Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data

Dimitris Menemenlis,[1] Paul Fieguth,[2] Carl Wunsch,[1] and Alan Willsky[3]

Abstract. A fast, recently developed, multiscale optimal interpolation algorithm has been adapted to the mapping of hydrographic and other oceanographic data. This algorithm produces solution and error estimates which are consistent with those obtained from exact least squares methods, but at a small fraction of the computational cost. Problems whose solution would be completely impractical using exact least squares, that is, problems with tens or hundreds of thousands of measurements and estimation grid points, can easily be solved on a small workstation using the multiscale algorithm. In contrast to methods previously proposed for solving large least squares problems, our approach provides estimation error statistics while permitting long-range correlations, using all measurements, and permitting arbitrary measurement locations. The multiscale algorithm itself, published elsewhere, is not the focus of this paper. However, the algorithm requires statistical models having a very particular multiscale structure; it is the development of a class of multiscale statistical models, appropriate for oceanographic mapping problems, with which we concern ourselves in this paper. The approach is illustrated by mapping temperature in the northeastern Pacific. The number of hydrographic stations is kept deliberately small to show that multiscale and exact least squares results are comparable. A portion of the data were not used in the analysis; these data serve to test the multiscale estimates. A major advantage of the present approach is the ability to repeat the estimation procedure a large number of times for sensitivity studies, parameter estimation, and model testing. We have made available by anonymous Ftp a set of MATLAB-callable routines which implement the multiscale algorithm and the statistical models developed in this paper.

1. Introduction

As the hydrographic component of the World Ocean Circulation Experiment (WOCE) nears completion, a large number of new hydrographic observations are becoming available. These observations are typical of several modern, global-scale data sets which are commonly used in gridded form in combination with dynamical models in a process sometimes known as "synthesis," "assimilation," or "fusion." With such global-scale data sets, the problems faced by the data analyst are daunting: they include the large number of measurements, the enormous number of estimation grid points, the irregular sampling and varying spatial quality of the measurements, and the lack of spatial stationarity of the ob-

[1]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge.

[2]Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.

[3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge.

served fields. In this paper we address these difficulties by introducing a scheme which permits the production of maps and error estimates at a very modest computational cost.

The oceanographic community has come to rely heavily upon gridded fields of tracer properties (e.g., heat, salinity, oxygen, nitrogen, etc.) for obtaining quantitative estimates of various aspects of the oceanic general circulation. For example, gridded fields are routinely used in numerical ocean modeling for initial and boundary conditions [e.g., *Semtner and Chervin*, 1992] and for comparisons [e.g., *Stammer et al.*, 1996]. Gridded fields are also used to make diagnostic calculations of the general circulation [e.g., *Bogden et al.*, 1993; *Marotzke and Wunsch*, 1993; *Schiller and Willebrand*, 1995]; to understand the production, transformation, spreading, and associated forcing mechanisms of water masses [e.g., *Worthington*, 1981]; to plan experimental surveys [e.g., *Bretherton et al.*, 1976]; etc. Most of these applications require, in addition to the estimates, a quantitative description of the spatially varying reliability of the gridded fields.

Gridded hydrographic fields are typically produced either by ad hoc interpolation methods or, preferably, through objective mapping (or optimal interpolation (OI)) [e.g., *Gandin*, 1965; *Bretherton et al.*, 1976; *Thiébaux and Pedder*, 1987; *Daley*, 1991]. One important advantage of OI over ad hoc interpolation is that OI provides uncertainty estimates. OI estimates can be obtained by solving an equivalent least squares minimization problem, but for the very large data and grid sizes in the global ocean the requisite brute-force matrix inversion is impractical. In practice, OI estimates are often produced with very restricted subsets of the data (thus throwing away useful information), or without error estimates, or both.

Because the solution of large, two-dimensional, least squares problems is of considerable interest to data analysts in the wider physical sciences, a number of efficient least squares solution methods have been previously proposed. Each of these methods, however, has some limitation which is too restrictive in the context of our hydrographic example. Among known algorithms, (1) fast Fourier transform (FFT) methods [e.g., *Chellappa and Kashyap*, 1982] require that the measurements be densely sampled on a regular grid; (2) local methods [e.g., *Thiébaux and Pedder*, 1987, section 2.4] use measurements only in the local vicinity of the estimate; consequently any long-range correlations in the prior model are ignored and information is lost; and (3) hierarchical methods, such as multigrid [e.g., *Hackbusch*, 1995; *McCormick*, 1989], and simple iterative methods, such as conjugate gradient or successive overrelaxation [e.g., *Dahlquist and Bjorck*, 1974], cannot supply estimation

error statistics (except by brute force). Other efficient methods for dealing with large oceanographic problems in multidimensional spaces have been proposed [e.g., *Bennett*, 1992, section 1.7; *Wunsch*, 1996, section 5.3, and references therein], but these methods do not directly address the gridding problem.

Here we describe an estimation scheme which provides results comparable to those obtained from exact OI, but at a fraction of the computational cost. The scheme is based on a "multiscale" algorithm similar to that used by *Fieguth et al.* [1995] for the analysis of TOPEX/POSEIDON satellite altimetry data. The term "multiscale" here refers to a hierarchical description of the statistical process under study (which, as in the present scheme, does not necessarily imply a description of the statistical process at multiple resolutions). *Fieguth et al.* [1995] chose a scalar description for each state and picked model parameters to match the observed spectral characteristics of the altimeter data. We have adapted their algorithm for use with hydrographic data by developing a totally different statistical model which, while remaining consistent with the multiscale structure, allows the specification of arbitrary correlation functions. The speed of this new approach makes possible sensitivity studies, parameter estimation, and model testing, all of which require a large number of repetitions of the estimation algorithm.

The approach is illustrated by mapping recent hydrographic data in the region of the Pacific Ocean depicted on Figure 1. The choice of this particular region is motivated by an ongoing effort to study oceanic climate and climate drift using acoustic tomography (the Acoustic Thermometry of Ocean Climate (ATOC) project). Specifically, our goal is to obtain estimates of the mean and covariance of the sound speed field for comparison with ATOC results and for initializing time-dependent estimation studies [*Menemenlis and Wunsch*, 1997]. Temperature, which is of more immediate oceanographic interest, is used as a proxy for sound speed [*Munk et al.*, 1995].

Figure 1.

The remainder of this paper is organized as follows. Section 2 is a discussion of statistical models and the standard OI algorithm. The multiscale approach is described in general terms in section 3, with an emphasis on algorithmic differences from the approach used by *Fieguth et al.* [1995]. Section 4 provides a hydrographic mapping example: three-dimensional temperature maps are obtained using the multiscale approach, multiscale and exact OI results are compared, and the multiscale estimates are checked against an independent data set. A discussion of the relative merits and limitations of the multiscale approach follows. Finally, a set of MATLAB-callable multiscale OI routines which implement the algorithm discussed in the manuscript are

described in Appendix A. These routines are available on an anonymous ftp server for the convenience of the oceanographic community.

2. Optimal Interpolation

Consider a set of noisy measurements, represented by a column vector $\mathbf{y}$, which are linearly related to some random process, column vector $\mathbf{x}$, so that

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \tag{1}$$

The matrix $\mathbf{H}$ is the measurement model and $\mathbf{n}$ is the measurement error. Without loss of generality, we consider the case where known biases and correlations of $\mathbf{x}$ and $\mathbf{n}$ have been removed, that is,

$$\langle \mathbf{x} \rangle = \langle \mathbf{n} \rangle = \langle \mathbf{x}\mathbf{n}^T \rangle = 0. \tag{2}$$

The second moments of $\mathbf{x}$ and $\mathbf{n}$ are specified by covariance matrices

$$\mathbf{S} = \langle \mathbf{x}\mathbf{x}^T \rangle, \qquad \mathbf{R} = \langle \mathbf{n}\mathbf{n}^T \rangle, \tag{3}$$

respectively. Optimal interpolation is commonly defined [*Daley*, 1991, section 4.2] as the interpolation which produces the minimum variance solution, $\widehat{\mathbf{x}}$, i.e., the solution that minimizes the individual diagonal elements of the uncertainty matrix,

$$\mathbf{P} \equiv \langle (\widehat{\mathbf{x}} - \mathbf{x})(\widehat{\mathbf{x}} - \mathbf{x})^T \rangle. \tag{4}$$

One form of the solution is

$$\widehat{\mathbf{x}} = \mathbf{P}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} \tag{5}$$

with

$$\mathbf{P} = (\mathbf{S}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}. \tag{6}$$

An alternate form is

$$\widehat{\mathbf{x}} = \mathbf{S}\mathbf{H}^T(\mathbf{H}\mathbf{S}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{y}, \tag{7}$$

$$\mathbf{P} = \mathbf{S} - \mathbf{S}\mathbf{H}^T(\mathbf{H}\mathbf{S}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{S}. \tag{8}$$

It is readily shown [e.g., *Wunsch*, 1996] that least squares minimization of the objective function,

$$J = \mathbf{x}^T\mathbf{S}^{-1}\mathbf{x} + \mathbf{n}^T\mathbf{R}^{-1}\mathbf{n}, \tag{9}$$

produces the same value of $\hat{\mathbf{x}}$ as does OI. For Gaussian fields, $\widehat{\mathbf{x}}$ is the maximum likelihood solution as well. Solutions (5)–(6) or (7)–(8) can be used interchangeably depending on their respective computational cost (which is a function of the relative dimensions of $\mathbf{x}$ and $\mathbf{y}$). Nevertheless, both forms require the inversion or multiplication of matrices that have dimensions of the

number of measurements, or of the number of estimates, or both.

