Marine Geodesy

ISSN: 0149-0419 (Print) 1521-060X (Online) Journal homepage: http://www.tandfonline.com/loi/umgd20

Development of an Uncertainty Propagation Equation for Scalar Fields

Brian Calder & Paul Elmore

To cite this article: Brian Calder & Paul Elmore (2017) Development of an Uncertainty Propagation Equation for Scalar Fields, Marine Geodesy, 40:5, 341-360, DOI: 10.1080/01490419.2017.1345811

To link to this article: https://doi.org/10.1080/01490419.2017.1345811

Accepted author version posted online: 05 Jul 2017.
Published online: 01 Aug 2017.

Submit your article to this journal

Article views: 34

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=umgd20

Download by: [University of New Hampshire] **Date:** 12 December 2017, At: 11:21

MARINE GEODESY
2017, VOL. 40, NO. 5, 341–360
https://doi.org/10.1080/01490419.2017.1345811

Taylor & Francis
Taylor & Francis Group

Check for updates

Development of an Uncertainty Propagation Equation for Scalar Fields

Brian Calder[a] and Paul Elmore[b]

[a]Jere A. Chase Ocean Engineering Laboratory, Center for Coastal & Ocean Mapping/Joint Hydrographic Center, University of New Hampshire, Durham, New Hampshire, USA; [b]Mapping Charting and Geodesy Branch, Naval Research Laboratory, John C. Stennis Space Center, Hancock County, Mississippi, USA

ABSTRACT

The uncertainty of a scalar field is essential structuring information for any estimation problem. Establishing the uncertainty in a dense gridded product from sparse or random uncertainty-attributed input data is not, however, routine. This manuscript develops an equation that propagates the uncertainty of individual observations, arbitrarily distributed in $\mathbb{R}^2$, to a common estimation location at which they can be used to determine the composite uncertainty of the output field. The equation includes the effect of the distance between the observation and estimation locations, the field and horizontal uncertainty of the observation, and user-parameters to control the expected variability in the field as a function of distance. Two computational versions of the equation, a lower cost conservative approach and a higher cost mean-distance approach, are developed and evaluated for computational cost and resulting accuracy in numerical experiments over simulated bathymetric data. The mean-distance approach is more accurate, but more costly; suitable numerical approximations are proposed to control computational costs. A benefit of the work described is flexibility and enhancement for applications of the model, such as the Combined Uncertainty and Bathymetry Estimator algorithm, which is used as a demonstration of the difference between the two versions of the equation.

ARTICLE HISTORY
Received 20 October 2016
Accepted 14 June 2017

KEYWORDS
DEM; mapping; hydrographic survey; ocean mapping; bathymetry; multibeam sonar

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

Introduction

Consider the following computational task: estimate output interpolation values from input points in $\mathbb{R}^2$ space. In polar coordinates, the interpolated value is $\widehat{z}_i = \mathcal{I}_i\{z_j(r_j, \theta_j)\}$. The operator $\mathcal{I}_i\{\cdots\}$ is an interpolation over the set of input points $\{z_j(r_j, \theta_j) : j \in \mathcal{N}(i)\} = [z_1(r_1, \theta_1), ..., z_J(r_J, \theta_J)]$, where $\mathcal{N}(i)$ is a local neighborhood, that yields $\widehat{z}_i$. Further, let there be azimuthal symmetry so that we can simplify $\widehat{z}_i = \mathcal{I}_j\{z_j(r_j)\}$. The following model of propagation of uncertainty from input data to an interpolated output point is based on that developed for the Combined Uncertainty and Bathymetry Estimator (CUBE) algorithm (Calder and Mayer 2003).

CONTACT Brian Calder brc@ccom.unh.edu Jere A. Chase Ocean Engineering Laboratory, Center for Coastal and Ocean Mapping/Joint Hydrographic Center, University of New Hampshire, 24 Colovos Road, Durham, NH 03824, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umgd.

© 2017 Taylor & Francis Group, LLC

The objective of the interpolation operator is to propagate uncertainty inherent in an observation from the location of the observation to the interpolation location. This process allows the inherent uncertainty of the observation to be used for estimating the composite uncertainty of the field being estimated. Uncertainty at the observation could be estimated through a formal propagation of measurement uncertainties, some defined uncertainty estimation protocol, or an empirical estimation process.

Let each measurement $\mathbf{m}_j$ have horizontal uncertainty $\sigma_{h,\,j}$ and vertical uncertainty $\sigma_{v,\,j}$. Calder and Mayer (2003) give a model of propagation of uncertainty, $(\sigma_{h,\,j}, \sigma_{v,\,j})$, from the jth input data point to the ith interpolated output point as

$$\sigma_{ij}^2 = \sigma_{v,j}^2 \left(1 + \left[\frac{r_{ij} + s_h \sigma_{h,j}}{\Delta_g}\right]^{\alpha}\right) \tag{1}$$

where Δ_g is the output grid spacing, and s_h and α are user-parameters that govern the confidence interval desired and growth rate for uncertainty. For the case of bathymetry with a sloped seafloor, Zambo et al. (2015) (q.v. Bourgeois et al. 2016) add an extra term to account for extra uncertainty when the seafloor has slope φ_i at the interpolation point

$$\sigma_{ij}{}^2 = \sigma_{v,j}^2 \left(1 + \left[\frac{r_{ij} + s_h \sigma_{h,j}}{\Delta_g}\right]^{\alpha}\right) + \sigma_{h,j}^2 \tan^2 \varphi i \tag{2}$$

(using the slope at the interpolation point provides smoothed uncertainty estimation for the slope contribution; see (Zambo et al. 2015; Bourgeois et al. 2016) for further discussion).

It is important to note that, while (1) and (2) were developed for bathymetry, both are applicable to scalar fields in general. For (2), generalization is achieved by replacing the second term with $|\nabla \widehat{z}_i|^2$ since $|\nabla \widehat{z}_i| = \sigma_{H_j} \tan \varphi i$ (Zambo et al. 2015).

Whatever the eventual use of the predicted uncertainty, the method by which it is generated should, for consistency, follow some general principles. This paper presents these principles, provides an analytic basis for (1) and (2), generalizes this propagated uncertainty equation to scalar fields, and outlines a newer approach that reduces uncertainty compared to the previous generation of uncertainty modeling (Calder and Mayer 2003; Zambo et al. 2015).

In the remaining part, "Base Equation" section provides principles that enable families of uncertainty propagation equations to be proposed, and examines some properties of these functions. "Conservative Approach" section shows how a single equation with three end-user adjustable parameters emerges from this family; (1) comes from this family with a specific choice of parameters.

The "Mean-distance Approach" section develops a new version of the model that reduces the number of user-specified parameters needed from three to two by considering the expected distance that an observation will be propagated, given its horizontal uncertainty, rather than the worst-case distance as was used in previous versions. Due to the Rician distribution of distance (under Gaussian horizontal uncertainty), the expected value has a solution through Laguerre functions, discussed in "Analytic Expression Using $L_{1/2}(x)$" section. "The Evaluation of $L_{1/2}(x)$" and "Chebyshev Series Fit to ${}_1F_1(-1/2, 1; x)$" sections discuss numerical techniques using the confluent hypergeometric function and Chebyshev polynomials to efficiently compute the expected propagation distance.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

"The Example: Estimating Depth" section provides a numerical example for estimation of depth and uncertainty using the models developed in "The Mean-distance Approach" section. The paper ends with "Discussion" and "Conclusions" sections.

Base equation

Uncertainty is defined here to be standard deviation, denoted σ (Taylor and Kuyatt 1994). For convenience, however, the equations are written in terms of variance, denoted σ^2. Methods for predicted uncertainty of scalar fields of interest in this paper follow these general principles:

1. The value of the information in an observation should decay away from its nominal location; that is, the uncertainty should increase with distance from the nominal observation location.
2. Information inherent in an observation with higher horizontal uncertainty should be more diffuse, so that the uncertainty predicted after propagation should be higher than that for an observation with lower horizontal uncertainty.
3. Propagation should scale with the resolution of analysis so that the distance considered significant is measured against the grid resolution.
4. Since the uncertainty is being transferred to a point for consideration, the final propagated uncertainty should have no horizontal component, since the location is precisely defined. Consequently, the equation must be able to combine the effects of positional and field uncertainty in the observation into a purely field uncertainty at the reference point.

