Multiresolution Optimal Interpolation and Statistical Analysis of TOPEX/POSEIDON Satellite Altimetry

Paul W. Fieguth[1]

William C. Karl[1]

Alan S. Willsky[1]

Carl Wunsch[2]

October 1, 1994

Abstract

A recently developed multiresolution estimation framework offers the possibility of highly efficient statistical analysis, interpolation, and smoothing of extremely large data sets in a multiscale fashion. This framework enjoys a number of advantages not shared by other statistically-based methods. In particular, the algorithms resulting from this framework have complexity that scales only linearly with problem size, yielding constant complexity load per grid point independent of problem size. Furthermore these algorithms directly provide interpolated estimates at multiple resolutions, accompanying error variance statistics of use in assessing resolution / accuracy trade-offs and in detecting statistically significant anomalies, and maximum likelihood estimates of parameters such as spectral power law coefficients. Moreover, the efficiency of these algorithms is completely insensitive to irregularities in the sampling or spatial distribution of measurements and to heterogeneities in measurement errors or model parameters. For these reasons this approach has the potential of being an effective tool in a variety of remote sensing problems. In this paper, we demonstrate a realization of this potential by applying the multiresolution framework to a problem of considerable current interest – the interpolation and statistical analysis of ocean surface data from the TOPEX / POSEIDON altimeter.

I. INTRODUCTION

This paper describes and uses a new statistically optimal approach to the interpolation and analysis of large, irregularly sampled geophysical data sets. The vehicle which we use to illustrate the methodology is the particular problem of estimating the shape of the ocean surface from satellite altimetry measurements. This application is of considerable current interest both because of its importance in global ocean modeling and climate studies and because of the relatively recent launch of the joint American/French TOPEX/POSEIDON altimeter[7, 16, 27], a satellite-based platform capable of measuring ocean height to an unprecedented accuracy of approximately 5 cm.

The availability of data of this quality and coverage makes it possible to address a variety of scientific questions ranging from producing regularly gridded maps of ocean height (to be used, for example, in global ocean modeling studies) to the accurate estimation of the spatial spectrum of ocean height variations. Achieving objectives such as these, however, presents daunting challenges to the data analyst, in particular in terms of the enormous size of the problems to be solved. This paper presents a methodology that permits the production of statistically optimal results, with computational loads that are extremely modest.

The TOPEX/POSEIDON altimetry and ocean height analysis problem encompasses many of the issues and characteristics of a broader class of geophysical data analysis problems that have motivated our work, and we have used a specific case study here to illustrate these issues as well as our methodology. Figure 1 depicts a region of the northeastern Pacific from Hawaii to Alaska. Overlaid on this region is the distribution of TOPEX/POSEIDON satellite measurements[18] over a typical ten day cycle. Successive measurements along a track are separated by approximately 7km or 0.06 degrees; the spacing between adjacent tracks is approximately 270km. For many reasons gridded images of the ocean are required at fine scales, both in order to observe features of interest, and to produce numerical values compatible with fine scale ocean models. Even for the comparatively modest portion of the ocean shown in Figure 1, we must estimate ocean surface heights at more than $100,000$ grid points based on roughly $20,000$ altimetric measurements. For a full ocean basin, or the entire global surface, the problem is of truly formidable proportions.

However the size of data analysis problems such as this is not their only significant challenge. First of all in many cases, including the one of interest here, significant spatial nonstationarities are present for several possible reasons:

1. The sampling pattern of the data is frequently nonuniform and irregular, including occasional periods of data dropout as shown in Figure 1.

2. The sensed phenomenon is itself nonstationary, exhibiting differing spatial scales and magnitudes of variability in different regions. Ocean surface statistics, for example, differ between regions containing vigorous currents such as the Kuroshio or the Gulf stream and those regions which are comparatively quiet such as the northeast Pacific.