The current generation of desktop workstations can routinely invert matrices with dimensions of a few thousand. For example, the storage of a dense 2000 by 2000 matrix, i.e., the uncertainty matrix of a 45 by 45 estimation grid, requires 32 Mbytes and the conventional inversion of such a matrix requires on the order of 16 Gflops, or 1.5 hours processing time on a SPARC-2 workstation. Storage requirements grow as the matrix dimension squared and the floating point operations grow as the matrix dimension cubed. For the basin-scale mapping studies which are of interest here, one needs to map tens of thousands of measurements onto tens of thousands of grid points, and the brute-force solution of the exact OI equations, (5)–(8), is impractical.

3. Multiscale Method

The multiscale approach is based on a stochastic process $\mathbf{z}(s)$, modeled on a tree, which satisfies the statistical recursion

$$\mathbf{z}(s) = \mathbf{A}(s)\,\mathbf{z}(s\bar{\gamma}) + \mathbf{B}(s)\,\mathbf{w}(s), \qquad (10)$$

$$\mathbf{y}(s) = \mathbf{C}(s)\,\mathbf{z}(s) + \mathbf{v}(s). \qquad (11)$$

Here s indexes the nodes of the multiscale tree as in Figure 2, $s\bar{\gamma}$ represents the parent node of s, $\mathbf{w}(s)$ and $\mathbf{v}(s)$ are white-noise processes with covariance $\mathbf{I}$ and $\mathbf{R}(s)$, respectively, $\mathbf{y}(s)$ represents the measurement process, and $\mathbf{A}(s)$, $\mathbf{B}(s)$, and $\mathbf{C}(s)$ are matrices to be defined later. Each node s on the tree is associated with tree level $m(s)$, where the level counts from zero at the root node $m(0) = 0$ and increases to finer scales. The stochastic process $\mathbf{z}(s)$ also satisfies

Figure 2.

$$\langle \mathbf{z}(0) \rangle = 0, \qquad \langle \mathbf{z}(0)\mathbf{z}(0)^T \rangle = \mathbf{P}_o, \qquad (12)$$

at the root (or coarsest) node of the tree. Each parent node has four "children," $s\alpha_1, \ldots, s\alpha_4$, so that each successive level on the tree has 4 times as many nodes as the previous level.

This particular structure (10)–(12) is motivated by the fact that it admits a very fast sequential estimation algorithm. Readers familiar with the Kalman filter will recognize in (10)–(12) the state-space description of a linear system, but now written in terms of scale rather than time. The noise processes $\mathbf{w}(s)$ and $\mathbf{v}(s)$ are analogous to the Kalman system noise and measurement noise, respectively. The estimation algorithm can therefore be implemented in two sweeps of the multiscale tree. First, a fine-to-coarse recursion, a generalization of the Kalman filter, results in the calculation at each node s of the best linear estimate of $\mathbf{z}(s)$ based on

all the data in the subtree below s. Next a coarse-to-fine recursion, a generalization of the Rauch-Tung-Striebel (RTS) smoother [*Rauch et al.*, 1965], produces the best estimate, $\widehat{\mathbf{z}}(s)$, and error covariance,

$$\mathbf{P}(s) = \langle [\widehat{\mathbf{z}}(s) - \mathbf{z}(s)][\widehat{\mathbf{z}}(s) - \mathbf{z}(s)]^T \rangle, \qquad (13)$$

at each node of the tree based on all of the measurements on the tree. Our algorithm differs from the standard Kalman filter and RTS smoother in that our generalization of the Kalman filter is initialized at the finest scale whereas the model, (10)–(12), is defined from coarse to fine scales. Details are given by *Chou et al.* [1994] or in the software (Appendix A). Briefly, the algorithm has the following properties: the computational effort per tree node is proportional to the cube of the dimension of $\mathbf{z}(s)$, and the resulting estimates are exact; that is, the estimation problem (10)–(12) is solved exactly. The covariance of the entire process, i.e., all terms of the form

$$\langle [\widehat{\mathbf{z}}(s_i) - \mathbf{z}(s_i)][\widehat{\mathbf{z}}(s_j) - \mathbf{z}(s_j)]^T \rangle, \qquad (14)$$

is not automatically computed, but specific terms of interest can be calculated efficiently [*Luettgen and Willsky*, 1995].

We represent the hydrographic process, $\mathbf{x}$, at the finest scale of the tree: the finest scale defines an $N \times N$ grid, and each measurement is associated with the nearest grid point. Therefore the measurement matrix $\mathbf{C}(s)$ in (11), which relates the state vector $\mathbf{z}(s)$ to observations $\mathbf{y}(s)$, is a sparse matrix with $\mathbf{C}(s) = 1$ at those grid points where measurements are available. The fundamental challenge then is the selection of appropriate matrices $\mathbf{A}(s)$ and $\mathbf{B}(s)$ in (10) such that nodes at the finest scale of the tree possess the desired (or close to the desired) statistical covariance $\mathbf{S}$ in (3).

Multiscale processes (10)–(11) were first applied to oceanographic estimation problems by *Fieguth et al.* [1995], who used a class of relatively simple $1/f$-like models in which coarse-scale nodes essentially represent coarse averages of the fine-scale process of interest. These $1/f$-like models do not, however, generalize to the types of prior models of interest here, so we propose an alternative class of models motivated by the method of canonical correlations [*Irving et al.*, 1994; *Irving*, 1995], leading to models in which coarse-scale nodes possess abstract interpretations, serving primarily to produce the desired fine-scale statistics.

Specifically, we propose a model in which the state $\mathbf{z}(s)$, at each node s, equals a subset of the process $\mathbf{x}$ sampled along the perimeters of the children of s (as illustrated in Figure 3). That is,

$$\mathbf{z}(s) = \mathbf{W}(s)\,\mathbf{x}, \qquad (15)$$

Figure 3.

where $\mathbf{W}(s)$ is a matrix, sampling every k^{th} pixel of $\mathbf{x}$ along the perimeters of the children of s, where $k = l/\rho$: l is the correlation length of $\mathbf{x}$ and ρ is a parameter under user control. In the limit where $\mathbf{W}(s)$ samples every pixel of x along the boundaries of the children of node s, only first-order Markov processes can be expected to be modeled exactly; however, we have found that the choice of multiscale model outlined above gives excellent results for a variety of monotonic correlation functions, as long as the samples $\mathbf{W}(s)\,\mathbf{x}$ are spaced by somewhat less than the correlation length of the prior statistics $\mathbf{S}$. Although $\mathbf{z}(s)$ is a hierarchical process, from (15) and Figure 3 it is clear that $\mathbf{z}(s)$ does not contain a multiresolution representation of $\mathbf{x}$; that is, $\mathbf{z}(s)$ does not model the process $\mathbf{x}$ of interest at multiple resolutions.

Once $\mathbf{W}(s)$ has been chosen for each tree node, the multiscale model follows immediately:

$$\mathbf{A}(s) = \left[\mathbf{W}(s)\mathbf{S}\mathbf{W}^T(s\bar{\gamma})\right]\left[\mathbf{W}(s\bar{\gamma})\mathbf{S}\mathbf{W}^T(s\bar{\gamma})\right]^{-1}, \quad (16)$$

$$\mathbf{B}(s)\mathbf{B}^T(s) = \left\langle \mathbf{z}(s)\mathbf{z}^T(s)\right\rangle - \mathbf{A}(s)\left\langle \mathbf{z}(s\bar{\gamma})\mathbf{z}^T(s\bar{\gamma})\right\rangle \mathbf{A}^T(s). \quad (17)$$

This class of models leads to the following trade-off, under explicit control of the user via parameter ρ: the more densely $\mathbf{W}(s)$ samples $\mathbf{x}$ along the perimeters of the children of s, i.e., the larger the value of ρ and the higher the dimension of $\mathbf{z}(s)$, the greater the statistical fidelity of the resulting estimates and posterior error, but the greater the computational burden. Statistical fidelity here refers to the degree to which the multiscale model is a faithful approximation of the desired prior model $\mathbf{S}$. Thus in the event that $\mathbf{S}$ were only approximate, the user might opt for a relatively small state dimension (i.e., low value of ρ) for rapid estimation.

