3-2013

Distribution-free, Variable Resolution Depth Estimation with Composite Uncertainty

Brian R. Calder
University of New Hampshire, Durham, brian.calder@unh.edu

Distribution-Free, Variable Resolution Depth Estimation with Composite Uncertainty

B. R. Calder

***Abstract*—Recent algorithms for processing hydrographic data have treated the problem of achievable resolution by constructing grids of fixed resolution, a composite grid of variable resolution, recursive sub-division in a quad-tree, or by relying on a comprehensive TIN of the original points. These algorithms all impose some artificial structure on the data to allow for efficient computation, however, which this paper attempts to address.**

A scheme is outlined which provides a robust estimate of depth and associated uncertainty that makes as few assumptions as possible. Using a non-uniform spectral analysis, it estimates the spatial scales at which the data are consistent so it can estimate within the Nyquist limit for the underlying surface. Kernel density techniques estimate the most likely depth, and density partitioning estimates the observational and modeling uncertainty. After correcting for potential biases the uncertainty is augmented using a modified Hare-Godin-Mayer system integration uncertainty and a sound speed profile variability due to Beaudoin et al. The result is a robust, distribution-free, continuously variable-resolution estimate of depth with an associated uncertainty.

This algorithm is illustrated by estimating the depth (and uncertainty) of Challenger Deep, and the paper then provides some perspectives on efficiency, extensibility and adaptability of this algorithm in the hydrographic context.

***Index Terms*—Bathymetric Estimation, Variable Resolution, Data Density Estimation, Data-driven Estimation, Distribution-Free Depth Estimation, Lomb-Scargle Periodogram, Bias Estimation**

I. Introduction

As a general principle, the depth estimation problem can be stated as: obtain the best estimate of depth possible given the observations in an area, plus its uncertainty, but do so in a manner than makes as few assumptions about the data as possible. Most of the research on this area in the last decade has revolved around computer-assisted methods of estimating plausible depths from raw data, with the goal of doing the majority of the processing required without user interaction, while maintaining operator interaction to assess the results of the algorithm, typically on the small subset of data that really needs attention. All of the methods that have been developed, however, make numerous assumptions about the data's properties, and induce a strong structure in the data that the data itself does not have. For example, the CUBE algorithm [1] assumes that the data is locally constant and independent, and, at least in most implementations, that a locally uniform grid of depth estimates may be used to represent the depths. None of these are strictly true. The CHRT algorithm [2] relaxes the constant grid requirement in favor of a piecewise defined composite grid, but still assumes independence and a locally constant data form. The CHARM algorithm [3] assumes that the data points can be modeled in a given area by a low order polynomial surface (which avoids a local constancy assumption) but still assumes independence (and uniform uncertainty), and induces a quad-tree structure [4] over the data. The method of Arge *et al.*[5] avoids the use of a gridded structure by constructing a segmented TIN of all of the data, but makes assumptions about data properties in order to segment the TINs being constructed. Other methods, such as kriging [6] have been attempted, which require strong statistical assumptions about the second order properties of the data [7].

For those methods that estimate an uncertainty associated with the declaration of depth, there is an additional difficulty in which uncertainty to report. The most common method is to report some estimate of the sample variance of the data used in the estimate (in the reference implementation of CUBE, for example, the uncertainty reported is a sequential estimate of standard deviation derived from the observations used to construct the depth hypothesis being reported; definitions vary between implementations). This method reflects the apparent 'noise' in the observations that a user would estimate by visual inspection and is easy to compute and explain, but more correctly reflects the repeatability of measurements than their uncertainty. (This is, of course, what is required for processing at this level, although perhaps not what the end-user expects to see.) What uncertainty should be reported, however, is a matter of the question that the end-user is likely to ask. For example, in the case of an interpretation of the data for geological purposes, or for initial processing, a measure of relative uncertainty (i.e., the repeatability) is acceptable. Any common-mode uncertainty, such as a datum uncertainty, would not be reflected in such a value, however, possibly leading to an under-estimate of the potential uncertainty being reported. If the end-user is more interested in the absolute uncertainty (i.e., the probability of being within a given vertical distance of the unique true depth in the area, as measured by an independent instrument), then an uncertainty value that reflected all final vertical correctors that are not averaged down by repeated measures might be more appropriate.

The influence of the data processing method on the depth and uncertainty declared is not well studied. As an obvious example, consider the case of 'perfect' data (i.e., data without outliers or systematic blunders) over which a depth estimate must be computed. Assuming that some region about any given point in the area of interest will have to be considered for source observations with which to construct an estimate

B. Calder is with the Center for Coastal and Ocean Mapping and NOAA-UNH Joint Hydrographic Center, University of New Hampshire. Address: Chase Ocean Engineering Lab, 24 Colovos Road, Durham NH 03824. E-Mail: brc@ccom.unh.edu. Phone: +1-603-862-0526.

of depth (whether that is by selecting an observation, or computing a suitably weighted average of the observations), it is clear that the choice of the region, or even simply the size of the region, could have a considerable effect on the depth estimate constructed. With more complex algorithms it is less obvious what the effects are likely to be, but the fact of a modeling uncertainty remains: the choice of algorithm, and how well it conforms to the underlying truth of the bathymetric structure of the real world, is important.