3. The *quality* of measurements may also be nonstationary. In particular, the TOPEX/POSEIDON altimeter provides direct measurements of the distance from the satellite to the ocean surface. What is actually desired, however, is a measurement of ocean height relative to the gravitational equipotential surface[22] known as the geoid; the ocean current field may be inferred

from the derivative of this relative surface. The complex and nonstationary error structure of the geoid model[21] thus translates directly into nonstationary errors in the corrected altimetry measurements.

Such nonstationarities or irregularities in the data pattern present a major challenge[8, 32], as there is no regular structure that can be used to advantage. In particular Fourier methods, with their noteworthy efficiencies, cannot be applied directly or without significant approximations and idealizations.

Furthermore, in addition to the estimation of quantities such as ocean surface height, there are compelling reasons for desiring a characterization of the errors in these estimates. In particular, to assess the value of a set of estimates we must have a measure of their accuracy, requiring at the very least the calculation of error variances. Moreover, there are strong motivations for the characterization of the spatial correlation structure in the estimation errors. For example, the assimilation of ocean surface estimates into global circulation models[8], which effects a blending of the surface measurements and the underlying science, in principle requires the full specification of the error correlations so that accurate model/data combinations can be effected.

In addition, error covariance calculations are useful for a variety of other scientific reasons. For example, geoid estimates have errors due to unresolved, spatially localized perturbations such as sea mounts or trenches. Such errors can manifest themselves as outliers in the data, or more precisely in the residuals (data minus estimates); the availability of error statistics permits the identification of statistically significant outliers and the estimation of localized geoid corrections implied by these residuals.

However the brute-force explicit computation and storage of a dense $100,000 \times 100,000$ estimation error covariance matrix is entirely impractical. Moreover, other statistical calculations place equally severe demands on data analysis. In particular, an important problem in oceanography which has received considerable attention, but remains of current interest is that of estimating the spatial statistical structure of ocean height variations - e.g., the structure of the spectral density of ocean height as a function of wavenumber. Frequently parametric models for such spectra are posited[8], but the statistically optimal estimation of these parameters, e.g., using methods of maximum likelihood estimation, and the associated specification of the accuracy in these estimates also present significant computational challenges.

Finally an important characteristic of many remote sensing problems, including the one examined here, is that the phenomenon under study exhibits behavior across a broad range of scales. For example, global ocean models predict behavior at (and interactions among) a vast range of spatial scales. Indeed, models for ocean height spectra[8] are typically described in terms of inverse power-law relationships. Such a spectral description corresponds directly to a scaling relationship between the expected amplitude and spatial scale of ocean features - i.e., it corresponds to a fractal model. Statistical modeling of the ocean surface and the processing of ocean height data must account for this multi-scale structure.

A number of smoothing and data assimilation algorithms (e.g., objective analysis[5], kriging[25]) have been developed, each of which has emphasized varying degrees of statistical structure or computational efficiency. The combination of the issues we have mentioned – problem size, non-

stationarity, statistical characterization of errors, and accounting for correlation structures over a range of scales – has generally required that compromises be made in the statistical consistency and optimality of the results. The method that we describe and illustrate here avoids the need to make such compromises.

The key to this new approach is that we begin by focusing explicitly on scale. In particular, rather than starting with the statistical description of the phenomenon to be estimated at a single, fine scale of resolution we describe its statistical structure at a hierarchy of scales. That is, we directly capture the multi-scale correlation structure of the phenomenon through a *scale-recursive* statistical model, proceeding from coarser representations to finer ones. With such a model we can meet the challenges presented in all of the issues mentioned previously. In particular, the algorithm that we describe directly deals with the issue of scale, has a total computational complexity per grid point independent of the size of the grid, can accommodate nonstationarities in the model of the phenomenon or the data, and allows the complete characterization of error statistics and the calculation of maximum likelihood parameter estimates. The results presented later in this paper for the processing of data over the region shown in Figure 1 take seconds on a current generation single processor workstation. Moreover, our approach produces estimates at a hierarchy of scales, facilitating resolution/accuracy tradeoffs and the direct extraction of estimates of coarser scale features. Furthermore, although we do not use it here, this framework also supports the fusion of data of differing resolution and coverage with no change in algorithmic structure.