Many possible equations could be proposed that follow these principles, but one of the simplest is

$$\sigma^2 = a + bd^{\alpha} \tag{3}$$

where $a, b \in \mathbb{R}_{>0}$ are constants, $\alpha \in \mathbb{R}_{\geq 1}$, and d is the effective distance between the observation and reference locations. It should be clear from the form of (3) that the distance term bd^{α} with $\alpha \geq 1$ satisfies the first general principle, the constant term satisfies the second general principle, and the whole form satisfies the fourth. The third general principle is induced by considering appropriate boundary conditions. That is, for consistency, at $d = 0$, set $\sigma^2 = \sigma^2_f$ where σ^2_f is the field variance of the observation, and hence, $a = \sigma^2_f$. In order to be able to control the speed of uncertainty increase with distance, assume a secondary boundary condition such that $\sigma^2 = k\sigma^2_f$ at some suitable distance, such as the grid sample spacing, Δ_g. Hence

$$\begin{aligned} k\sigma_f^2 &= \sigma_f^2 + b\Delta_g^{\alpha} \\ \Rightarrow \quad b &= \sigma_f^2 \frac{k-1}{\Delta_g^{\alpha}} \end{aligned} \tag{4}$$

Substituting in (3) yields

$$\begin{aligned} \sigma^2 &= \sigma_f^2 + \sigma_f^2 (k-1) \frac{d^{\alpha}}{\Delta_g^{\alpha}} \\ &= \sigma_f^2 \left(1 + (k-1) \left(\frac{d}{\Delta_g} \right)^{\alpha} \right) \end{aligned} \tag{5}$$

Eq. (5) forms the base model for which we propose two methods of computation, which differ in the way they compute parameter d. The distance d used has to reflect the actual

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

distance traveled by the observation to the interpolation point rather than a nominal distance, d_0 between the most likely location of the observation and the interpolation location. For tractability, it is assumed that the horizontal observation distribution is Gaussian.

Conservative approach

The simplest model to include the effects of σ_h on d is to assume that d must be greater than d_0 by some factor that is a function of σ_h. The "conservative" approach is to create this function so that uncertainty is maximized. Due to positivity of d and σ_h, if $s_h > 0$, this corresponds to $d = d_0 + s_h\sigma_h$ so that (5) becomes

$$\sigma^2 = \sigma_f^2\left(1 + (k-1)\left(\frac{d_0 + s_h\sigma_h}{\Delta_g}\right)^{\alpha}\right) \tag{6}$$

This equation has the disadvantage of having three adjustable parameters: α, s_h, and k. If $k \equiv 2$, (6) can be simplified since the uncertainty in the vertical doubles within one effective sample spacing, irrespective of α, so that

$$\sigma^2 = \sigma_f^2\left(1 + \left(\frac{d_0 + s_h\sigma_h}{\Delta_g}\right)^{\alpha}\right) \tag{7}$$

Eq. (7) has the same form as (1), which was developed for the CUBE algorithm (Calder and Mayer 2003). It has the benefit of simplicity in that the scale factor s_h modifies the rate at which horizontal uncertainty is converted to field uncertainty, and the value of α controls the shape of the transfer curve (Figure 1). The rate of increase of uncertainty (Figure 2) can be significant, however, particularly for what might be considered "reasonable" scale factors corresponding to, e.g., the 95% CI for the observation.

Adjusting the scale factor k can adjust this behavior somewhat, but at the cost of another parameter which interacts with the others to make it more difficult for the algorithm designer, or user, to adjust the transfer behavior to match the specific problem being tackled. This makes the propagation equation difficult to adjust in practice.

Mean-distance approach

Instead of assuming the worst-case transfer distance for all observations, it is possible to compute the expected distance instead. Assuming isotropic Gaussian distribution in the horizontal, the distance is Rician distributed so that the expected distance can be computed as an analytic expression via a (generalized) Laguerre function. Implementing this computation is more expensive than the "conservative" function, but has the benefit of avoiding the s_h parameter.

Evaluation of the expected distance must be implemented numerically using the following stages:

1. Analytic expression using the Laguerre function, $L_{1/2}(x)$, with $x = (-d^2{}_0/2{\sigma_h}^2)$; a numerical solution is required.
2. One-time numerical computation of $L_\nu(x)$ via the confluent hypergeometric function (CHGF), ${}_1F_1(\alpha, \beta; x)$; this computation can be slow and intensive.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

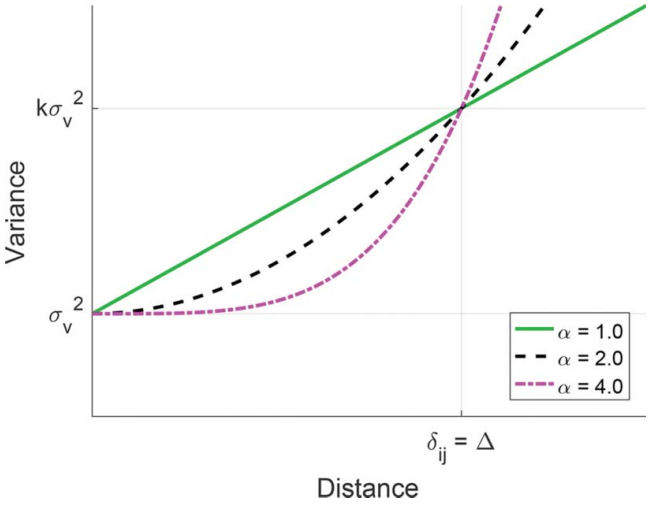

Figure 1. Increase in propagated vertical uncertainty as a function of effective distance and control parameter α for (7). Here, $s_h = 1$, $\Delta_g = 1.0$, and $\sigma^2{}_f = 1.0$ to normalize the graph. The amplification at one grid spacing is controlled by the form of the equation, while the control parameter manipulates the shape.

3. Fit of the CHGF using Chebyshev polynomials of the first kind, $T_n(x)$, for a rapid approximation to the one-time CHGF computation result.

In-depth discussions of these functions can be found in Abramowicz and Stegun (1965); see also National Institute of Standards and Technology (2015), Olver et al. (2010), and Arfken et al. (2012, Ch. 18).

Analytic expression using $L_{1/2}(x)$

For a circularly symmetric Gaussian distribution in $\mathbb{R}^2$ with non-zero mean vector, the distribution of distance from the origin (equivalently, here, the distance the observation travels) is given by a Rician distribution (Rice 1945)

$$p\left(d; d_0, \sigma_h^2\right) = \frac{d}{\sigma_h^2} I_0\left(\frac{d_0 d}{\sigma_h^2}\right) exp\left\{ - \frac{d_0^2 + d^2}{2\sigma_h^2}\right\} \tag{8}$$

where $I_0(x)$ is a modified Bessel function of the first kind of order zero, and d_0 is the nominal distance (magnitude of the mean vector). If $d_0 \equiv 0$, this reduces to the Rayleigh distribution, but in general the mean value is given by

$$\overline{d}\left(d_0, \sigma_h^2\right) = \mathbb{E}[d] = \sigma_h \sqrt{\frac{\pi}{2}} L_{1/2}(x) \tag{9}$$

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

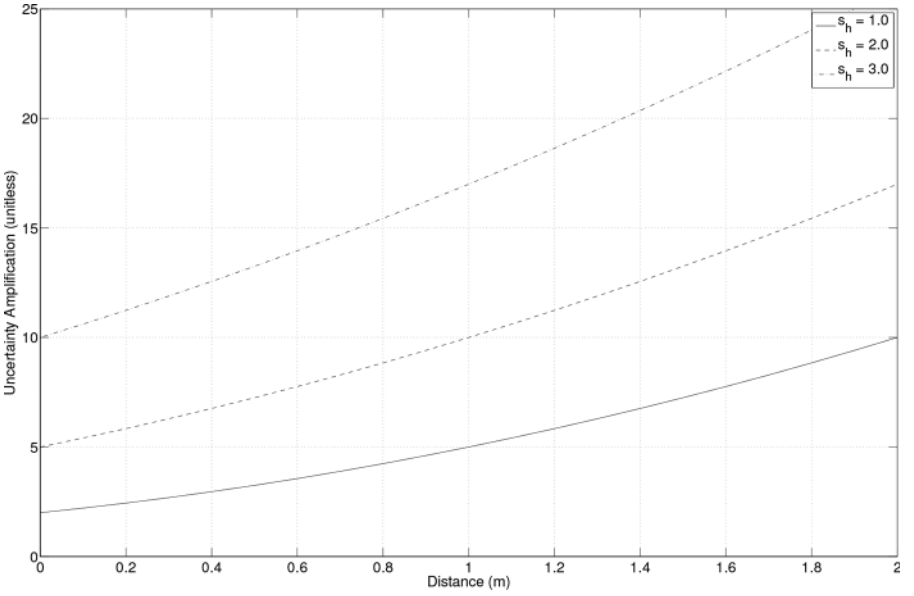

Figure 2. Increase in propagated field uncertainty as a function of nominal distance between observation and reference location, and the horizontal uncertainty scale factor, s_h, for a unit horizontal uncertainty. Increasing the scale factor changes the rate of increase of uncertainty, but also offsets the zero-distance uncertainty, often by very significant amounts.