This paper proposes a method to estimate the depth of the ocean which makes as few assumptions as possible about the data, and that allows the expression of both a relative and absolute uncertainty. The key idea is to allow the data to define the local structure at which to process by computing an estimate of the spatial spectral density of the data (from the raw observations) in the region of analysis, and using it to determine at what resolution the estimate can be constructed. This avoids having to estimate a resolution *a priori*, and makes the representation of continuously variable resolution, and data adaptive. The method then uses a non-parametric estimate of the distribution of depths in the data-adaptive region to estimate the most likely depth (hence avoiding outliers), and an analysis of the modes of the distribution to determine the combined observational and modeling uncertainty for the data used. Finally, the method uses a modified version of the Hare-Godin-Mayer uncertainty model [8] to provide an estimate of the systemic uncertainty associated with the observations used to construct the depth estimate (the model is adjusted to avoid the effects of observational uncertainty, and that of sound speed variability in the water column), and an estimate of the sound speed variability-induced uncertainty in the area using the method of Beaudoin *et al.* [9] to complete the uncertainty estimate. Constructed in this manner, the analysis provides a data-adaptive estimate of depth, the relative uncertainty due to the observables, and an estimate of the absolute uncertainty of the depth reported, corrected for potential bias effects and taking into account the distribution of the observations given the limited number that are typically available in data-adaptive approaches.

The paper is organized as follows. The methods used in the implementation of the proposed algorithm are outlined in section II. An example of the use of the method to determine the depth and uncertainty of the deepest part of the world ocean in the Challenger Deep section of the Marianas Trench is given in section III. The algorithm has a number of advantages, but some concomitant difficulties; these are outlined in the discussion in section IV. Finally, we attempt to set the algorithm and its potential in perspective in the conclusions of section V.

II. Methods

A. Resolution Determination

Determining the resolution at which to work is a key question in any depth estimation algorithm. Previous approaches have included implicit estimation based on data density [2], [10], and observation of model fit degeneration [3]. Most fundamentally, however, the resolution at which the depths should be estimated is that at which all observable features of the spatial structure of the region can be represented. In particular, the method should sample at the Nyquist frequency for the observations so as to avoid aliasing; the difficulty, of course, is that the spatial spectrum of the area is unknown before processing starts. (Note that this is the Nyquist frequency for the observations, not for the seafloor itself. The Nyquist rate for the observations can be significantly lower than that for the seafloor due to the limited angular resolution of the observing instrument, along-track data density, etc.)

The spatial spectrum is readily estimated using Fourier techniques given a regularly spaced sample set, but in the case of raw data the samples are semi-randomly distributed (i.e., according to the acquisition geometry) and any attempt to generate a regular set, e.g., by interpolation or gridding, would result in an induced structure. The method here therefore uses the Lomb-Scargle periodogram [11], which is a particular example of the more general non-uniform Fourier analysis techniques [12]–[14], which are also sometimes described as Least-Squares Spectra [15]. In one dimension, the Lomb-Scargle periodogram is defined as

$$P_z(\omega) = \frac{1}{2}\left\{ \frac{\left[\sum_j z_j \cos(\omega(x_j-\chi))\right]^2}{\sum_j \cos^2(\omega(x_j-\chi))} + \frac{\left[\sum_j z_j \sin(\omega(x_j-\chi))\right]^2}{\sum_j \sin^2(\omega(x_j-\chi))} \right\} \quad (1)$$

where the inputs are samples at (x_j, z_j) and the position parameter χ is defined as

$$\tan(2\omega\chi) = \frac{\sum_j \sin(2\omega x_j)}{\sum_j \cos(2\omega x_j)}. \quad (2)$$

This formulation is different from the conventional non-uniformly sampled periodogram,

$$P_z(\omega) = \frac{1}{N}\left[\left(\sum_j z_j \cos(\omega x_j)\right)^2 + \left(\sum_j z_j \sin(\omega x_j)\right)^2\right] \quad (3)$$

but is equivalent to the conventional periodogram if the sampling is uniform, has simpler statistical properties, and is equivalent to the least-squares solution of sinusoid fitting to data.

The interpretation of the Lomb-Scargle periodogram is the same as for a conventional periodogram, except that there is no strong definition for the highest frequency which can be represented. (This is due to the non-uniform nature of the input points.) Here, the periodogram is constructed at the analysis point where the depth is required, and an assessment is made as to the maximum frequency with significant energy as a means to determine the sampling rate (or analysis region size) at which to work for the remainder of the estimation procedure. In practice, although it would be possible to construct a two dimensional periodogram to assist in determining the maximum significant frequency, the algorithm assumes that the

spectrum is locally isotropic, and therefore that constructing a spectrum in the cardinal directions is sufficient.

The only implementation difficulty is that the computation of (1) can be quite slow since it cannot benefit from the normal Fast Fourier Transform (FFT). Fast algorithms for computing non-uniform Fourier transforms exist, however, and the method here uses the code from [16], which relies on a conventional FFT library from [17].

