a summary statistic from the observations used to compute the estimated depth which reflects this human observer intuition on the "natural" uncertainty of the data. There are a number of potential methods for this, some more robust to outliers than others, but for estimates based on robust estimators like CUBE or CHRT the simplest and most efficient method is to compute the variance of only the observations that are used to compute the estimated depth. In an area where there is no noise, this statistics will match the "fuzziness" of the data that would be observed in a point cloud display (since all of the data will be integrated in one estimate), and will automatically accommodate for higher uncertainties on slopes as would be expected. In areas where there are significant noise levels,

[9] This paper assumes that an appropriate and well configured uncertainty model (i.e. TPU) is used.
[10] With the caveat that the observations are not truly independent estimates, of which most methods do not compensate, so as to have a feasible computation

more than one potential reconstruction of depth may be present, and the uncertainty estimated for those reconstructions that are generated by blunders might be larger or smaller than expected for real data. However, because this technique focuses only on the observations used for a single reconstruction, assuming that there is a valid estimate of depth in the potential reconstructions (as determined by the algorithm or a human operator), then the reported uncertainty will still robustly report the "fuzziness" of the observations that are believed to be valid estimates of the real depth (and not blunders). This estimate should therefore be robust even in noisy data, at least to the level of robustness of the underlying estimation algorithm[11].

A potential concern is in areas where there are very few observations, and the estimation algorithm is attempting to reconstruct from a very limited number of observations. Areas on the edges of surveys tend to be made up of observations from the outer edges of a MBES swath, and therefore tend to have higher levels of noise and greater potential for being blunders. In this case, the depth estimation may have only one or two valid estimates of depth to use, and in such circumstances it might be impossible to reliably estimate the uncertainty statistics. In the past, some approaches to this used a combination of the statistics and source uncertainty to stabilize the estimate, but more recent practice has demonstrated that these estimates are not reliable, and make the estimation process more complex than required. The simpler solution is to return to the basic survey maxim that if you really wanted to know what the depth was there, you should have collected more data. In practice, just dropping the edges of the survey is much more effective.

Dynamic seafloors on a scale smaller than the capture radius of a group of observations may also result in a limited number of observations to support a variance statistic. An underlying assumption of the averaging of many observations is that the observations can be considered representations of the same object. Averaging bags of sugar with a few bags of flour thrown in does not improve the estimate of bags of sugar. Robust estimators, such as CUBE and CHRT, can compensate by separating observations into distinct hypotheses (assuming an appropriate apriori system uncertainty model), but this results in fewer estimates contributing to the selected estimate.

Even if all of the limited data is valid, the statistical summary might be a significant underestimate of the uncertainty value that would be seen in more dense data of the same type. This is a consequence of a statistical effect where most data tends to come from near the mean value and therefore tends to cluster, reducing the estimated uncertainty, until you have a sufficient number of samples to work from (typically order 30-100 or so). This is usually called a small-sample statistic. Although there are statistical methods to compensate for this under-estimation, in practice the simplest solution is not to work in areas of sparse data. Since the Coast Survey spatial density requirements often require observations counts on the order of 5-10, however, this may not be an option. Three alternatives are possible. First, the problem can

---

[11] If a poorly defined uncertainty model (i.e. TPU) were used, it may cause CUBE/CHRT to over produce hypotheses. This may cause the resulting grid to bounce between inconsistent hypotheses in adjacent nodes and report a lower-than-it-should-be uncertainty value, as fewer unique soundings contributed to the variance of the selected hypothesis.

just be ignored. This will result in under-estimates of uncertainty, on order 30% at N=5, dropping to 13% at N=10, and <5% at N=30. Since the uncertainty is simply an indicator of performance, these values may be acceptable, so long as the method is clearly documented. Second, the estimation algorithm could record the number of observations used for the statistic computation, which can be used to compensate. This would result in extra data that needed to be recorded, but would allow for better post-processing capability. Finally, the estimation algorithm could compensate directly for the small sample effects when reporting the uncertainty. This would preclude reprocessing the uncertainty later, but would obviate the need to record extra data.