where $x = -d^2{}_0/(2\sigma_h^2)$ and $L_v(x)$ is a generalized Laguerre function (Abramowicz and Stegun 1965, Ch. 22). Thus, propagated variance is

$$\begin{aligned}\sigma^2 &= \sigma_f^2\left(1+(k-1)\left(\frac{\overline{d}\left(d_0,\sigma_h^2\right)}{\Delta_g}\right)^{\alpha}\right)\\ &= \sigma_f^2\left(1+(k-1)\left(\frac{\sigma_h\sqrt{\frac{\pi}{2}}L_{1/2}(x)}{\Delta_g}\right)^{\alpha}\right)\end{aligned} \tag{10}$$

With (10), the uncertainty model now has only two manually set parameters, k and α, easing the specification of behavior for users.

Evaluation of $L_{1/2}(x)$

Computation of $L_{1/2}(x)$ in (9) and (10) is non-trivial (Thompson 1997; Muller 2001; Pearson et al. 2014) because of the non-integer order of the Laguerre function. Since the present problem requires frequent numerical evaluation of the Laguerre function $L_{1/2}(x)$, computational efficiency is essential for implementation.

Modified Bessel function expansion

Through fractional differentiation (Herrmann 2014) of the defining equation for the Laguerre function (Andrews et al. 1999), it is possible to show (for example by expanding

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

both sides as power series, or through a computer-algebra system) that $L_{1/2}(x)$ can be expressed in terms of modified Bessel functions of the first kind (Andrews et al. 1999) of zeroth and first order, *viz.*:

$$L_{1/2}(x) = exp(x/2)(\mathcal{I}_0(x/2) + x\mathcal{I}_1(x/2) - x\mathcal{I}_0(x/2)) \tag{11}$$

where for efficiency, the two nominal evaluations of the modified Bessel function of order zero could be trivially collapsed if required (i.e., computing $(1 - x)I_0(x/2) + xI_1(x/2)$). While this approach has a number of theoretical benefits, most particularly if further manipulation of the formulae were required, evaluation of modified Bessel functions is itself a significant computational cost, making this approach less attractive where many evaluations are required, as here.

Use of ${}_1\mathrm{F}_1(\alpha, \beta; x)$ for $\mathrm{L}_\nu(x)$

Common procedures for computation of $L_\nu(x); \nu \notin \mathbb{Z}$ begin by rewriting it in terms of Kummer's CHGF (Abramowicz and Stegun 1965; Arfken et al. 2012; Buchholz 1969; National Institute of Standards and Technology 2015) as

$$L_n^{(m)}(x) := \binom{m+n}{n} {}_1F_1(-n, m+1; x) \tag{12}$$

where Kummer's CHGF is defined by the series

$$\begin{aligned} {}_1F_1(\alpha, \beta; x) &= 1 + \frac{\alpha}{\beta}x + \frac{\alpha(\alpha+1)}{\beta(\beta+1)}\frac{x^2}{2!} + \frac{\alpha(\alpha+1)(\alpha+2)}{\beta(\beta+1)(\beta+2)}\frac{x^3}{3!} + \cdots \\ &= \sum_{k=0}^{\infty} \frac{(\alpha)_k}{(\beta)_k}\frac{x^k}{k!} \end{aligned} \tag{13}$$

and $(\alpha)_n = \alpha(\alpha + 1)(\alpha + 2)\cdots(\alpha + n - 1)$ is the Pochhammer symbol (or rising factorial); $(\alpha)_0 \equiv 1$. With $n = 1/2$ and $m = 0$, and hence $\alpha = -1/2$, and $\beta = 1$, (12) becomes

$$L_{1/2}(x) = {}_1F_1(-1/2, 1; x) \tag{14}$$

Thus, use of (9) and (14) provides $\overline{d}$,

$$\overline{d}(d_0, \sigma_h^2) = \sqrt{2\sigma_h^2}\Gamma\left(\frac{3}{2}\right){}_1F_1\left(-\frac{1}{2}, 1; x\right) \tag{15}$$

which can be evaluated by direct substitution as

$$\begin{aligned} {}_1F_1\left(-\frac{1}{2}, 1; x\right) &= \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}\right)_k}{(1)_k}\frac{x^k}{k!} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{\left(-\frac{1}{2}\right)_k}{(k!)^2} x^k \end{aligned} \tag{16}$$

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

(since $(1)_k = k!$). Subsequent terms of (16) alternate in sign and analysis of the ratio of terms (Abramowicz and Stegun 1965; National Institute of Standards and Technology 2015) shows that the sequences converges, but the terms do not reduce in magnitude as $k \to \infty$ and therefore the rate of convergence depends on the argument. Practically, therefore, direct evaluation is only feasible for small values of $x \lesssim 40$–50, even using a double-precision floating-point representation.

Computation of ${}_1\mathrm{F}_1(-1/2, 1; x)$

In general, there is no optimal computational procedure for ${}_1F_1(\alpha, \beta; x)$ (Lopez and Ester 2014; Pearson et al. 2014). As discussed in depth in Pearson et al. (2014), differing strategies exist depending on the values of the two parameters and functional argument. For the present case, $\alpha = -1/2, \beta = 1, x \in \mathbb{R}, x \leq 0$, and hence from (Pearson et al. 2014, Section 3), for values of $x \leq 10$, summation of terms in (16) is efficient. For large arguments (e.g., for $x > 500$), Slater's expansion (Abramowicz and Stegun 1965; National Institute of Standards and Technology 2015; Slater 1953)

$$\frac{{}_1F_1(\alpha, \beta; z)}{\Gamma(\beta)} = \frac{e^x x^{\alpha-\beta}}{\Gamma(\alpha)} \left\{ \sum_{n=0}^{S-1} \frac{(\beta-\alpha)_n (1-\alpha)_n}{n!} x^{-n} \right\} \tag{17}$$

is effective (National Institute of Standards and Technology 2015, Section 13.29(i)). Practically, however, evaluation of either (16) or (17) would be prohibitively expensive for large datasets where the expected distance has to be computed for each observation at least once. Consequently, a more efficient computational scheme is required.

Chebyshev series fit to ${}_1\mathrm{F}_1(-1/2, 1; x)$

As recommended in Pearson et al. (2014), an efficient approach to computing ${}_1F_1(-1/2, 1; x)$ is to perform a one-time computation using (16) for $x < 40$ and (17) for $x \geq 40$. Then, approximate the result using a suitable polynomial series,

$${}_1F_1(\alpha, \beta; x) \approx \sum_{n=0}^{N-1} C_n P_n(x) \tag{18}$$

where $P_n(x)$ is a polynomial function of degree n and the choice of $N < \infty$ reflects the desired level of approximation.

For the current application, Chebyshev polynomials of the first kind, $T_n(x)$, form a pragmatic computational solution. $T_n(x)$ provide numerical advantages over alternatives, such as a power series, for rapid convergence and compact representation (i.e., the sequence can be computed to user specified accuracy with fewer terms) in the domain $[-1, 1]$ (Arfken et al. 2012; Boyd 2001; Press et al. 2007). Within this domain, Chebyshev polynomials are orthogonal and have range $[-1, 1]$. For Chebyshev polynomials, the coefficients required for (18)

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

are

$$C_n = \frac{2}{J}\sum_{j=0}^{J-1} {}_1F_1\left(\alpha, \beta; x_j\right) T_n\left(x_j\right) \tag{19}$$

and x_j is the set of domain points used to numerically compute the CHGF.