B. Optimal Depth Estimation

As described above, the method by which the observations are combined in order to construct the depth estimate can have a significant effect on the depth estimate obtained. Under the assumption that the size of the analysis region is adapted to the spectral structure of the data at the point of interest (i.e., according to the spectral analysis described previously), however, then the data should show (absent any outliers) a consistent distribution about the depth of interest. In this case, a kernel density estimate of the probability density function of depth may be constructed. For the set of data points $\mathcal{X} = \{\vec{x}_j\} = \{(\vec{p}_j, z_j)\}$, the kernel density estimate is

$$p(z; \vec{s}, r) = \frac{1}{|\mathsf{N}(\vec{s}, r)|} \sum_{\vec{x} \in \mathsf{N}(\vec{s}, r)} K(z; (\vec{x})_z, u) \tag{4}$$

where

$$K(z; z_0, u) = \frac{1}{\sqrt{2\pi u^2}} \exp\left\{ -\frac{(z - z_0)^2}{2u^2} \right\} \tag{5}$$

is the estimation kernel, and $\mathsf{N}(\vec{s}, r)$ is the set of observations in the neighborhood of estimation point $\vec{s} \in \mathbb{R}^2$, which is taken to be a closed ball $B(\vec{s}, r), r > 0$ of radius r about $\vec{s}$. The kernel smoothing bandwidth u is chosen as the optimal value with respect to asymptotic square integrated mean error (ASIME, [18]) under the assumption that the distribution of data is Gaussian. Note, however, that there is no requirement that the distribution is Gaussian (this only affects the claim to optimality of the bandwidth estimation), and in practice it is expected that the distribution will be distinctly non-Gaussian, and possibly multi-modal, due to the presence of outliers in the dataset and the potential for the bathymetry to have some features that are of higher frequencies than might be expected from the periodogram.

The estimate $p(z; \vec{s}, r)$ is a potentially multi-modal distribution. A mode is readily identifiable as a zero of the first differential of the estimate, so long as the zero is concave down, as determined by the second derivative. Concave up zeros are valleys between the modes, and therefore an analysis of the first and second derivatives of the estimate allows the algorithm to identify the position of all of the modes, and the boundaries between them. Let Z be the set of depths associated with the modes, let the domain of support of the distribution be $\mathbb{D} = [\min_j\{z_j\} - 3u, \max_j\{z_j\} + 3u]$, and let V be the set of depths associated with the valleys between the modes (i.e., $v_i \in \mathsf{V}$ is the valley between mode i and mode $i+1$, appending $v_0 = \inf \mathbb{D}$ and $v_\infty = \sup \mathbb{D}$ to provide outer limit markers for the domain of definition). Then, the domain can be partitioned into non-overlapping regions $D(z_i) = [v_{i-1}, v_i), 1 \leq i \leq |\mathsf{Z}|$ so that $\bigcup_{z_i \in \mathsf{Z}} D(z_i) = \mathbb{D}$, and the most likely mode can be selected simply as

$$j = \arg\max_i p(z_i; \vec{s}, r), \quad 1 \leq i \leq |\mathsf{Z}|. \tag{6}$$

C. Uncertainty Estimation

1) Observation Uncertainty: With the definitions above, the observations associated with the i^{th} mode as simply the set $\mathsf{M}_i = \{\vec{x} \in \mathsf{N}(\vec{s}, r) : (\vec{x})_z \in D(z_i)\}$, and therefore the variability of the observations that support the mode is simply the sample variance about the mode value,

$$v_i := \frac{1}{|\mathsf{M}|_i - 1} \sum_{\vec{x} \in \mathsf{M}_i} \left((\vec{x})_z - z_i\right)^2 \tag{7}$$

.

This value takes into account the fundamental observational uncertainty of the data, and any modeling uncertainty induced by local slopes, etc.

2) Bias Reduction: In the presence of biases, the sample estimate of variance can over-estimate the magnitude of the variability. For example, if there are two or more passes in an area, a small vertical offset between them due to difficulties in establishing a water-level model would result in an increased estimate of variability. In general, biases are difficult to assess, since there is no means to know in which direction the offset was applied. It is possible, however, to estimate the magnitude of the variability increase, at least in simple cases. Consider the simplest case of two passes over the data, and for simplicity of expression, consider observations resolved into the coordinate frame of the echosounder. A plausible model for the depth observed in beam n is

$$z(n) = z_0 + \rho(z_0, \theta_i, \vec{c}) + \epsilon(z_0, \theta_n) \tag{8}$$

where z_0 is the true depth, θ_i is the mean beam angle relative to the swath at the analysis site, $\vec{c}$ is the vector of sound speeds in the water column, θ_n is the beam angle associated with the observation, and $\epsilon(z_0, \theta_n) \sim \mathcal{N}(0, \sigma^2(z_0, \theta_n))$ is the stochastic component representing measurement uncertainty of the sonar. For two passes over the region of interest, the observations will contain a mixture of points from each pass, with different weights, and therefore the probability density function, estimated through the kernel density, will be a mixture distribution, $p(z) = w_1 p_1(z) + w_2 p_2(z)$. Ignoring outliers, the expected depth is simply

$$\begin{aligned} \mathbb{E}[z] &= w_1 \mathbb{E}[z_1] + w_2 \mathbb{E}[z_2] && (9)\\ &= z_0 + w_1 \delta\rho(\theta_1, \theta_2) + \rho(z_0, \theta_2, \vec{c}) && (10)\\ &= z_0 - w_2 \delta\rho(\theta_1, \theta_2) + \rho(z_0, \theta_1, \vec{c}) && (11) \end{aligned}$$