The next section will illustrate the application of this multiscale approach for Gaussian correlation functions and will compare the accuracy of the multiscale approach to exact least squares methods.

4. Hydrographic Example

We illustrate and compare the two mapping algorithms discussed in sections 2 and 3 by constructing a recent temperature climatology in the Pacific Ocean, 175°–250°E, 5°–50°N. The data are from 976 high-resolution vertical temperature profiles obtained within the last 10 years (Figure 1). The profiles retained are of known and consistent quality and span the water column in regions where the ocean depth is greater than 1600 dbar. There is further discussion of the data and data sources in Appendix B.

4.1. Methodology

Our approach is similar to that used by *Fukumori and Wunsch* [1991] and by *Bindoff and Wunsch* [1992]

in that we make use of singular value decomposition to obtain a set of vertical empirical orthogonal functions (EOFs). The EOF decomposition reduces the three-dimensional problem to a two-dimensional mapping exercise for each vertical EOF. We start by projecting temperature onto 35 standard depths (0:50:300, 400:100:1500, 1750:250:5000, 5500, 6000 dbar). The first depth is assigned the shallowest measurement of each cast. The remaining samples are obtained by averaging the high-resolution profiles 20 dbar above and below each standard depth. The resulting temperature profiles are normalized by subtracting the mean and dividing by the standard deviation at each depth (Figure 4). Unless otherwise noted, potential temperature is reported, and depths are in units of pressure.

Figure 4.

A matrix $\mathbf{D}$ is then constructed so that each column corresponds to a particular hydrographic station and each row to a standard depth. Standard depths below the bottom are padded with zeros to make all columns of identical length. (By construction, $\mathbf{D}$ has zero mean at each depth, so this corresponds to replacing values below the bottom with the mean value at that depth. Possible problems of this approach, with suggested solutions, are discussed by *Fukumori and Wunsch* [1991]). By the singular value decomposition, $\mathbf{D}$ is decomposed as $\mathbf{D} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix and the columns of $\mathbf{U}$ and of $\mathbf{V}$ are the vertical and the horizontal EOFs of the hydrographic data set, respectively. The diagonal elements of $\boldsymbol{\Lambda}$ are called singular values and their square measures the contribution of each corresponding pair of EOFs to the variance of $\mathbf{D}$. The EOFs are ordered so that each successive EOF explains less variance than the preceding one. The singular values and the cumulative explained variance are displayed on Figure 5. Note the rapid drop-off of the singular values; for example, the first five EOFs explain more than 95% of the variance of $\mathbf{D}$. The first 13 vertical EOFs are displayed on Figure 6, and, in general, the higher modes are seen to be associated with progressively higher vertical wavenumbers. The statistics of each EOF are uncorrelated and therefore horizontal maps for the coefficients of each vertical EOF can be computed in parallel. However, before applying the OI methods described in sections 2 and 3, the second statistical moments of the signal and the noise, $\mathbf{S}$ and $\mathbf{R}$ in (3), need to be specified.

Figure 5.

Figure 6.

4.2. Determination of a priori Statistics

Strictly speaking, a priori statistics must be obtained independently from the data set used in the OI analysis. In practice, however, independent a priori statistics are seldom available and they must be determined from the data using adaptive filter methods [e.g., *Blanchet et al.*, 1997] or otherwise. Our approach here is one of trial

and error whereby we seek a simple set of statistical assumptions which are consistent with the available data and we use a small fraction of the available degrees of freedom (3 out of 976) to estimate the signal variance, the noise variance, and a correlation length scale. This trial and error approach is made possible by the efficiency of the multiscale interpolation algorithm which allows the estimation procedure to be repeated a large number of times.

The simplest model consistent with the available data is that of a stationary field x with horizontally isotropic Gaussian covariance,

$$\mathrm{Cov}(\mathbf{x}) = \sigma^2 \exp[-(r/l)^2], \tag{18}$$

where r is the horizontal spatial separation, σ^2 is the signal variance, and l is the correlation length scale. This particular form is chosen to represent the correlation structure both because the associated spectrum is everywhere positive and because the resulting covariance matrix, $\mathbf{S}$ in (3), is positive definite; that is, all the eigenvalues of the covariance matrix are positive (see discussion by *Bretherton et al.* [1976]). The measurement noise is modeled as white and horizontally homogeneous with variance n^2. Signal and noise are assumed uncorrelated so that the data covariance can be written as the sum of the field and noise covariances,

$$\mathrm{Cov}(\mathbf{y}) = \sigma^2 \exp[-(r/l)^2] + n^2, \tag{19}$$

consistent with (1) and (2). These assumptions were tested a posteriori to verify that they are consistent with the data.

Estimates of σ^2 and l were obtained by least squares fit of (18) to the data for $r > 0$. The noise variance is estimated as the difference between the measurement variance and σ^2. Figure 7 displays the signal to noise ratio and the correlation length scale associated with vertical EOFs 1 through 15. Note that in general EOFs with larger vertical scales have larger horizontal correlation length scales. For the present analysis we have, conservatively, chosen to consider EOFs greater than 13 as part of the noise. (13 EOFs explain more than 99% of the data variance, and the signal to noise ratio and correlation length scales are both relatively small for EOFs greater than 6.)

Figure 7.

4.3. Multiscale Optimal Interpolation

We can use the second-order priori statistics specified above to compute OI estimates and uncertainty. Specifically, the prior statistics determine $\mathbf{S}$, which in turn determines the multiscale model via (15)–(17), once the sampling parameter ρ has been specified by the user. We will map the coefficients of each of the 13 vertical EOFs on a regular 256 by 256 grid; that is, our tree

has nine levels, and the coarsest (root) node on the tree represents a 256×256 pixel or $75^\circ \times 75^\circ$ area. We will illustrate the trade-off between statistical fidelity and computational burden, resulting from particular choices of ρ, by comparing multiscale and exact OI results.

Since it is impossible to solve the exact OI equations, (5)–(8), at all 65,000 grid locations, to directly compare the multiscale and the exact OI algorithms, we have deliberately limited the computation of estimates to a small number of locations (along WOCE line P2, at the locations marked by open circles on Figure 1). Figures 8 and 9 compare exact (i.e., by brute-force matrix inversion) and multiscale OI results for the fourth vertical EOF. (EOF 4 was chosen because it happens to be particularly effective at illustrating the effect of varying the sampling density parameter ρ.) The multiscale estimates of Figure 8 were obtained using a low sampling density, $\rho = 0.2$, and required 56 Mflops, or 40 s of processing time on a SPARC-2. Even at this low sampling density, the multiscale estimates are mostly contained within the 95% confidence interval of the exact OI estimates. Figure 9 shows results for $\rho = 0.8$, 165 Mflops, or 60 s on a SPARC-2. At this higher sampling density, the multiscale results are almost indistinguishable from those of the exact OI algorithm. It must be emphasized that the computation times quoted here refer to computing multiscale estimates and error statistics, comparable in quality to exact OI, on the whole 65,000 point grid, not just the few samples shown in Figures 8 and 9.

Figure 8.

Figure 9.

4.4. Temperature Map and Uncertainty

After choosing appropriate values for the sampling density parameters, $\rho = 2.3$ for EOF 1 and $\rho = 0.8$ for EOFs 2–13, we computed multiscale OI estimates on a 256 by 256 grid for the first 13 vertical EOFs. The gridded EOF coefficients were then combined to obtain a three-dimensional temperature map and the associated uncertainty variance. A horizontal slice at 400 dbar through the resulting fields is displayed on Figure 10. The uncertainty map was obtained under the assumption that the vertical EOFs are statistically uncorrelated so that the total uncertainty variance at a particular location is the sum of contributions from each EOF.

Figure 10.

4.5. Oceanographic Model Testing

There are many statistical models, e.g., (18), which are consistent with the available data. The best we can do is to show that our results are statistically consistent with our assumptions and with data which were not used in the analysis. We have compared the resulting temperature estimates with data from the WOCE hy-

drographic section P2 at the locations marked by open circles on Figure 10. These data were used neither during the construction of the oceanographic model nor during the OI analysis. Figure 11 compares the multiscale OI estimates with the test data. The estimates are everywhere consistent with the data within 2 standard deviations of the measurement error.