With appropriate substitutions, then, (10) can be rewritten as,

$$\sigma^2 \approx \sigma_f^2\left(1 + (k-1)\left[\frac{\sqrt{\frac{\pi}{2}}\sigma_h}{\Delta_g}\sum_{n=0}^{N-1} C_n T_n(x)\right]^{\alpha}\right) \tag{20}$$

giving a computable form for the propagated uncertainty.

Direct evaluation of (20) is possible, but suffers from significantly higher errors as $x \to 0$ (Figures 3 and 4). To avoid this, the Chebyshev polynomials can be fitted to the pre-warped domain $x'(\gamma) = x^{\gamma}$, $0 < \gamma < 1$, which stretches out the region close to $x = 0$, and results in having to evaluate significantly fewer coefficients for any given level of approximation (Figures 5 and 6).

Comparison of implementation methods

A full comparison of the implementation efficiency of the two methods proposed is essentially impossible, since that of (11) depends strongly on the implementation of the standard library function for the modified Bessel function of the first kind. An informal assessment of

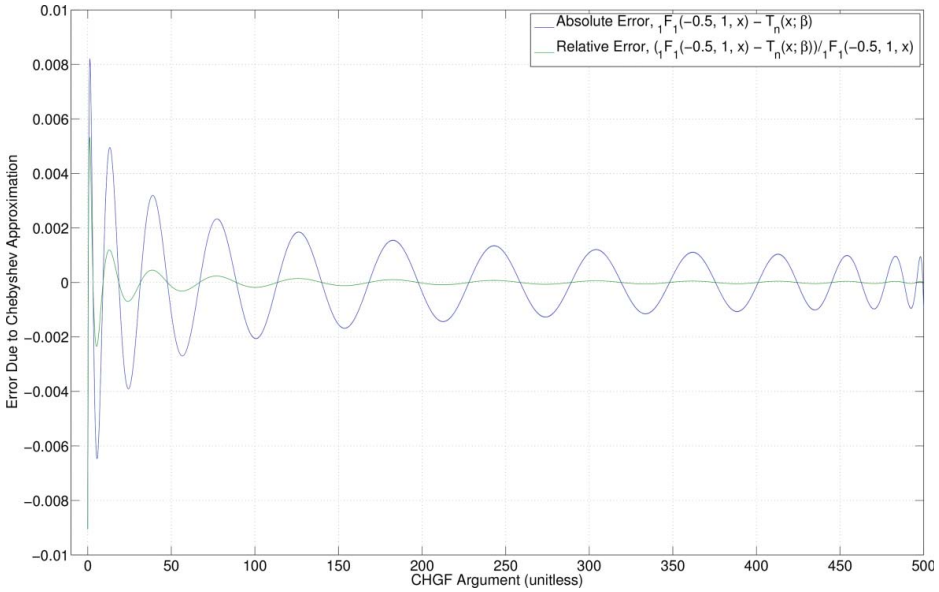

Figure 3. Absolute and relative errors in approximating the CHGF with a Chebyshev polynomial (of 26 coefficients, designed to give absolute accuracy to within 0.01 of the computed CHGF value). The shape of the CHGF as $x \to 0$ leads to bunching of error ripples around $x = 0$.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

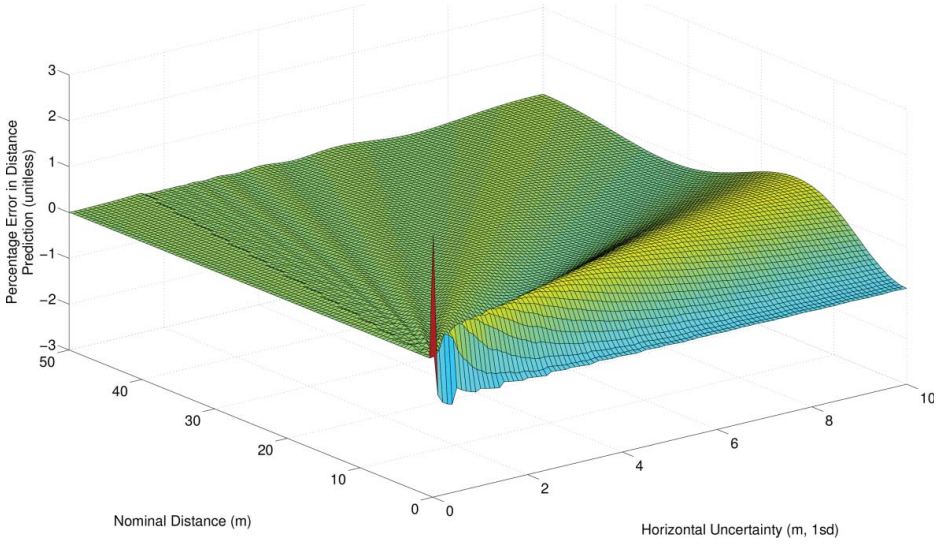

Figure 4. Percentage relative error in distance prediction by approximating the CHGF with a Chebyshev polynomial (of 26 coefficients, designed to give absolute accuracy to within 0.01 of the computed CHGF value).

likely efficiency can be conducted, however, by considering the code generated from trial implementations. In this regard, modified Bessel function implementations from the GNU Scientific Library (Galassi et al. 2009) and Numerical Recipes (Press et al. 2007) were compared to an implementation of the mean distance computation coded for the examples following.

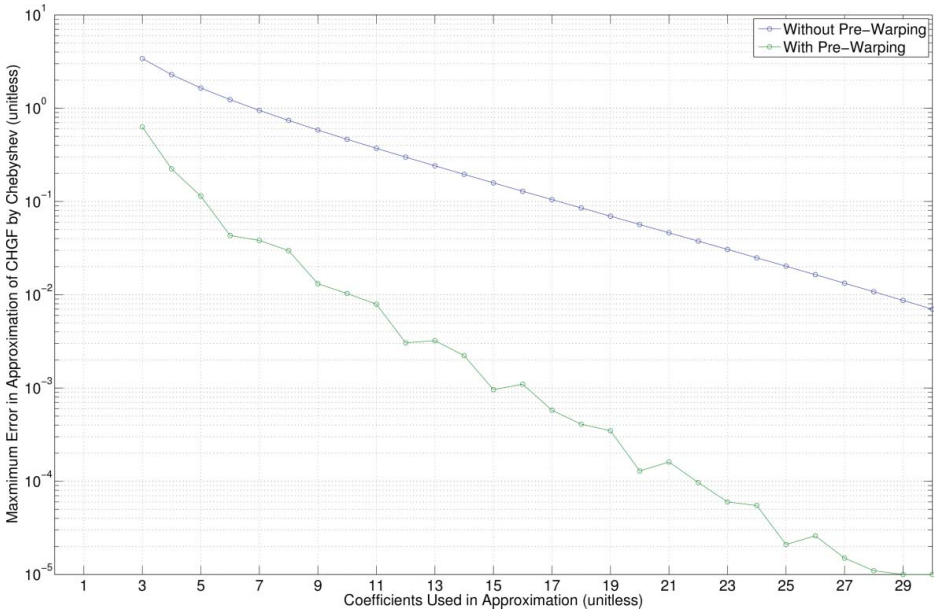

Figure 5. Approximation error in evaluating the CHGF via a Chebyshev polynomial as a function of the number of coefficients used, with and without domain pre-warping. Pre-warping the evaluation domain significantly improves the approximation performance.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