(where $\delta\rho(\theta_1, \theta_2) := \rho(z_0, \theta_1, \vec{c}) - \rho(z_0, \theta_2, \vec{c})$) with corresponding variance

$$\begin{aligned} \mathbb{V}[z] &= \mathbb{E}[(z - \mu)^2] \\ &= w_1 \sigma_1^2 + (1 - w_1)\sigma_2^2 \\ &\quad + w_1(1 - w_1)\delta\rho^2(\theta_1, \theta_2) \end{aligned} \tag{12}$$

so that the weighted blend of variance, and a truncated version of the bias, may be constructed. (This may be estimated

through the sample variance v_i for each mode where a bias might occur.) Normally, separating out the terms of this expression would be impossible since the membership of the observations is typically not know. Here, however, all of the observations are labelled with their acquisition line number, and therefore it is clear that $\mathbb{E}[\delta\rho(\theta_1,\theta_2)] = \mathbb{E}[\rho(z_0,\theta_1,\vec{c})] - \mathbb{E}[\rho(z_0,\theta_2,\vec{c})] = m_1 - m_2$ (where $m_i = N_i^{-1}\sum_j (z_i(n))_j$ is the sample mean of the soundings from line i associated with the mode in question), with weights $w_i := N_i/\sum_j N_j$. Since σ_1^2 and σ_2^2 can be computed separately, an estimate of $\delta\rho^2(\theta_1,\theta_2)$ can be extracted, given a sample estimate of the overall variance of the weighted dataset, and the sample variance adjusted accordingly.

3) Systemic Uncertainty: The uncertainty associated with a depth estimate depends strongly on the survey system in addition to the measurement uncertainty of the echosounder itself. Unknown variability in the patch test alignment estimates, measurement uncertainty in the horizontal and vertical offsets between the components of the system, etc., cause extra uncertainties when propagated into the position and depth estimates of each observation. The most commonly used model for this is due to Hare *et al.*[8], [19], which is adopted here. However, since the algorithm estimates the measurement uncertainty of the observations (i.e., the uncertainty due to the echosounder itself) and that due to sound speed variability (see below) separately, the model was adjusted to ignore those components. In implementation, the algorithm is run for the known configuration of the survey system (as recorded in the survey metadata) at the most likely mode depth, z_i, and the results corresponding to the most common beam in M_i are selected to provide the vertical uncertainty due to the survey system integration.

4) Refraction Uncertainty: Possibly one of the most significant uncertainties in the absolute uncertainty of a depth estimate is that induced by unobserved spatio-temporal variability of the sound speed field in the ocean. This is typically significantly undersampled in space and time (although very accurate samples can be made at a particular point), and therefore the corrections used for any given observation are typically at the wrong place, at the wrong time. Here, the algorithm uses the method of Beaudoin *et al.* [9] to estimate the variability that might be induced for a given depth and observation angle (relative to the echosounder), given the observations of sound speed profile that were made during the survey (again, as recorded in the survey metadata). In implementation, the analysis is done for all depths and angles, and an 'uncertainty wedge' is generated. Once the depth estimate is established at any analysis point, a simple lookup is conducted, using the angle corresponding to the most common beam in M_i, to establish the refraction uncertainty.

5) Combined and Expanded Uncertainty: The algorithm assumes that the uncertainties due to measurement uncertainty, systemic uncertainty and refraction uncertainty are essentially independent, and therefore the combined uncertainty of the depth estimate is simply

$$\sigma_z^2 = \sigma_m^2 + \sigma_s^2 + \sigma_r^2 \tag{13}$$

where σ_m^2 is the measurement uncertainty, σ_s^2 is the systemic uncertainty, and σ_r^2 is the refraction uncertainty. Since setting the analysis radius by the methods described here often means that few samples are present from which to estimate the depth, it cannot be assumed that a sufficient number of samples are present to make the Gaussian approximation for confidence intervals associated with the depth estimate. The measurement uncertainty has $\nu_m = |\mathsf{M}_i| - 1$ degrees of freedom, and we assume that both the systemic uncertainty and refraction uncertainty are effectively of infinite degrees of freedom. The adjusted degrees of freedom for the combined uncertainty may therefore be computed using the Welch-Satterthwaite formula [20] as

$$\nu_z = \nu_m \left(\frac{\sigma_z^2}{\sigma_m^2}\right)^2 \tag{14}$$

and the expanded uncertainty for a 95% CI is therefore

$$U_z = T(0.05,\nu_z)\sqrt{\sigma_z^2} \tag{15}$$

where $T(\alpha,\nu_z)$ is the critical value function for Student's T-distribution.

III. Example

A. Background and Data

The data used in this example was collected as part of the U.S. Extended Continental Shelf project [21], and in particular to map the Mariana Trench and the surrounding area [22], [23]. The data collection, conducted in two phases using the *USNS Sumner*, was carried out in 2010 using a $1^\circ \times 1^\circ$ Kongsberg EM122 multibeam echosounder, operating in dual-ping mode so as to generate 864 potential observations in two pings per sounding cycle. The remainder of the survey system consisted of an Applanix POS/MV 320 V.4 with global differential correctors; sound speed profiles were provided by expendable bathythermograph probes launched at least every six hours, and more often if required, during the survey.

Coincidentally to the primary mapping mission, the second leg of the cruise took the ship over what is generally accepted to be the deepest part of the world ocean, the Challenger Deep. For no other reason than that it is believed to be the deepest part of the ocean, this area has been visited many times [24], with the primary question being just how deep the area is. The data from this second cruise is used here to illustrate the proposed algorithm.

The data in the vicinity of Challenger Deep consists of 2,051,371 observations from eight survey lines. The data were converted from raw Kongsberg datagrams into CARIS/HIPS, clipped to an approximate bounding box of $[11.2606, 11.4453]^\circ\text{N} \times [141.9761, 142.8753]^\circ\text{E}$, and then output as a simple text file for further processing. Position, depth, source line, ping and beam identifiers were retained. The data were then projected to Universal Transverse Mercator projection (zone 54N) with respect to WGS-84, loaded into MATLAB and processed with custom scripts.