Figure 11.

5. Summary and Concluding Remarks

The goal of the present paper was to demonstrate the use of a new highly efficient multiscale OI algorithm for the mapping of large hydrographic and other oceanographic data sets. We have applied the multiscale algorithm to a limited number of hydrographic profiles to map temperature in the North Pacific region depicted on Figure 1. We also solved the exact OI algorithm at a small number of locations and showed that there is good agreement between the multiscale approach and the exact OI algorithm. The number of measurements used in the present analysis was kept deliberately small to make the above comparison possible. However, in contrast to the exact OI algorithm, the multiscale approach can accommodate much larger data sets and estimation grid points at little additional computational cost.

The particular mapping example discussed in this paper assumes isotropic, horizontally homogeneous prior models, and local measurements with uniform uncorrelated measurement error. But the multiscale approach itself is more general and can be applied to the solution of a wide variety of problems of the form (1)–(9). The challenge lies in selecting appropriate matrices $\mathbf{A}(s)$, $\mathbf{B}(s)$, and $\mathbf{C}(s)$ for the multiscale model (10)–(11). Without further modifications, the multiscale model-building routines which have been made available (see Appendix A) admit anisotropic prior models and nonuniform measurement errors. The multiscale mapping approach has also been applied to oceanographic processes with inhomogeneous prior statistics [*Fieguth et al.*, 1995] and with correlated measurement errors [*Fieguth et al.*, 1997].

Finally, we have shown that the multiscale approach provides estimates of uncertainty variance (the diagonal terms of the uncertainty matrix) which are consistent with the exact OI uncertainty. Although the computation of the full error covariance matrix remains prohibitively expensive, the multiscale approach allows the efficient calculation of specific off-diagonal terms [*Luettgen and Willsky*, 1995]. It is therefore possible to selectively sample the uncertainty matrix and construct a simplified model for the correlated error structures, a step which is crucial to the production of dynamically and observationally consistent analyses. An example is the large-scale correlated errors in surface temper-

ature climatologies and surface buoyancy fluxes which are often used in combination to drive general circulation models. Data assimilation studies require that the different correlated errors be known and used in order to achieve acceptable solutions.

Appendix A: MATLAB-Callable Optimal Interpolation Code

The implementation of the multiscale estimation algorithm [*Chou et al.*, 1994; *Fieguth et al.*, 1995] in its full generality is a rather complicated undertaking. In the interest of promoting the use of this algorithm and enabling interested researchers to apply it to problems of their own, we are making this code publicly available. Two programs are provided, one of which is tailored specifically to reproduce the results shown in section 4, and the other of which is more general and permits researchers to apply the multiscale estimation method of this paper to problems of their own.

The code is written in MATLAB scripts and in C: the front-end visible to the user is written in MATLAB, and the multiscale computational engine is written in C. No programming experience is needed to use the software, although at least a rudimentary understanding of MATLAB scripts would be required to customize our programs for a different application.

Anyone interested in compiling and running our code will require MATLAB 4.x software and an ANSI-compatible C compiler (precompiled versions of the code, not requiring any compilation, are available for Sun-SPARC platforms). Most workstations should have ample computational power; all of our results were computed on a Sun-SPARC 2 with 48 Mbytes of core memory. A large multiscale tree requires a large amount of memory. At least 20 Mbytes is required to run the programs for small test cases; 128 Mbytes or more is recommended for serious research applications.

The programs may be obtained via anonymous Ftp to lids.mit.edu (IP Address 18.78.0.101) from directory pub/ssg/code/Hydrography. The file README describes the purpose of each program and how to get started.

Appendix B: Hydrographic Data Set

Hydrographic data used in this paper include a subset of WOCE lines P1, P2, P3, P4, P14, P16, and P17. Additional data sources include the 1991–1993 INPOC trans-Pacific lines, the 1986–1987 reciprocal tomography experiment, and the 1984 North Pacific section along 175° W by the Woods Hole Oceanographic Institution conductivity-temperature-depth (CTD) group. Most of the data, cruise reports, and other useful information is available on anonymous ftp nemo.ucsd.edu and on the world-wide-web at www.cms.udel.edu. Preliminary WOCE line P2 data were kindly made available to us by Hideki Kinoshita of the Japan Hydrographic Office.

Acknowledgments. We would like to acknowledge the extensive comments provided by the anonymous reviewers which led to a variety of improvements in the paper. Financial support was provided by SERDP/ARPA as part of the ATOC project (University of California SIO contract PO 10037358), by NASA grant NAGW–1048, by ONR grant N00014–91–J–1004, and by NSF grant DMS–9316624.

References

Bennett, A. F., *Inverse Methods in Physical Oceanography*, *Cambridge Monographs on Mechanics and Applied Mathematics*, Cambridge Univ. Press, New York, 1992.

Bindoff, N. L., and C. Wunsch, Comparison of synoptic and climatologically mapped sections in the South Pacific Ocean, *J. Clim.*, *5*(6), 631–645, 1992.

Blanchet, I., C. Frankignoul, and M. A. Cane, A comparison of adaptive Kalman filters for a tropical Pacific Ocean model, *Mon. Weather Rev.*, *125*, 40–58, 1997.

Bogden, P. S., R. E. Davis, and R. Salmon, The North Atlantic circulation: Combining simplified dynamics with hydrographic data, *J. Mar. Res.*, *51*, 1–52, 1993.

Bretherton, F. P., R. E. Davis, and C. B. Fandry, A technique for objective analysis and design of oceanographic experiments applied to MODE-73, *Deep Sea Res.*, *23*, 559–582, 1976.

Chellappa, R., and R. Kashyap, Digital image restoration using spatial interaction models, *IEEE Trans. Acoust. Speech Signal Process.*, *30*(3), 461–472, 1982.

Chou, K., A. Willsky, and A. Benveniste, Multiscale recursive estimation, data fusion, and regularization, *IEEE Trans. Autom. Control*, *39*(3), 464–478, 1994.

Dahlquist, G., and A. Bjorck, *Numerical Methods*, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

Daley, R., *Atmospheric Data Analysis*, *Cambridge Atmospheric and Space Science Series*, Cambridge Univ. Press, New York, 1991.

Fieguth, P. W., W. C. Karl, A. S. Willsky, and C. Wunsch, Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry, *IEEE Trans. Geosci. Remote Sens.*, *33*(2), 280–292, 1995.

Fieguth, P. W., D. Menemenlis, T. Ho, A. S. Willsky, and C. Wunsch, Multiresolution mapping of altimeter data in the Mediterranean, *J. Atmos. Oceanic Technol.*, in press, 1997.

Fukumori, I., and C. Wunsch, Efficient representation of the North Atlantic hydrographic and chemical distributions, *Prog. Oceanogr.*, *27*, 111–195, 1991.

Gandin, L. S., *Objective Analysis of Meteorological Fields*, Isr. Program for Sci. Transl., Jerusalem, 1965.

Hackbusch, W., *Multi-Grid Methods and Applications*, Springer-Verlag, New York, 1995.

Irving, W., Multiscale stochastic realization and model identification with applications to large-scale estimation problems, Ph.D. thesis, Mass. Inst. of Technol., Cambridge, 1995.

Irving, W., W. Karl, and A. Willsky, A theory for multiscale stochastic realization, in *33rd Conference on Decision and Control, Orlando, FL*, pp. 655–662, 1994.

Luettgen, M., and A. Willsky, Multiscale smoothing error models, *IEEE Trans. Autom. Control*, *40*(1), 173–175, 1995.

Marotzke, J., and C. Wunsch, Finding the steady state of a general circulation model through data assimilation: Application to the North Atlantic Ocean, *J. Geophys. Res.*, *98*(C11), 20,149–20,167, 1993.

McCormick, S., *Multilevel Adaptive Methods for Partial Differential Equations*, Soc. for Ind. and Appl. Math., Philadelphia, Pa., 1989.

Menemenlis, D., and C. Wunsch, Linearization of an oceanic circulation model for data assimilation and climate studies, *J. Atmos. Oceanic Technol.*, in press, 1997.

Munk, W., P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography*, *Cambridge Monographs on Mechanics*, Cambridge Univ. Press, New York, 1995.

Rauch, H. E., F. Tung, and C. T. Striebel, Maximum likelihood estimates of linear dynamic systems, *AIAA J.*, *3*, 1445–1450, 1965.