The compensation for small-sample effects is closely related to the way that the uncertainty is reported to the end user (or recorded in an output, e.g., a BAG file). The most fundamental uncertainty statistics is the variance of the observations: technically the second central moment of the observations. In practice, this is computed by estimating the mean of the observations, and then estimating the average of the squared difference between the observations and the mean. Since the differences are squared, however, the units of variance as squared meters, and since the differences are often small, the squared differences are often very small (e.g., <0.01 m^2) which can make them hard to interpret.

Most often, therefore, the positive square root of the variance - the standard deviation - is used. This has units of meters, and therefore has approximately the same scale as the differences from the mean, and can be much more readily interpreted. This does not, however, reflect the "fuzziness" that the user would see in a point cloud representation of the data. Under some greatly simplified[12] assumptions, the distribution of the data can be modeled as a Gaussian, in which case the "fuzziness" is approximately a range of four standard deviations, or approximately 99.99% of the data, centered on the mean depth[13]. A natural way of reporting the uncertainty, therefore, is to scale the standard deviation by a known value to provide what is known as an expanded uncertainty. The choice of the scale factor is essentially arbitrary so long as it is documented, but a conventional approach is to multiply by 1.96 which - if the data is really Gaussian - provides a range that covers 95% of the expected observations. This provides a reasonable estimate of coverage that reflects what the user would see in the point cloud without over-expanding the uncertainty and limiting the value of the data artificially.

As is evident from the preceding, there are many different approaches that could be taken to estimate and report the uncertainty. In a general sense, any of the methods could be used, so long as the specifics of the method are sufficiently well documented in the metadata for the dataset. There is an advantage, however, in using a method that is easily documented and efficiently computed, and which reflects what the user would intuitively see in the data.

---

[12] While there are many variables that *could* be taken into account to get a more accurate result, this simplification provides the necessary resolution for our purposes without requiring additional complex math

[13] The full range is closer to six standard deviations, but humans are not good at estimating this visually.

NOAA’s MBES Uncertainty Calculation

Taking into consideration the aforementioned options for calculating uncertainty, the following approach has been identified for all Coast Survey commissioned surveys:

1) Bathymetry estimates must be computed using CUBE or CHRT

2) For each depth estimate reported, compute and report the variance (σ^2) of <u>only the observations used to generate the depth estimate</u>

    a) This will exclude blunders or outliers filtered out by the CUBE or CHRT algorithm

3) For each depth estimate reported, compute and report the expanded uncertainty, which is calculated as the positive square root of the variance, multiplied by a scale factor of 1.96 (Equation: $\sqrt{\sigma^2} * 1.96$)

    a) Scale factor of 1.96 represents 95% Confidence Interval

    b) The use of expanded uncertainty alleviates the need to utilize the computationally intensive small sample compensation processes, given the limited additional benefit that process would provide

    c) This will be the published uncertainty value populated as “noaaProduct_2024” for each associated depth estimate

4) The following grid metadata must be recorded in the BAG file

    a) Choice of algorithm (i.e. CUBE/CHRT)

    b) Value of the scale factor used to compute uncertainty

5) In addition, the following node specific metadata must be recorded in the BAG file for each depth value

    a) Number of observations used to estimate the resulting depth and uncertainty as “*number of soundings*”.

Conclusion

Revisiting the intent and background of this paper, there are three major goals that stand out; 1) Define a method that is easily documented and efficiently computed, 2) Represent the data in a manner that matches our intuitive understanding of the data when viewing graphically, and 3) Provide necessary context to Coast Survey's hydrographers to better understand this critical concept of uncertainty. These goals have been accomplished via a simple approach that relies on the basic principles of probability and statistics. While this paper alludes to opportunities for increased accuracies of uncertainty estimates (i.e. through integration of small sample correction equations), the benefit of these potential "improvements" is far outweighed by the additional computational resources (and resulting processing time) they would require. The simple method identified in this paper meets Coast Survey's needs, and does so in a computationally efficient manner.

Coast Survey will continue to monitor the effectiveness of this approach moving forward as hydrography continues to evolve. While future advancements may necessitate another revision, the path defined in this paper charts a clear course for the future, providing the necessary consistency and usability for Coast Survey's public products.