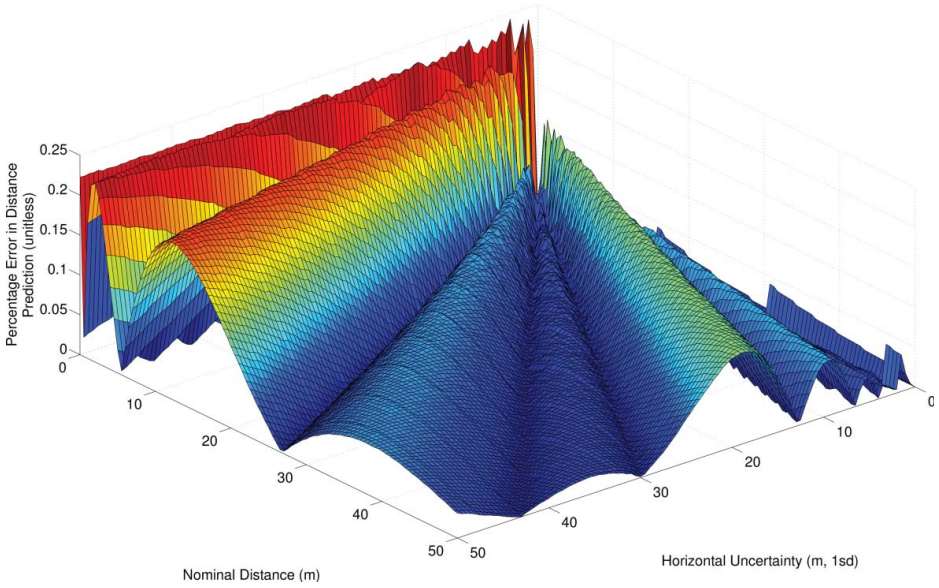

Figure 6. Percentage relative absolute error in approximating the CHGF with a Chebyshev polynomial (of 12 coefficients with pre-warping coefficient $\gamma = 0.2$). The pre-warping of the evaluation domain significantly reduces the number of coefficients required to achieve a given level of approximation, with concomitant improvement in evaluation speed.

The GNU Scientific Library code uses Chebyshev polynomials of various degrees to evaluate its modified Bessel functions, while the Numeric Recipes code uses a variety of polynomial ratios. The minimum degree of the Chebyshev approximations in the GSL was 12 coefficients for each modified Bessel function, the same as that used in the algorithm developed here. This means that it will take at least twice as long to evaluate a mean distance based on (11) as (20). An estimate of instruction latency for the Numerical Recipes implementation was constructed by generating an assembly listing of the C-code, and evaluating the latencies of the individual instructions. The estimated latency is approximately twice that of the code developed for this work, although there might be some variability due to the degree to which the CPU used has hardware to assist in computation of transcendental functions, particularly the exponential.

The pre-warped Chebyshev approximation of (20) therefore appears most likely to be the preferred implementation method. Of course, for maximum performance it would always be possible to invoke the classical trade-off of speed for memory requirement by pre-computing a look-up table for the pre-warped form of $L_{1/2}(x)$ and simply applying table look-up with interpolation. The trade-offs required would have to be evaluated *in situ* for any specific implementation.

Example: Estimating depth

A particular example of scalar field estimation is estimation of the depth of the ocean. Whether conducted for hydrographic mapping, oil and gas exploration, military mapping, or simply for scientific exploration purposes, the depth of water can be measured or

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

estimated by a variety of technologies ranging from laser time-of-flight measurements (Light Detection and Ranging, or lidar) for shallow water to satellite altimetry for the deep oceans. In many cases, however, acoustic means are used to estimate the depth of water using active-source time-of-flight reflectometry, or sonar (Sound Navigation and Ranging). Whatever the source of observations, the goal of the data processing algorithm is the same: estimate the depth from the randomly scattered observations, ideally dealing with outliers and noise points, and by preference estimating an uncertainty associated with the declaration of depth.

Mechanics of the CUBE algorithm

The CUBE algorithm (Calder and Mayer 2003) for which (7) was originally designed is a particular example of such a depth estimation algorithm. The algorithm operates on the principle of estimating the depth at a specific point ("estimation node") with respect to an arbitrarily defined, but fixed, coordinate frame. Each observation is qualified with a horizontal and vertical uncertainty, and this uncertainty is used both to determine which estimation nodes should receive the information on depth inherent in the observation, and how to weigh that information in assessing the depth estimate.

In order to accommodate outliers and noise points, CUBE continually monitors observations being added to the developing depth estimate, using the assessed uncertainty to identify when a new observation is inconsistent with the current depth estimate. Such inconsistent observations are segregated into an alternative hypothesis on depth, essentially forming an unsupervised clustering algorithm for the observations propagated to the estimation node. At any point (although most often after all of the observations have been used), the user can request the "most likely" reconstruction currently known, at which point the algorithm attempts to select one of the (potentially many) hypotheses that is most likely to be the true seafloor, qualifying this selection by a number of metrics that attempt to communicate the status of the reconstruction. "Most likely" is defined by a variety of rules that attempt to avoid mis-selection of noise hypotheses; the default, and most common, selection is to select the hypothesis with the most observations assimilated.

A core component of the CUBE algorithm is the method by which observations are propagated from their nominal location to the location of the surrounding estimation nodes (which are most often, although not necessarily, arranged in a uniform resolution grid), and how this affects the estimated uncertainty of the observation. Clearly, therefore, how to select the distance in (5) can affect the results of CUBE processing. The example here explores that difference in a set of three depth estimation scenarios.

Example depth estimation scenarios

In order to simplify the example, the CUBE algorithm was limited to considering estimation in a 1D line rather than a 2D grid; node spacing of 0.25 m was used for all examples (this is essentially arbitrary, but appropriate for a modern multibeam echosounder in shallow water), and a domain of 10.0 m total length was constructed. Three scenarios were considered:

1. A flat seafloor at 10.0 m nominal depth. This depth is arbitrary, but mid-range for high-resolution survey data used for hydrographic purposes.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

2. The same flat seafloor, with a 1.0 m vertical step half-way through the domain.
3. The same flat seafloor, with a 1.0 m wide, 1.0 m high square object imposed half-way through the domain. A 1.0 m cube is a common target for object detection tests in hydrographic surveying because of its reference in IHO survey specifications (International Hydrographic Bureau 2008).

The CUBE algorithm parameters not affected by choice of propagation equation were consistent over all test scenarios, and were set to the recommended algorithm defaults[1]. For both propagation algorithms, $k = 2.0$ and $\alpha = 2.0$ were selected (corresponding to CUBE default settings); for the conservative algorithm, $s_H = 1.96$ was selected (also the CUBE default), while for the mean distance algorithm a total of 12 Chebyshev parameters were retained with pre-warping factor $\gamma = 0.2$. This provides centimetric accuracy in the approximation of mean distance (Figure 5–6).

Synthetic data were generated as IID Gaussian variates using the IHO S.44 (International Hydrographic Bureau 2008) uncertainty model of an expanded uncertainty of $U = \sqrt{a^2 + (bz)^2}$ at 95% CI (for depth z) assuming Gaussian statistics. Vertical uncertainties were set to 0.25 m fixed and 0.75% depth-dependent; horizontal uncertainties were set to 0.4 m fixed and 1% depth-dependent. These values are typical of SBAS-aided[2] GNSS-controlled[3] hydrographic survey systems in shallow water. In all cases 100 observations uniformly distributed across the computational domain were used.

Results

As might be expected, in the simplest case of a flat seafloor, Figure 7, the two propagation equations lead to similar CUBE behaviors, with a single hypothesis being formed from all of the data presented at each node, and the uncertainty bounds predicted covering the true depth in each case. There are very small differences between the values generated by the estimation algorithm in each case, which are barely visible on the scale of Figure 7. This is due to the different uncertainties generated by the two propagation equations affecting the weighting of the observations in the estimation portion of the algorithm. As is clearly evident, however, the differences are statistically insignificant with respect to the uncertainty bounds predicted, and can be ignored for all practical purposes in this "ideal" case. Note that the bounds are at 95% CI according to the number of soundings used in the estimate of the hypothesis, based on the sample estimate of the variance of the data. Since there is only a single hypothesis, the hypothesis selection uncertainty is reported as 0.0, equivalent to the statement "the odds against the right hypothesis being other than reported are 5:1 or more."