Schiller, A., and J. Willebrand, A technique for the determination of surface heat and freshwater fluxes from hydrographic observations, using an approximate adjoint ocean crculation model, *J. Mar. Res.*, *53*, 433–451, 1995.

Semtner, A. J., Jr., and R. M. Chervin, Ocean general circulation from a global eddy-resolving model, *J. Geophys. Res.*, *97*(C4), 5493–5550, 1992.

Stammer, D., R. Tokmakian, A. Semtner, and C. Wunsch, How well does a 1/4° global circulation model simulate large-scale oceanic observations?, *J. Geophys. Res.*, *101*(C11), 25,779–25,811, 1996.

Thiébaux, H. J., and M. A. Pedder, *Spatial Objective Analysis with Applications in Atmospheric Science*, Academic, San Diego, Calif., 1987.

Worthington, L. V., The water masses of the world ocean: Some results of a fine-scale census, in *Evolution of Physical Oceanography*, edited by B. A. Warren, and C. Wunsch, chap. 2, pp. 42–69. MIT Press, Cambridge, Mass., 1981.

Wunsch, C., *The Ocean Circulation Inverse Problem*, Cambridge Univ. Press, New York, 1996.

P. Fieguth, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. (e-mail:pfieguth@sysoffice.watstar.uwaterloo.ca)

D. Menemenlis and C. Wunsch, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139. (e-mail: dimitri@gulf.mit.edu; carl@pond.mit.edu)

A. Willsky, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139. (e-mail: willsky@MIT.EDU)

(Received April 22, 1996; revised February 13, 1997; accepted February 19, 1997.)

Paper number 97JC00697.
0148-0227/97/97JC-00697$09.00

Figure Captions

Figure 1. Location of hydrographic stations relative to bathymetry in the northeastern Pacific Ocean. The solid dots indicate stations which were used in the OI analysis, while the open circles are stations which were left out of the analysis to be used for oceanographic model testing. Bathymetric contour intervals are in meters.

Figure 1. Location of hydrographic stations relative to bathymetry in the northeastern Pacific Ocean. The solid dots indicate stations which were used in the OI analysis, while the open circles are stations which were left out of the analysis to be used for oceanographic model testing. Bathymetric contour intervals are in meters.

Figure 2. Illustration of a hierarchical stochastic process, i.e., a process defined on a tree. The nodes of the tree are indexed by s, $s\bar{\gamma}$ represents the parent node of s, and $s\alpha_1, \ldots, s\alpha_4$ are the children of s. The scale of each node s on the tree is written $m(s)$, where the scale counts from zero at the root node, $m(0) = 0$, and increases to finer scales. As discussed in the text, very fast estimation algorithms exist for such processes.

Figure 2. Illustration of a hierarchical stochastic process, i.e., a process defined on a tree. The nodes of the tree are indexed by s, $s\bar{\gamma}$ represents the parent node of s, and $s\alpha_1, \ldots, s\alpha_4$ are the children of s. The scale of each node s on the tree is written $m(s)$, where the scale counts from zero at the root node, $m(0) = 0$, and increases to finer scales. As discussed in the text, very fast estimation algorithms exist for such processes.

Figure 3. An illustration of the definition for the state at each tree node: the state at each node s is made up of pixels (small circles) sampled along the boundaries of the children of node s.

Figure 3. An illustration of the definition for the state at each tree node: the state at each node s is made up of pixels (small circles) sampled along the boundaries of the children of node s.

Figure 4. Mean and standard deviation of the temperature profiles. The series of lines on the right of the figure indicate the 35 standard depths used to subsample the temperature profiles.

Figure 4. Mean and standard deviation of the temperature profiles. The series of lines on the right of the figure indicate the 35 standard depths used to subsample the temperature profiles.

Figure 5. Singular values (circles) of a matrix whose columns consist of temperature profiles at the locations indicated by dots on Figure 1 and sampled at the 35 standard depths of Figure 4. (Before the computation, the mean temperature was removed and the resulting perturbation was divided by the standard deviation at each depth.) The singular value decomposition also provides a set of empirical orthogonal functions (EOFs), and the associated cumulative explained variance is indicated in percent (asterisks).

Figure 5. Singular values (circles) of a matrix whose columns consist of temperature profiles at the locations indicated by dots on Figure 1 and sampled at the 35 standard depths of Figure 4. (Before the computation, the mean temperature was removed and the resulting perturbation was divided by the standard deviation at each depth.) The singular value decomposition also provides a set of empirical orthogonal functions (EOFs), and the associated cumulative explained variance is indicated in percent (asterisks).

Figure 6. Normalized vertical structure of EOFs 1 through 13.

Figure 6. Normalized vertical structure of EOFs 1 through 13.

Figure 7. A priori signal and noise statistics for vertical EOFs 1 through 15. (a) Signal to noise ratio, σ^2/n^2. (b) Characteristic correlation length scale, l.

Figure 7. A priori signal and noise statistics for vertical EOFs 1 through 15. (a) Signal to noise ratio, σ^2/n^2. (b) Characteristic correlation length scale, l.

Figure 8. Comparison of multiscale and exact OI results for EOF 4 at the locations indicated by open circles on Figure 1. (a) Estimates. (b) Standard deviation uncertainty. The multiscale estimates were first computed on a 256 by 256 grid and then interpolated onto the section. The sampling density parameter was set to $\rho = 0.2$ and the multiscale estimation required 56 Mflops, or 40 s of processing time on a SPARC-2. Even at this low sampling density, the multiscale estimates are mostly contained within the 95% confidence interval of the exact OI estimates. With increased computational effort, the discontinuities in the estimates can be reduced, as shown on Figure 9.

Figure 8. Comparison of multiscale and exact OI results for EOF 4 at the locations indicated by open circles on Figure 1. (a) Estimates. (b) Standard deviation uncertainty. The multiscale estimates were first computed on a 256 by 256 grid and then interpolated onto the section. The sampling density parameter was set to $\rho = 0.2$ and the multiscale estimation required 56 Mflops, or 40 s of processing time on a SPARC-2. Even at this low sampling density, the multiscale estimates are mostly contained within the 95% confidence interval of the exact OI estimates. With increased computational effort, the discontinuities in the estimates can be reduced, as shown on Figure 9.

Figure 9. Same as in Figure 8 but the multiscale estimates were obtained using a higher sampling density, $\rho = 0.8$, which required 165 Mflops, or 60 s on a SPARC-2. For this particular choice of ρ the multiscale estimates are almost indistinguishable from exact OI results.

Figure 9. Same as in Figure 8 but the multiscale estimates were obtained using a higher sampling density, $\rho = 0.8$, which required 165 Mflops, or 60 s on a SPARC-2. For this particular choice of ρ the multiscale estimates are almost indistinguishable from exact OI results.

Figure 10. Multiscale OI maps of potential temperature and uncertainty at 400 dbar. (a) Temperature contoured at 1°C intervals. (b) Standard deviation uncertainty contoured at 0.2°C intervals. The maps represent sums of multiscale estimates and uncertainty variance of the first 13 vertical EOFs. The solid dots indicate the location of hydrographic data used in the analysis, while the open circles indicate WOCE line P2 data used later for checking the oceanographic model.

Figure 10. Multiscale OI maps of potential temperature and uncertainty at 400 dbar. (a) Temperature contoured at 1°C intervals. (b) Standard deviation uncertainty contoured at 0.2°C intervals. The maps represent sums of multiscale estimates and uncertainty variance of the first 13 vertical EOFs. The solid dots indicate the location of hydrographic data used in the analysis, while the open circles indicate WOCE line P2 data used later for checking the oceanographic model.

Figure 11. Comparison of multiscale OI estimates at 400 dbar with data from WOCE hydrographic section P2 which were not used in the analysis. The error bars represent the 95% confidence (± 2 standard deviations) of the measurement error.

Figure 11. Comparison of multiscale OI estimates at 400 dbar with data from WOCE hydrographic section P2 which were not used in the analysis. The error bars represent the 95% confidence (± 2 standard deviations) of the measurement error.