B. Configuration

The survey system recorded offsets from the motion sensor to the transducer of $(18.34, 0.58, -4.08)$ m, and patch-test offsets of $(0.39, 0.95, -0.15)^\circ$ in roll, pitch and heading. We estimated the uncertainty of the offsets at 5 mm, the positioning accuracy at 2 m, the latency of the attitude and positioning information at 5 ms, the measurement uncertainty of attitude at 0.02°, heave at 0.05 m or 5%, and vertical offsets at 0.1 m for draft, dynamic draft and loading. (All values are standard deviations except positioning, which is $1\mathrm{d}_{\mathrm{rms}}$).

In order to speed up processing, the algorithm was first applied in a coarse search with 500 m resolution and an analysis radius of 500 m. This identified a most likely location of the deepest region around $[11.356760, 11.303109]^\circ\mathrm{N} \times [142.133747, 142.280569]^\circ\mathrm{E}$, at which point the proposed algorithm was used to estimate the resolution, depth and uncertainty as described below.

C. Resolution Estimation

A spectral analysis was conducted in the east-west direction about the middle of the focussed region, Figure 1, and about the north-south, Figure 2. In both cases there is clear evidence of very long wavelengths (second panel in each), but little evidence of anything except noise below approximately 500 m wavelength. This suggests that wavelengths under 500 m correspond to measurement uncertainty, and the analysis region radius was set to 250 m.

D. Observed Data

The remainder of the algorithm was then conducted for the region of interest with an analysis radius of 250 m. The results, Figure 3, show that there is a relatively even distribution of data in the region (as estimated through the number of degrees of freedom in the estimates of depth reported), that the uncertainty is typically on the order of 25 m (95%), and that it is not adversely affected by the bathymetry.

The array of depth estimates in the focussed region were used to find the location of the deepest estimate, for which the details of the algorithm are shown in Figure 4. Here, there are ten observations from two lines forming the sample, which is (just) unimodal but strongly asymmetric (top left panel); the kernel bandwidth was estimated as 19.1 m. The algorithm correctly detected the primary peak as the mode, making the deepest depth estimate 10,984 m, and estimating the uncertainty at 25 m (95%) on 9 d.f., including a systemic uncertainty of 4.2 m (95%) estimated for beam 250 (the most common beam in the estimates), and a refraction uncertainty of 3.1 m (95%) based on the same beam angle. The position was estimated from the centroid of the observations at (623,875.0, 1,252,789.5) m, which corresponds to $11.329903^\circ\mathrm{N}$, $142.199305^\circ\mathrm{E}$.

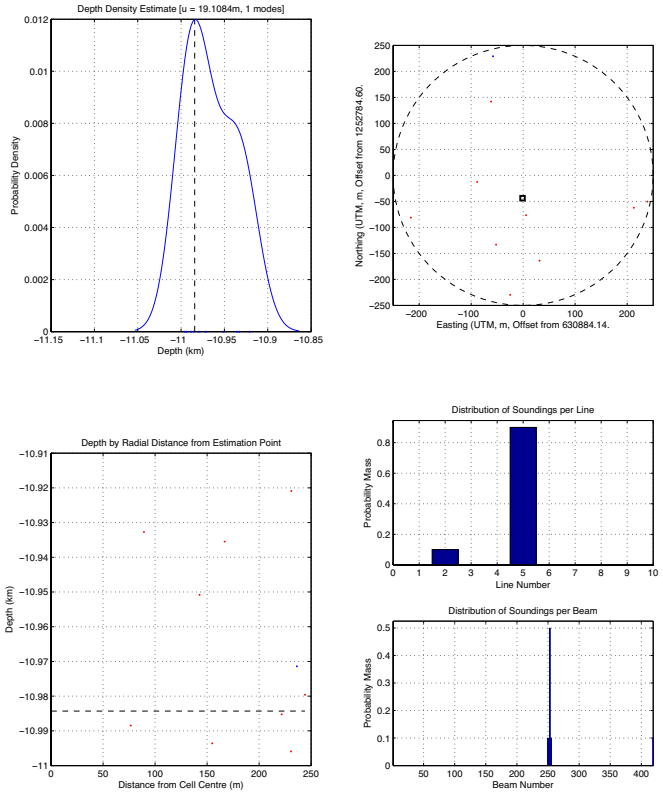

Fig. 4. Details of the algorithm's estimation process at the location of the deepest depth in the Challenger Deep. The kernel density estimate (top left) has a bandwidth of 19.1 m, which the single mode is supported by ten samples (9 d.f.) from two lines. Note the distribution of observations (top right panel), color coded by line, and the estimated position (black square) derived from the centroids.

IV. Discussion

The method proposed has several advantages over the available alternatives. First, it makes fewer assumptions about the statistical properties of the data, and adapts to the data resolution that is achievable in practice. This makes it less likely to induce structure in the data, and less likely to generate misleading results due to weakly supported assumptions such as independence and Gaussian distributions. Second, it will provide estimates of depth that are as high a resolution as can be supported by the data, which is always preferred. Third, the algorithm is commutative, in the sense that the ordering of the data does not affect the estimation process. This makes the algorithm significantly easier to use, and makes it more robust against outliers (since they are always taken in the context of all of the data, rather than being dealt with on limited information). Finally, the uncertainty model presented here avoids any theoretical model of the echosounder measurement uncertainty, and correctly represents the absolute uncertainty of the depth estimate provided, taking into account the refraction and systemic uncertainties that are not removed by the presence of more data. (The relative uncertainty is still available, of course.) This makes the uncertainty reported a much more realistic answer to the typical end-user question of the chances of a repeated measurement that they might make falling within the bounds predicted by the estimate.