A significant difference is, however, observed in the behavior of the algorithm when considered over the stepped seafloor of Figure 8. Here, the higher propagated uncertainty from the conservative propagation equation means that observations from both sides of the step appear to be consistent in depth within their (propagated) uncertainty bounds, and are

[1] Capture distance 5.0% for depth, IHO S.44 5ed Order 1A survey for maximum observation propagation distance, median prefilter queue of one sample, estimate offset threshold of 4.0, Bayes factor threshold of 0.135, sequential monitoring threshold of five samples, and prior (most observations assimilated) hypothesis selection.
[2] Satellite-Based Augmentation System, a system for providing GNSS correctors by satellite transponder, rather than having a local correction system.
[3] Global Navigation Satellite System, or colloquially "GPS" (Global Positioning System), most often the U.S.-based NavStar.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

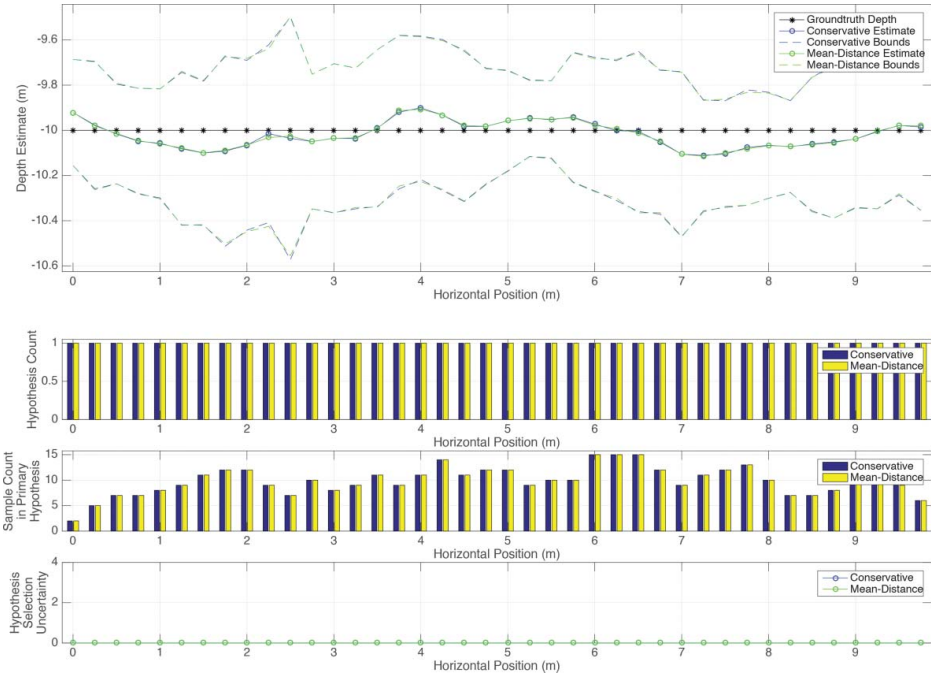

Figure 7. Example estimation with CUBE using the conservative (blue) and mean distance (green) uncertainty propagation equations over a flat seafloor of depth 10.0 m. The algorithms behave equivalently, although small differences in estimate, and associated uncertainty, can be seen due to the differences in the propagated uncertainty.

therefore combined by CUBE into a single hypothesis. Therefore, no second hypothesis is formed to segregate the two observation populations, and the algorithm effectively computes a moving average over the step, although the uncertainty bounds are still consistent with the true depth. In contrast, the reduced uncertainty provided by the mean distance propagation equation allows CUBE to correctly detect that there are two populations, and successfully segregate them, selecting the correct hypothesis to significantly improve the reconstruction, Figure 9. Note also that CUBE is able to more correctly indicate the hypothesis selection uncertainty in the step region for the mean distance propagation equation, showing that there is potentially significant confusion about which hypothesis should be selected in this region. In contrast, the higher propagated uncertainty generated by the conservative propagation equation convinces CUBE to form just one hypothesis, which by definition has to be "best." Avoiding over-confidence like this is a key to better user feedback on algorithm performance, which aids and accelerates user interaction time.

Finding and accurately representing small objects is a key issue for depth estimation algorithms, particularly when used for hydrographic purposes in shallow water. Figure 10 shows the reconstruction of a typically-sized target object, a 1.0 m cube, and readily highlights the distinct difference in behavior of the CUBE algorithm with the two uncertainty propagation equations. As for the step example, the more plausible uncertainty estimates generated by the mean distance algorithm mean that the CUBE algorithm can more readily distinguish between the two observation populations corresponding to the top of the object and the surrounding seafloor, resulting in more consistent generation of hypotheses, and therefore

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

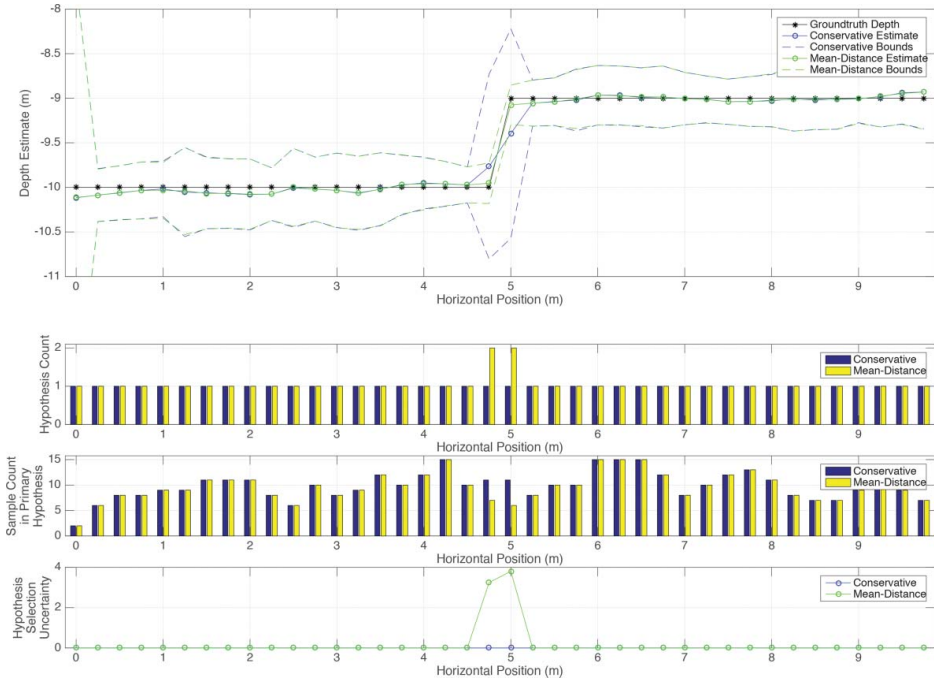

Figure 8. Example estimation with CUBE using the conservative (blue) and mean distance (green) uncertainty propagation equations over a seafloor that steps from 10.0 to 9.0 m half-way through the domain. Higher propagated uncertainty from the conservative algorithm convinces CUBE to form only one hypothesis in the transition region, leading to poor (although statistically correct) reconstruction.

better reconstruction of the shape of the object and much tighter sample uncertainty bounds since the two populations are less likely to be conflated within one hypothesis.

The reconstruction of the right-hand edge of the object (at $x = 5.5$ m) is not ideal when using the mean distance propagation equation, and the reported uncertainty bounds do not include the true depth. Here, the CUBE algorithm has generated two hypotheses, Figure 11, but has selected the wrong hypothesis for "best" reconstruction. The other hypothesis does however match the true depth within its uncertainty bounds. In practice, in this situation, the user's task is to confirm that the algorithm has selected the correct hypothesis, and this configuration of hypotheses would be cause for an intervention in the algorithm's reconstruction behaviors.