Figures

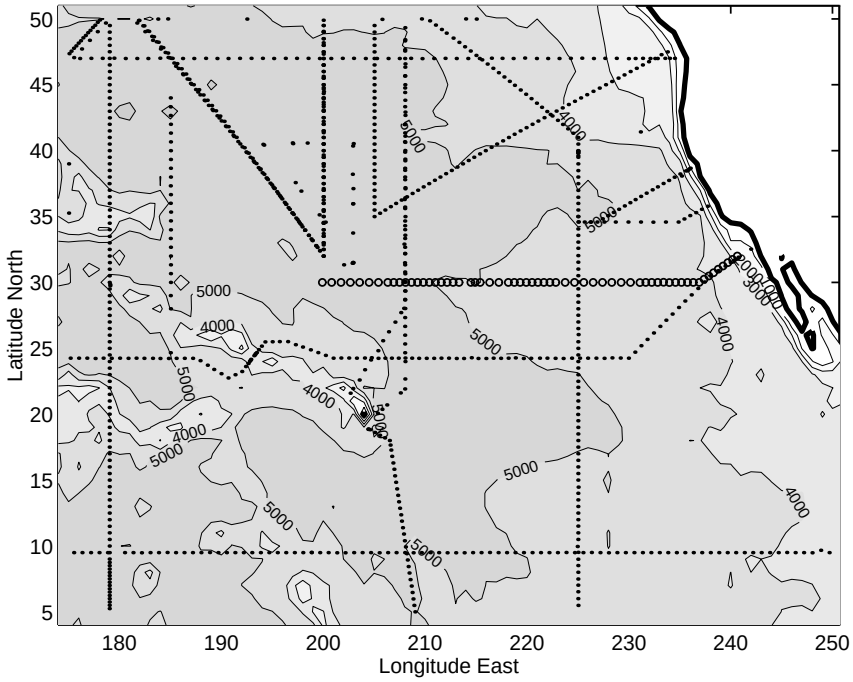

Figure 1. Location of hydrographic stations relative to bathymetry in the northeastern Pacific Ocean. The solid dots indicate stations which were used in the OI analysis, while the open circles are stations which were left out of the analysis to be used for oceanographic model testing. Bathymetric contour intervals are in meters.

Figure 1. Location of hydrographic stations relative to bathymetry in the northeastern Pacific Ocean. The solid dots indicate stations which were used in the OI analysis, while the open circles are stations which were left out of the analysis to be used for oceanographic model testing. Bathymetric contour intervals are in meters.

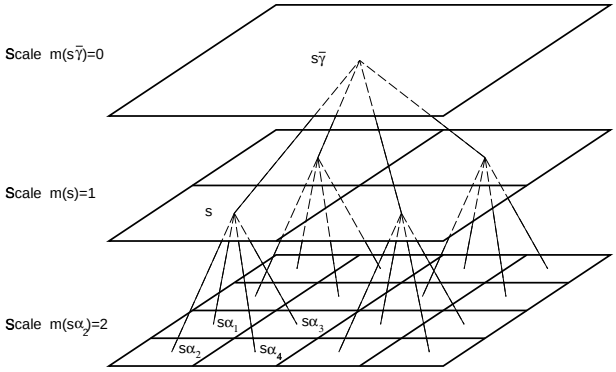

Figure 2. Illustration of a hierarchical stochastic process, i.e., a process defined on a tree. The nodes of the tree are indexed by s, $s\bar{\gamma}$ represents the parent node of s, and $s\alpha_1, \ldots, s\alpha_4$ are the children of s. The scale of each node s on the tree is written $m(s)$, where the scale counts from zero at the root node, $m(0) = 0$, and increases to finer scales. As discussed in the text, very fast estimation algorithms exist for such processes.

Figure 2. Illustration of a hierarchical stochastic process, i.e., a process defined on a tree. The nodes of the tree are indexed by s, $s\bar{\gamma}$ represents the parent node of s, and $s\alpha_1, \ldots, s\alpha_4$ are the children of s. The scale of each node s on the tree is written $m(s)$, where the scale counts from zero at the root node, $m(0) = 0$, and increases to finer scales. As discussed in the text, very fast estimation algorithms exist for such processes.

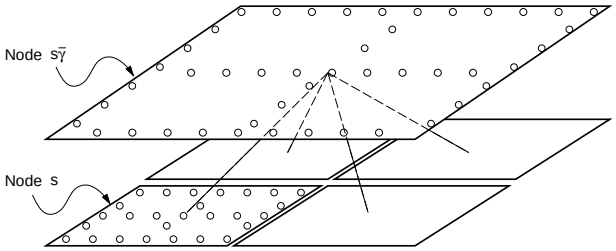

Figure 3. An illustration of the definition for the state at each tree node: the state at each node s is made up of pixels (small circles) sampled along the boundaries of the children of node s.

Figure 3. An illustration of the definition for the state at each tree node: the state at each node s is made up of pixels (small circles) sampled along the boundaries of the children of node s.

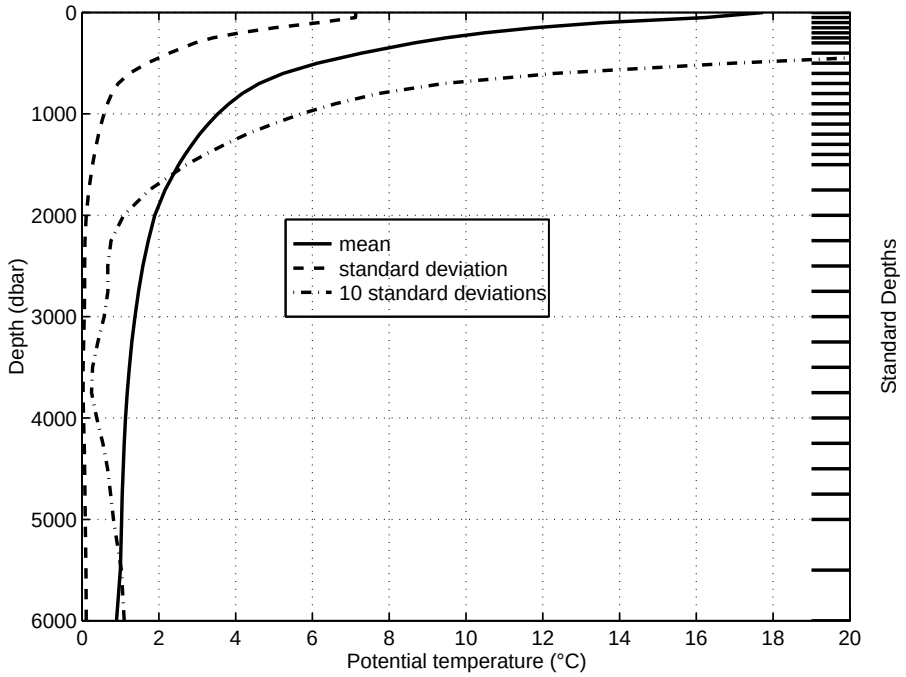

Figure 4. Mean and standard deviation of the temperature profiles. The series of lines on the right of the figure indicate the 35 standard depths used to subsample the temperature profiles.

Figure 4. Mean and standard deviation of the temperature profiles. The series of lines on the right of the figure indicate the 35 standard depths used to subsample the temperature profiles.

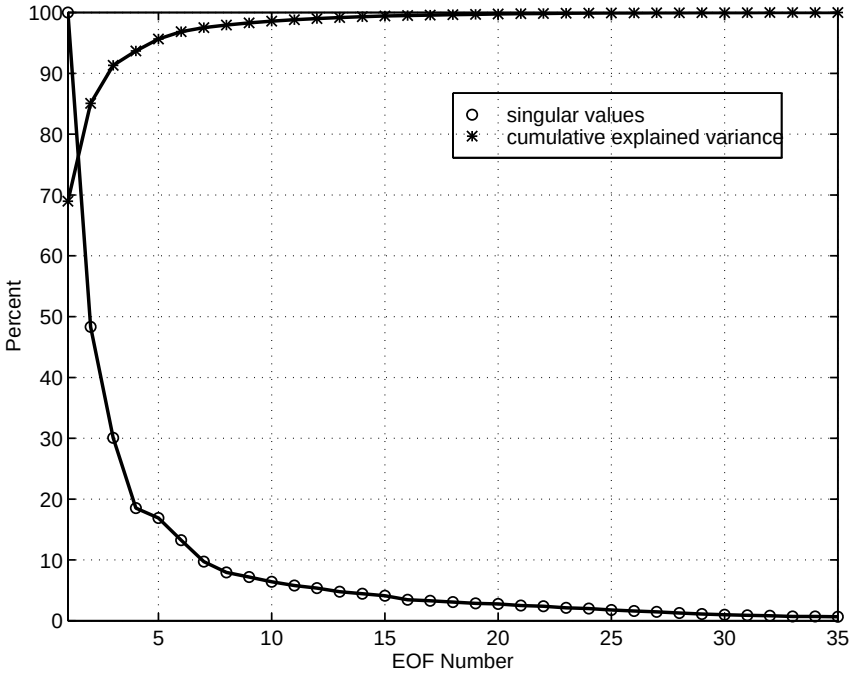

Figure 5. Singular values (circles) of a matrix whose columns consist of temperature profiles at the locations indicated by dots on Figure 1 and sampled at the 35 standard depths of Figure 4. (Before the computation, the mean temperature was removed and the resulting perturbation was divided by the standard deviation at each depth.) The singular value decomposition also provides a set of empirical orthogonal functions (EOFs), and the associated cumulative explained variance is indicated in percent (asterisks).