There are concomitant disadvantages. Generating estimates that are as high resolution as possible automatically implies that the algorithm will have to work with limited amounts

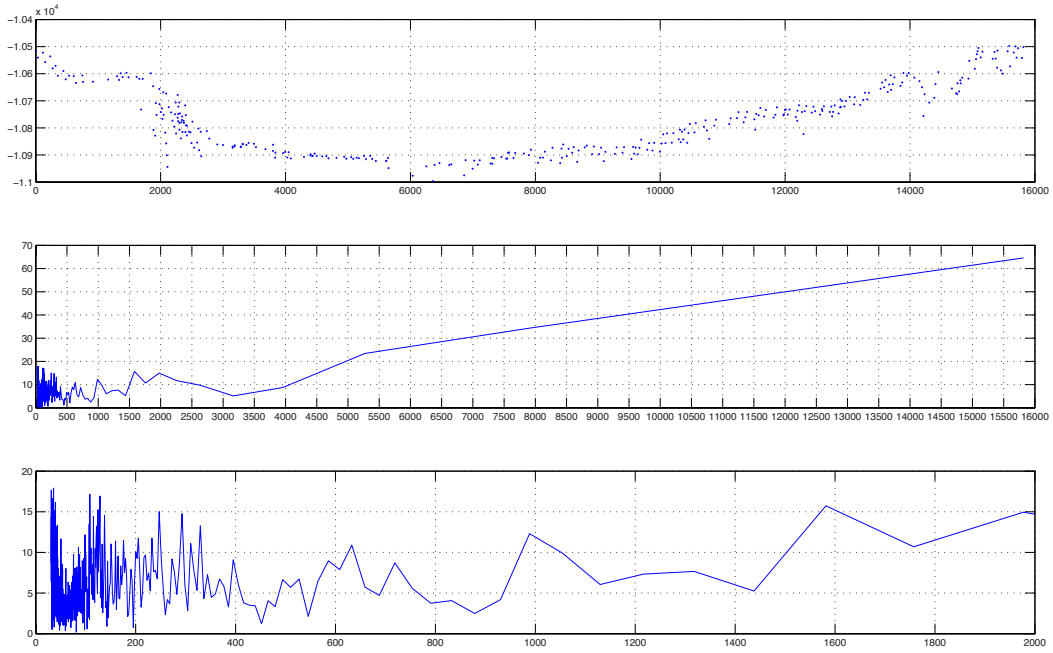

Fig. 1. Spatial spectrum in the east-west direction about the focussed area of analysis. Top panel: depth profile of the data as a function of along-transect distance (m); middle panel: un-normalized spectrum magnitude as a function of wavelength (m); bottom panel: detail about the short wavelength part of the middle panel.

of data. This means that the algorithm has to be much more careful to track the degrees of freedom associated with the estimates, for example, and must also use the appropriate small-sample distributions when computing expanded uncertainties. In a certain sense it might also be considered to make the algorithm slightly less robust than other alternatives, since estimates must be constructed from smaller amounts of data and are therefore more susceptible to outliers. Although a formal comparison has not been done in this instance, it is reasonable to assume that the reduction in robustness is not likely to be significant, however: most other techniques, presented with a similar lack of data, would likely misbehave also.

Secondly, the algorithm is inherently a post-processing methodology. That is, the algorithm requires all of the data to be available before the processing can begin, and therefore cannot be used as described to process data as it is being collected. It would of course be possible to simply recompute all of the algorithm steps when new data is added, but this would be very inefficient. It is possible that at least the kernel density estimate could be computed incrementally, at least approximately, and the algorithm could be adapted to either assume an analysis resolution or to update this on a regular basis as more data was being collected. The practicalities of these modifications are, however, unknown.

Finally, the spectral estimation component of the algorithm is relatively slow to compute. Although there are a number of fast algorithms to assist in approximate solutions to the problem, the cost of doing a non-uniform transform is high, and doing so for each analysis point would be prohibitively so. It is possible that the algorithm could make some assumptions about how quickly the resolution is likely to change within an area, and do fewer analyses, but this would start to induce a structure in the data that is not supported. Since the algorithm does not require the whole spectrum, however, just the section where the energy is dropping off, it is possible that it might be adapted to compute only part of the spectrum, which would lead to significant speed improvements. This might lead to instabilities, however, if the resolution changed dramatically, and would require further investigation.

At present, the algorithm is at a prototype stage, and requires further investigation to determine how to adapt it to higher resolution, more dense, data. The advantages of the algorithm are compelling, however. For example, having an estimate of the probability density function at each location allows for the possibility of data fusion between locations in order to reliably identify outlier clusters, and propagate consistency information through the entire dataset. Analysis of the width of the primary model should allow for some estimate of modeling uncertainty to be made, and used to adapt the algorithm. How well these translate in practice, however, is a subject for much further study.