This selection confusion is caused by the two hypotheses here, both having exactly the same number of observations incorporated. In this case, with the default hypothesis selection rules, which is selected as "best" is essentially arbitrary (due to the algorithm's implementation, the one that is spawned last is selected). However, the better uncertainty bounds at least allow CUBE to indicate this by declaring a hypothesis selection uncertainty of 4.0, which is equivalent to saying "the next-best choice is just as good as the hypothesis being reported here, and therefore this reconstruction should be viewed with suspicion." The conservative propagation equation, by contrast, only generates a single hypothesis, which it confidently, although mistakenly, marked as of low selection uncertainty. The uncertainty bounds do cover the true value of depth, so this is not a disaster, but it would fail to alert the user to a potential difficulty that might demand further attention.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

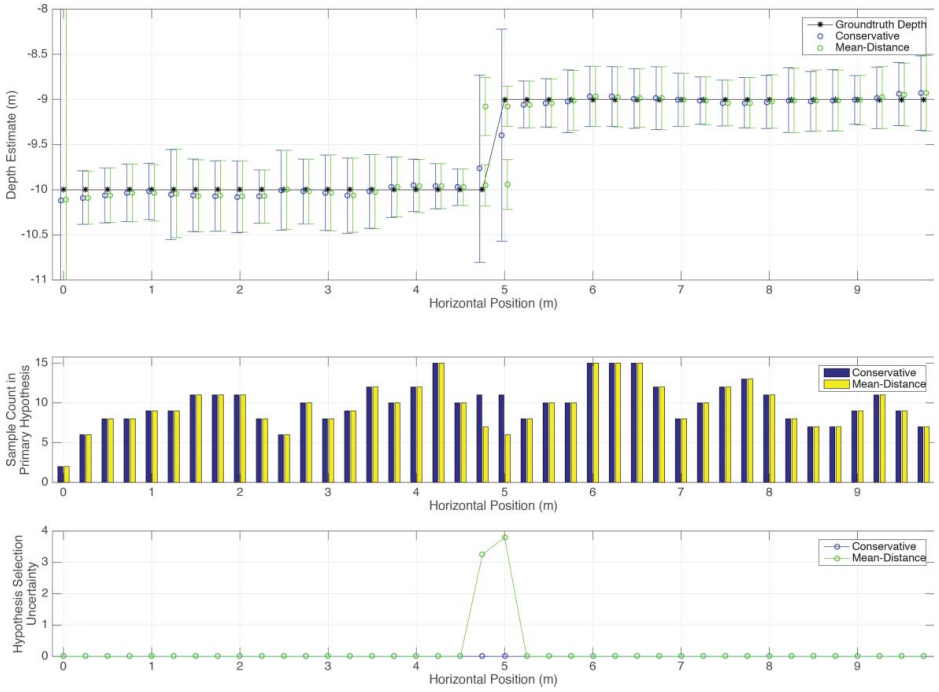

Figure 9. Example of depth hypotheses generated with CUBE using the conservative (blue) and mean distance (green) uncertainty propagation equations over a seafloor that steps from 10.0 to 9.0 m half-way through the domain. Note that for visualization purposes the hypotheses for the conservative and mean distance propagation equations are displaced slightly to left and right, respectively, of the position of the estimation node, as indicated by the true depth marker. More realistic propagated uncertainty from the mean distance propagation equation results in secondary hypotheses being formed in the step transition region, leading to better reconstruction, albeit with caveats in the form of the hypothesis selection uncertainty in that area.

These simple examples demonstrate that the choice of propagation equation, and to some extent the parameters for that equation, can have a significant effect on the depth reconstruction achieved, even using the same core depth estimation algorithm. The computational cost of these propagation equations are different, however, which is another factor in deciding which equation to use. In these tests, the average assimilation time across all examples on a particular system[4] for 100 samples with the conservative propagation equation was approximately 65 μs, while the assimilation time for the mean distance equation on average was 104 μs, leading to an average additional cost of approximately 61% of the original runtime to implement the mean distance equation. Further tests with more samples suggest that the costs remain approximately 60–65%. Not all of this extra runtime is due to the mean distance computation, however: when new hypotheses are formed, there is a non-trivial overhead time due to memory management. Consequently, this is expected to be a worst-case estimate for the extra computational cost involved.

[4] Apple iMac with Intel Core i7 at 4 GHz, 32 GB RAM, running OS X 10.11.6, running unoptimized code from XCode 7.3.1.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

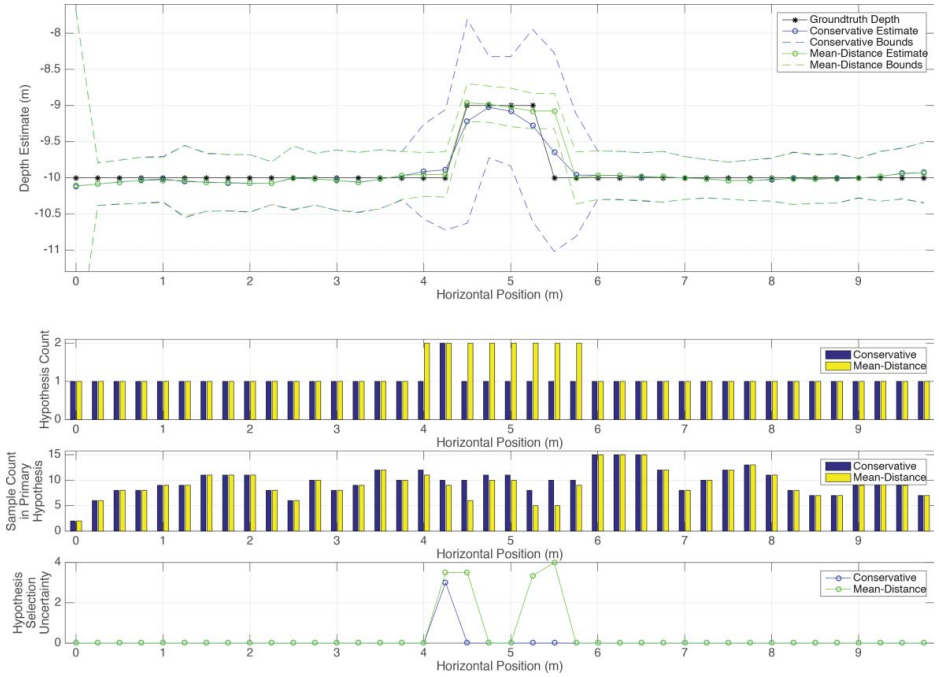

Figure 10. Example estimation with CUBE using the conservative (blue) and mean distance (green) uncertainty propagation equations over a seafloor containing a 1.0 m wide, 1.0 m high object half-way through the domain. With the conservative propagation equation, higher propagated uncertainties lead to limited hypothesis generation and poor object reconstruction. With the mean distance algorithm, lower propagated uncertainties make the different populations of observations more evident, leading to more frequent hypothesis generation and therefore better reconstruction.

Discussion

Using (7) and (20), the propagation equation can be set at two different information content levels. The tradeoffs are computational complexity, number of free parameters, and discontinuity detection sensitivity. Practitioners should decide which version of the propagation equation to use for their application on a case-by-case basis. They must weigh whether the extra computational cost is worth the information gained with the mean-distance approach versus the less-expensive conservative approach.

The conservative approach of (7) has the advantages of lower computational cost; the estimator is a simple algebraic equation. This equation, however, requires specification of three parameters. The mean-distance approach containing the Laguerre function requires two parameters instead of three. The computational cost, however, is higher than the conservative approach. The estimator requires nontrivial numerical evaluation of the CHGF and Chebyshev series fitting.

For applications that require sensitive discontinuity detection, the mean-distance approach provides higher sensitivity than the conservative approach. Within the CUBE algorithm, this version of the propagated uncertainty equation provides more candidate depths for hypothesis resolution. This behavior may be desired, for example, in a shipping area where navigation safety concerns are high. Although the conservative approach smooths over smaller discontinuities, it will still detect larger ones. For use with applications like

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

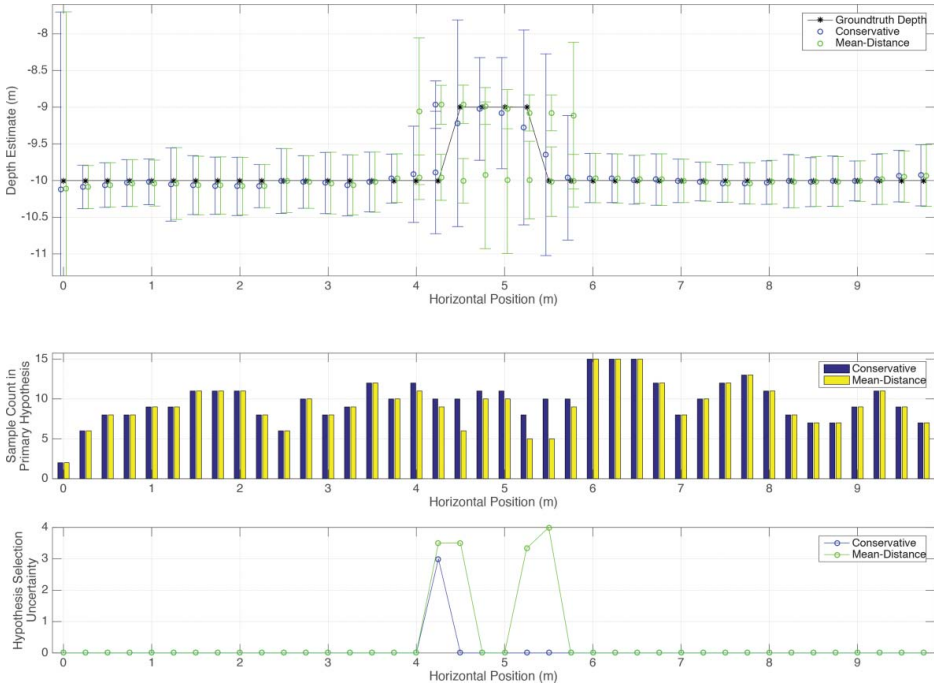

Figure 11. Example of depth hypotheses generated with CUBE using the conservative (blue) and mean distance (green) uncertainty propagation equations over a seafloor that steps from 10.0 to 9.0 m half-way through the domain. The results using the mean distance propagation equation are generally better, but the node at $x = 5.5$ m has two hypotheses that the algorithm had difficulty distinguishing at reconstruction time, as highlighted by the hypothesis selection uncertainty value of 4.0.