Figure 5. Singular values (circles) of a matrix whose columns consist of temperature profiles at the locations indicated by dots on Figure 1 and sampled at the 35 standard depths of Figure 4. (Before the computation, the mean temperature was removed and the resulting perturbation was divided by the standard deviation at each depth.) The singular value decomposition also provides a set of empirical orthogonal functions (EOFs), and the associated cumulative explained variance is indicated in percent (asterisks).

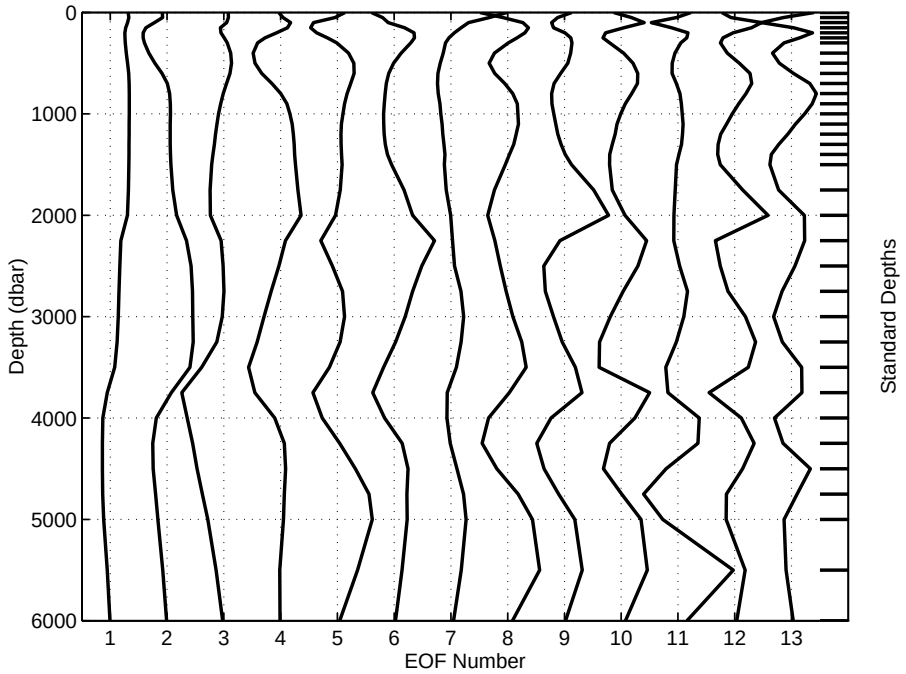

Figure 6. Normalized vertical structure of EOFs 1 through 13.

Figure 6. Normalized vertical structure of EOFs 1 through 13.

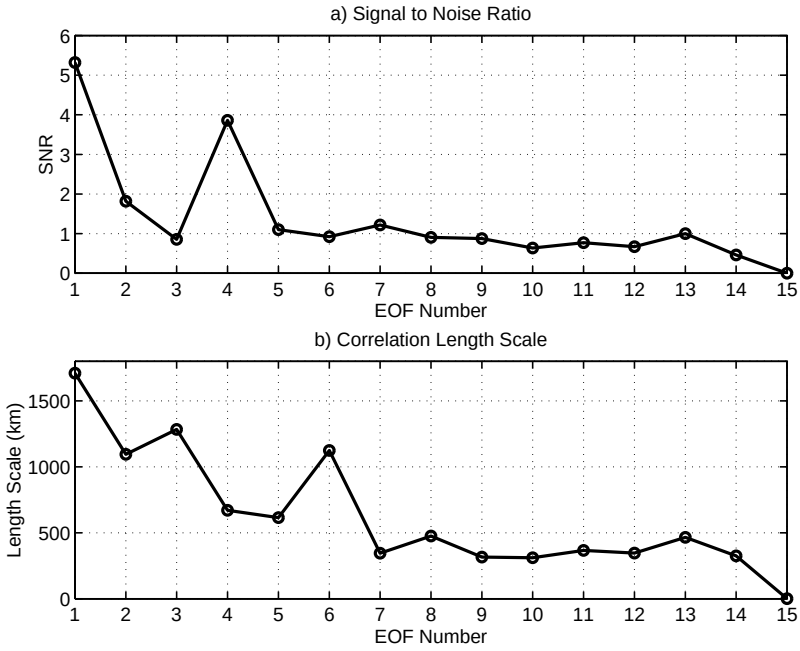

Figure 7. A priori signal and noise statistics for vertical EOFs 1 through 15. (a) Signal to noise ratio, σ^2/n^2. (b) Characteristic correlation length scale, l.

Figure 7. A priori signal and noise statistics for vertical EOFs 1 through 15. (a) Signal to noise ratio, σ^2/n^2. (b) Characteristic correlation length scale, l.

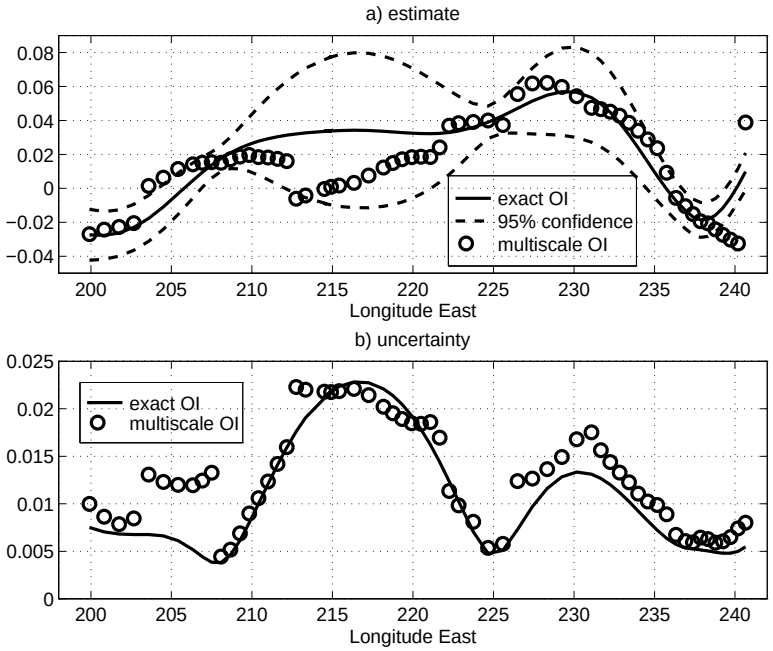

Figure 8. Comparison of multiscale and exact OI results for EOF 4 at the locations indicated by open circles on Figure 1. (a) Estimates. (b) Standard deviation uncertainty. The multiscale estimates were first computed on a 256 by 256 grid and then interpolated onto the section. The sampling density parameter was set to $\rho = 0.2$ and the multiscale estimation required 56 Mflops, or 40 s of processing time on a SPARC-2. Even at this low sampling density, the multiscale estimates are mostly contained within the 95% confidence interval of the exact OI estimates. With increased computational effort, the discontinuities in the estimates can be reduced, as shown on Figure 9.

Figure 8. Comparison of multiscale and exact OI results for EOF 4 at the locations indicated by open circles on Figure 1. (a) Estimates. (b) Standard deviation uncertainty. The multiscale estimates were first computed on a 256 by 256 grid and then interpolated onto the section. The sampling density parameter was set to $\rho = 0.2$ and the multiscale estimation required 56 Mflops, or 40 s of processing time on a SPARC-2. Even at this low sampling density, the multiscale estimates are mostly contained within the 95% confidence interval of the exact OI estimates. With increased computational effort, the discontinuities in the estimates can be reduced, as shown on Figure 9.