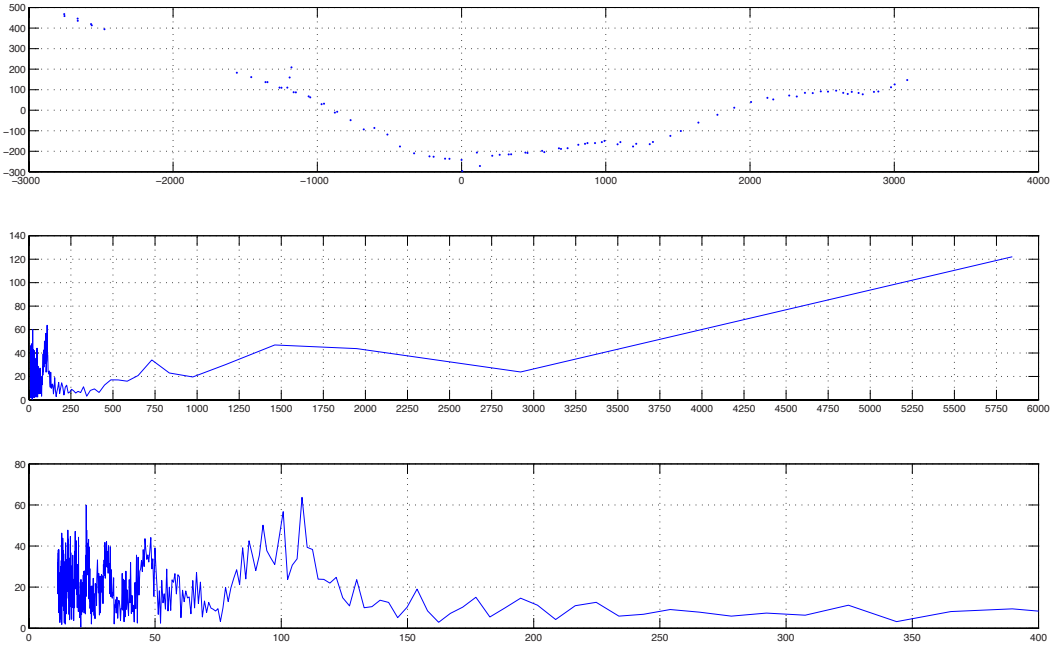

Fig. 2. Spatial spectrum in the north-south direction about the focussed area of analysis. Top panel: depth profile of the data as a function of along-transect distance (m); middle panel: un-normalized spectrum magnitude as a function of wavelength (m); bottom panel: detail about the short wavelength part of the middle panel.

V. Conclusions

This paper has outlined a method for estimating the depth of the seafloor from raw data points which avoids many of the artificial constraints employed by other methods of similar goal. In particular, it adapts continuously to the resolution that is best supported by the data, makes few assumptions about the statistical properties of the data, and provides a robust estimate of uncertainty that reflects a plausible real-world interpretation of what uncertainty should mean. The algorithm was illustrated on some deep-water multibeam echosounder data.

The algorithm has a number of advantages, particularly in its adaptability, but has limitations in that the core algorithm component to drive the adaptation is slow to compute. The extent to which this algorithm may be configured for use in denser data, therefore, is a subject of future research.

Acknowledgements

This work is supported by NOAA grant NA10NOS4000073. The author would also like to thank Dr. J. V. Gardner (CCOM) and Capt A. A. Armstrong, NOAA (ret.) (NOAA) for the data used in this paper, and Dr. J. Beaudoin (CCOM) for his assistance with processing the sound speed profiles used to provide refraction uncertainties in the example. The data used in this work was collected in support of the U.S. Extended Continental Shelf project; more details (and the data) can be found on CCOM's website, `http://www.ccom.unh.edu/theme/law-sea`.

References

[1] B. R. Calder and L. A. Mayer, "Automatic processing of high-rate, high-density multibeam echosounder data," *Geochem., Geophys. and Geosystems (G3) DID 10.1029/2002GC000486*, vol. 4, no. 6, 2003.

[2] B. R. Calder and G. Rice, "Design and implementation of an extensible variable resolution bathymetric estimator," in *Proc. US Hydro. Conf.* Hydro. Soc. Am., April 2011.

[3] N. Debese, R. Moitié, and N. Seube, "Multibeam echosounder data cleaning through a hierarchic adapative and robust local surfacing," *Computers and Geosciences*, vol. 46, pp. 330–339, September 2012.

[4] H. Samet, *Foundations of Multidimensional and Metric Data Structures*. San Fransisco, CA: Morgan Kaufmann, 2006.

[5] L. Arge, K. G. Larsen, T. Mølhave, and F. van Walderveen, "Cleaning massive sonar point clouds," in *Proc. ACM Int. Conf. on Advances in GIS*, A. el Abbadi, D. Agrawal, M. Mokbel, and P. Zhang, Eds., Assoc. Comp. Machinery. San Jose, CA, USA: Assoc. Comp. Machinery, November 2010, pp. 152–161.

[6] P. Bottelier, C. Briese, N. Hennis, R. Lindenberg, and N. Pfeifer, *Distinguishing Features from Outliers in Automatic Kriging-based Filtering of MBES Data: a Comparative Study*. Springer, 2005, pp. 403–414.

[7] N. A. C. Cressie, *Statistics for Spatial Data*. 900 p.: Wiley, 1993.

[8] R. Hare, "Error budget analysis for the Naval Hydrographic Office (NAVOCEANO) hydrographic survey systes," Univ. Southern Mississippi, 119 p., Tech. Rep., 2001.

[9] J. Beaudoin, B. Calder, J. Hiebert, and G. Imahori, "Estimation of sounding uncertainty from measurements of water mass variability," *Int. Hydro. Review*, no. 2, pp. 20–38, November 2009.

[10] G. Rice and B. Calder, "A quantitative approach to the resolution of bathymetric representation," in *Proc. US Hydro. Conf.* Norfolk, VA: The Hydrographic Society of America, May 2009.