CURVE, where uncertainty is being estimated for fast interpolation algorithms, the conservative approach provides smoothing behavior and an uncertainty estimate that an end-user may find satisfactory for their interpolation task at lowered computational cost.

The lowest uncertainty estimate for σ^2 is driven by the term σ_f and the amount of data obtained for a single node. As discussed in Bourgeois et al. (2016), the conservative form is at a minimum for $d_0 = 0$. This minimum also holds for the mean distance approach by inspection of (10) and monotonic behavior of (13).

As with any parametric equation, choice of parameters for either (7) or (20) can be problematic and problem specific, making generic guidance difficult. In applications, setting $\alpha = 2.0$ appears to be most common, corresponding to some notion of squared distance weighting in estimation. For the original CUBE algorithm, setting $s_H \approx 2.0$ was a default choice, reflecting in a general sense a couple of standard deviations "worst case" extra distance due to the horizontal uncertainty, but this appears to generate too rapid an increase in uncertainty with distance, and many field users set s_H significantly lower in order to avoid this; in practice, using the mean distance algorithm is probably a better choice if this is significant to a particular problem. The uncertainty scale factor k controls the rate at which the horizontal uncertainty and propagation distance affect the propagated vertical uncertainty; setting $k = 2.0$ gives the default CUBE behavior, although since this results in the uncertainty doubling after one grid spacing, this may be too much for many applications. Values in the open range (1, 2) are therefore recommended as an initial starting point for experimentation.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

Conclusion

This paper considered the modeling of propagated uncertainty in a scalar field. The model provides an uncertainty estimate for an interpolation point estimated from the positional and field values of the observations. The uncertainty in the observation is magnified when used for the computation of the interpolated output point, increasing as the distance between the input and output positions lengthen. Following four principles, we provided a formal equation for this model of uncertainty in (5).

We provided two approaches for evaluation of (5). The conservative approach, (6), provides a simple algebraic expression that requires specification of three parameters. The mean-distance approach, (10), requires non-trival numerical computation of the Laguerre function, $L_{1/2}(x)$. The higher computational cost provides the benefit of only two-parameter selection and lowered estimation of σ^2. We provided a computational methodology using Chebyshev polynomial fitting to a one-time computation of (10) to address the computational cost, which is expressed in (20).

From our numerical experiments, we concluded that the use of (20) enables enhanced sensitivity to rapidly changing field values compared to (5) due to lowering of the propagated σ^2. This capability is important for use in the Combined Uncertainty Bathymetry Estimation (CUBE) application (Calder and Mayer 2003) where the best estimation of water depth for navigation safety is crucial. As demonstrated in the numerical experiments, the mean-distance approach provides closer reconstruction of the field than the conservative approach. The conservative approach demonstrated a smoothing behavior in the reconstructed field, but correctly captures the inaccuracy within a higher estimate of output variance.

As elaborated in the "Discussion" section, these approaches have their tradeoffs. Selection depends upon the objectives of end-user. An overall benefit of this work is new computational flexibilities and enhancements for practitioners of this propagated uncertainty model in differing applications.

Acknowledgements

We are grateful to one of the anonymous reviewers for bringing to our attention the potential for (11) as an alternative implementation method.

Funding

This work was supported under National Oceanic and Atmospheric Administration grant NA15NOS4000200, and by the Office of Naval Research under the Naval Research Lab Base Program, both of which are gratefully acknowledged.

References

Abramowicz, M. and I. A. Stegun. 1965. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Mineola, NY: Dover.

Andrews, G., R. Askey, and R. Roy. 1999. *Special Functions,* Volume 71 of *Encyclopedia of Mathematics and its Applications*. Cambridge, UK: Cambridge University Press.

Arfken, G. B., H. J. Weber, and F. E. Harris. 2012. *Mathematical Methods for Physicists: A Comprehensive Guide*. Amsterdam, Netherlands: Academic Press.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017

Bourgeois, B. S., P. A. Elmore, W. E. Avera, and S. J. Zambo. 2016. Achieving comparable uncertainty estimates with Kalman filters or linear smoothers for bathymetry data. *Geochemistry, Geophysics and Geosystems (G3)* 17. doi: 10.1002/2015GC006239

Boyd, J. P. 2001. *Chebyshev and Fourier Spectral Methods*. Mineola, NY: Dover.

Buchholz, H. 1969. *The Confluent Hypergeometric Function,* Volume 15 of *Springer Tracks in Natural Philosophy*. Berlin/Heidelberg: Springer-Verlag.

Calder, B. R. and L. A. Mayer. 2003. Automatic processing of high-rate, high-density multibeam echosounder data. *Geochemistry, Geophysics and Geosystems (G3)* 4(6). doi: 10.1029/2002GC000486

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi. 2009. *GNU Scientific Library Reference Manual* (3 ed.). Network Theory Ltd.

Herrmann, R. 2014. *Fractional Calculus* (2 ed.). Singapore: World Scientific Publishing Company.

International Hydrographic Bureau. 2008. IHO standards for hydrographic surveys, 5ed. Technical Report S-44, International Hydrographic Bureau, 4, quai Antoine 1er, B.P. 445-MC 98011 MONACO Cedex.

Lopez, J. L. and P. S. Ester. 2014. New series expansions for the confluent hypergeometric function M (a, b, z). *Applied Mathematics and Computation* 235:26–31.

Muller, K. E. 2001. Computing the confluent hypergeometric function, *m*(*a, b, x*). *Numerische Mathematik* 90:179–196.

National Institute of Standards and Technology. 2015. NIST digital library of mathematical functions. http://dlmf.nist.gov/, Release 1.0.10 of 2015-08-07. Online companion to Olver et al. (2010).

Olver, F. W. J., D. W. Lozier, R. F. Boisvert, and C. W. Clark. 2010. *NIST Handbook of Mathematical Functions*. New York, NY: Cambridge University Press. Print companion to National Institute of Standards and Technology (2015).

Pearson, J. W., S. Olver, and M. A. Porter. 2014. Numerical methods for the computation of the confluent and Gauss hypergeometric functions. *ArXiv e-prints*.

Press, W. H., S. A. Teukolsky, Q. T. Vetterling, and B. P. Flannery. 2007. *Numerical Recipies* (3rd ed.). Cambridge, UK: Cambridge University Press.

Rice, S. O. 1945. Mathematical analysis of random noise. *Bell System Technical Journal* 24(1):46–156.

Slater, L. J. 1953. On the evaluation of the confluent hypergeometric function. *Proceedings of the Cambridge Philosophical Society* 49(4):612–622.

Taylor, B. N. and C. E. Kuyatt. 1994. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Technical Report 1297, National Institute of Standards and Technology.

Thompson, W. J. 1997. *Atlas for Computing Mathematical Functions*. Wiley.

Zambo, S. J., P. A. Elmore, A. L. Perkins, and B. S. Bourgeois. 2015. Uncertainty estimation for sparse data gridding algorithms. In *Proc. U.S. Hydro. Conf.*, National Harbor, M.D. The Hydrographic Society of America.

Downloaded by [University of New Hampshire] at 11:21 12 December 2017