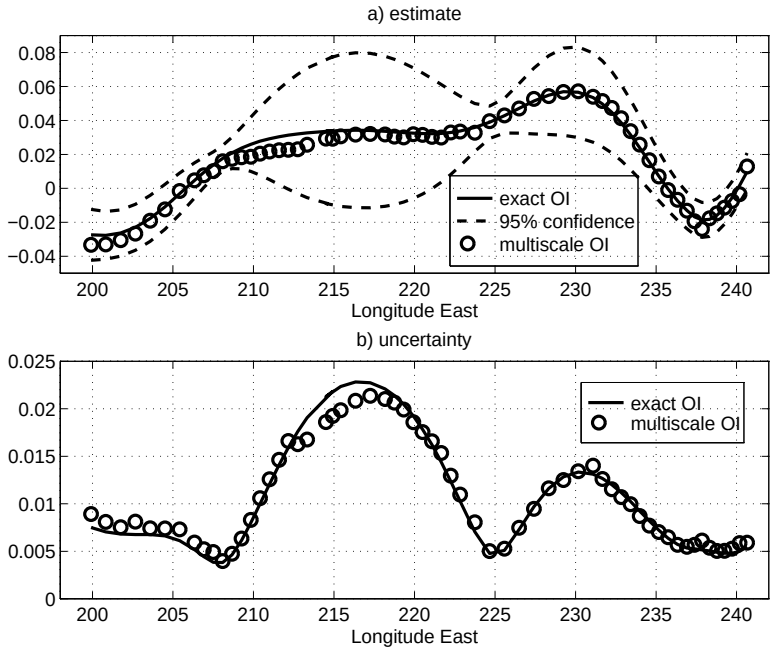

Figure 9. Same as in Figure 8 but the multiscale estimates were obtained using a higher sampling density, $\rho = 0.8$, which required 165 Mflops, or 60 s on a SPARC-2. For this particular choice of ρ the multiscale estimates are almost indistinguishable from exact OI results.

Figure 9. Same as in Figure 8 but the multiscale estimates were obtained using a higher sampling density, $\rho = 0.8$, which required 165 Mflops, or 60 s on a SPARC-2. For this particular choice of ρ the multiscale estimates are almost indistinguishable from exact OI results.

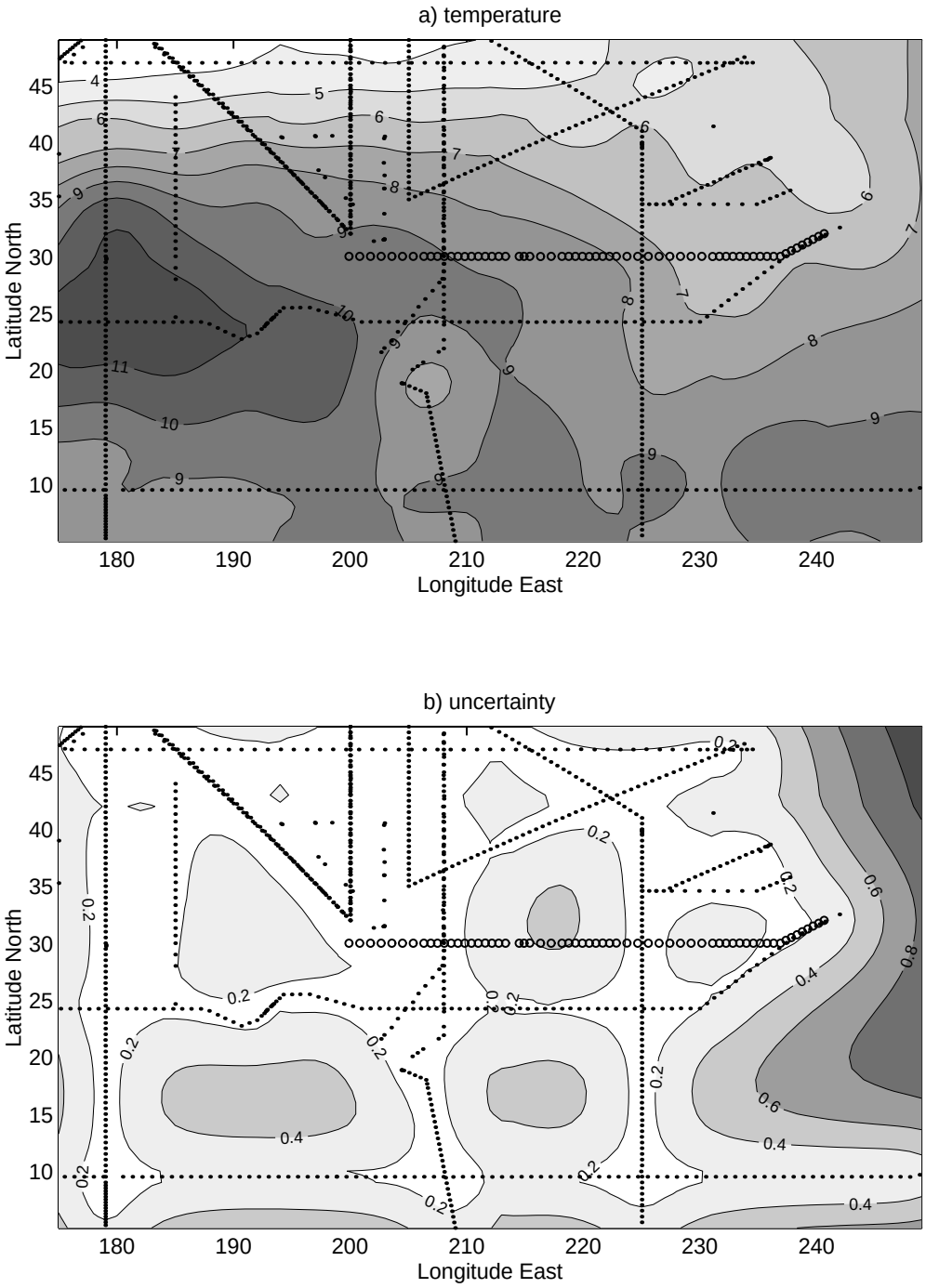

Figure 10. Multiscale OI maps of potential temperature and uncertainty at 400 dbar. (a) Temperature contoured at 1°C intervals. (b) Standard deviation uncertainty contoured at 0.2°C intervals. The maps represent sums of multiscale estimates and uncertainty variance of the first 13 vertical EOFs. The solid dots indicate the location of hydrographic data used in the analysis, while the open circles indicate WOCE line P2 data used later for checking the oceanographic model.

Figure 10. Multiscale OI maps of potential temperature and uncertainty at 400 dbar. (a) Temperature contoured at 1°C intervals. (b) Standard deviation uncertainty contoured at 0.2°C intervals. The maps represent sums of multiscale estimates and uncertainty variance of the first 13 vertical EOFs. The solid dots indicate the location of hydrographic data used in the analysis, while the open circles indicate WOCE line P2 data used later for checking the oceanographic model.

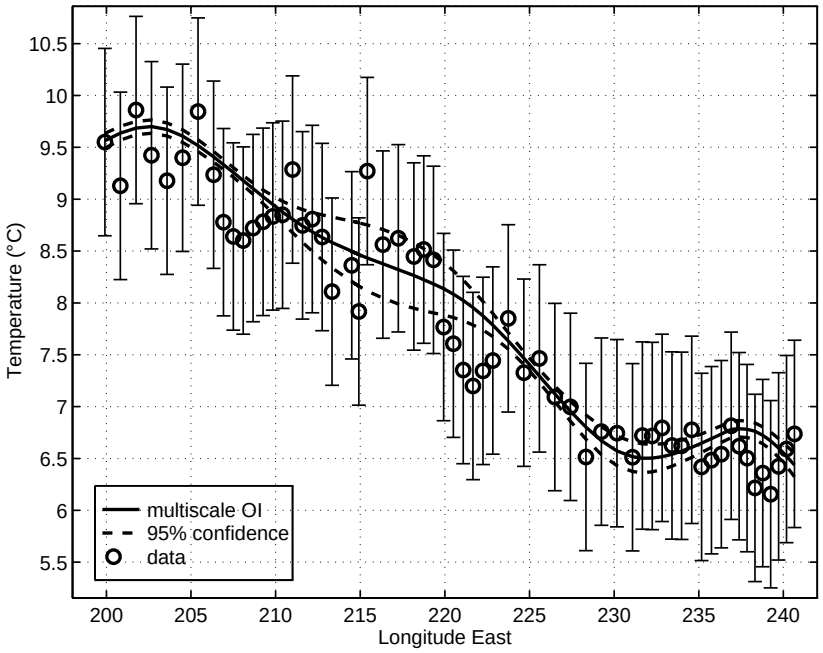

Figure 11. Comparison of multiscale OI estimates at 400 dbar with data from WOCE hydrographic section P2 which were not used in the analysis. The error bars represent the 95% confidence (± 2 standard deviations) of the measurement error.

Figure 11. Comparison of multiscale OI estimates at 400 dbar with data from WOCE hydrographic section P2 which were not used in the analysis. The error bars represent the 95% confidence (± 2 standard deviations) of the measurement error.