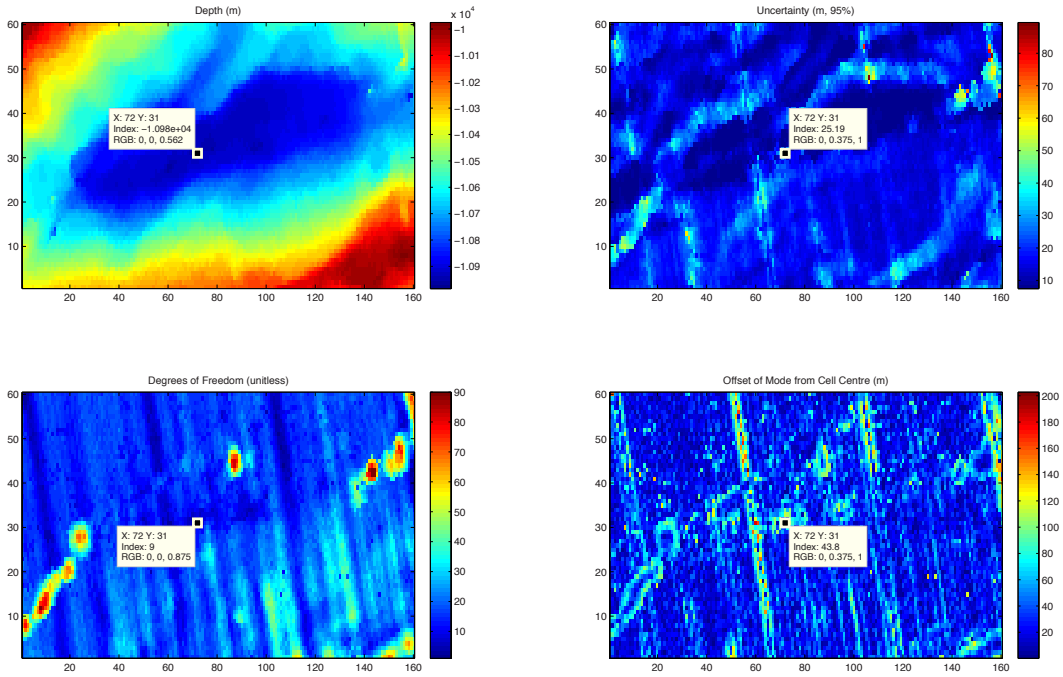

Fig. 3. Observed data around the deepest area of the world ocean, Challenger Deep. Top left panel: depth in metres; top right panel: uncertainty of depth in metres at 95% CI; bottom left panel: number of degrees of freedom in the estimate of depth; bottom right panel: offset of the centroid of depth from the nominal analysis location, in metres.

[11] J. D. Scargle, "Studies in astronomical time series analysis ii - statistical aspects of spectral analysis of unevenly spaced data," *Astrophysical J. (Part 1)*, vol. 263, pp. 835–853, 1982.

[12] A. Dutt and V. Rokhlin, "Fast fourier transforms for nonequispaced data, II," *Appl. and Comp. Harm. Anal.*, vol. 2, pp. 85–100, 1995.

[13] A. F. Ware, "Fast approximate fourier transforms for irregularly spaced data," *SIAM Review*, vol. 40, no. 4, pp. 838–856, December 1998.

[14] J. A. Fessler and B. P. Sutton, "Nonuniform fast fourier transforms using min-max interpolation," *IEEE Trans. Sig. Proc.*, vol. 51, no. 2, pp. 560–574, February 2003.

[15] M. R. Craymer, "The least squares spectrum, its inverse transform and autocorrelation function: Theory and some applications in geodesy," Ph.D. dissertation, U. Toronto, Dept. Civil Eng., 1998.

[16] J. Keiner, S. Kunis, and D. Potts, "Using NFFT3 - a software library for various nonequispaced fast fourier transforms," *ACM Trans. Math. Soft.*, vol. 36, no. 4, pp. 19:1–19:30, August 2009.

[17] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," *Proc. IEEE*, vol. 93, no. 2, pp. 216–231, 2005.

[18] D. W. Scott, *Multivariate Density Estimation*, ser. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley and Sons Inc, 1992.

[19] R. Hare, A. Godin, and L. A. Mayer, "Accuracy estimation of Canadian swath (multibeam) and sweep (multitransducer) sounding systems," Canadian Hydrographic Service, Tech. Rep., 1995.

[20] R. H. Dieck, *Measurement Uncertainty (Methods and Applications)*, 3rd ed. 252 p.: The Instrumentation, Systems and Automation Society, 2002.

[21] J. V. Gardner and A. A. Armstrong, "Mapping supports potential submission to u.n. law of the sea," *Eos, Trans. Am. Geophys. Union*, vol. 87, pp. 157–160, 2006.

[22] J. V. Gardner, "U.s. law of the sea cruises to map sections of the mariana trench and the eastern and southern insular margins of guam and the northern mariana islands," Univ. New Hampshire, CCOM/JHC, http://www.ccom.unh.edu/theme/law-sea, Tech. Rep. 10-003, 2010.

[23] A. A. Armstrong, "U.s. extended continental shelf cruise to map sections of the mariana trench and the eastern and southern insular margins of guam and the northern mariana islands," Univ. New Hampshire, CCOM/JHC, http://www.ccom.unh.edu/theme/law-sea, Tech. Rep. 11-002, 2011.

[24] J. V. Gardner, A. A. Armstrong, B. R. Calder, and J. Beaudoin, "So, how deep is the mariana trench?" *Submitted to Marine Geodesy*, 2013.