Section II summarizes the general problem of optimal estimation of interest here in order to establish terminology and to specify the precise statistical problems to be solved. Section III contains an outline of our multiscale estimation framework, a detailed description of which is contained in the appendix. Section IV describes the TOPEX/POSEIDON altimetric interpolation problem and uses it to illustrate our methods and the types of scientific questions which maybe addressed. Section V lists conclusions and summarizes ongoing efforts.

II. STATISTICAL MODELS AND OPTIMAL ESTIMATION

The starting point is the basic problem of estimating a collection of random variables, represented abstractly by the vector $\boldsymbol{x}$, based on a set of noise-corrupted measurements, represented by $\boldsymbol{y}$:

$$\boldsymbol{y} = C\boldsymbol{x} + \boldsymbol{v} \qquad E\left[\boldsymbol{v}\right] = 0 \qquad E\left[\boldsymbol{v}\boldsymbol{x}^T\right] = 0 \qquad E\left[\boldsymbol{v}\boldsymbol{v}^T\right] = R \qquad (1)$$

where $\boldsymbol{v}$ represents the measurement noise or error. For the problem to be considered in Section IV the components of $\boldsymbol{v}$ are assumed to be uncorrelated but possibly with non-constant variances – i.e., R is diagonal but not a multiple of the identity. The matrix C describes the nature of the measurement process. For the problems considered here, the components of $\boldsymbol{x}$ includes a full grid of ocean height values, and C is a "selection matrix" indicating which of the components of $\boldsymbol{x}$ are measured and which x_i corresponds to each y_j.

We can view our estimation problem as estimating the deviations of $\boldsymbol{x}$ from its mean, thus for simplicity, we assume that $\boldsymbol{x}$ is zero-mean and has prior covariance

$$E\left[\boldsymbol{x}\boldsymbol{x}^T\right] = P_x \qquad (2)$$

For problems of substantial size, the explicit specification of the correlation structure of x through the full covariance matrix P_x is neither feasible nor useful unless P_x is extremely sparse with known structure – e.g., if P_x is banded, implying only local correlation among the components of x. However, such sparse or banded structures are not particularly appropriate or useful for problems of interest here, as we are interested in representing phenomena possessing correlations at many (and not just local) scales. Furthermore, as we will see, banded or sparse covariance structures do not necessarily lead to simple algorithms for statistical data analysis.

Consequently we are led instead to construct an implicit model of the statistical structure of x of the form

$$M\boldsymbol{x} = \boldsymbol{w} \tag{3}$$

$$P_x^{-1} = M^T P_w^{-1} M \tag{4}$$

where P_w is the covariance of $\boldsymbol{w}$. There are several reasons why representations as in (3),(4) can be attractive. One is that processes with complex correlation structures can be represented in a very compact manner. For example consider the linear state space model

$$x(t+1) = Ax(t) + w(t) \qquad E\left[x(0)w^T(t)\right] = 0 \tag{5}$$

If we construct the vectors

$$\boldsymbol{x}^T = \begin{bmatrix} x^T(0) & x^T(1) & x^T(2) & \dots \end{bmatrix} \qquad \boldsymbol{w}^T = \begin{bmatrix} x^T(0) & w^T(0) & w^T(1) & \dots \end{bmatrix} \tag{6}$$

then we obtain a representation as in (3) with P_w block diagonal and M lower bidiagonal:

$$M = \begin{bmatrix} I & 0 & 0 & 0 & \dots \\ -A & I & 0 & 0 & \dots \\ 0 & -A & I & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{7}$$

As we now show, it is the *inverse* of P_x, which according to (4) involves only M and P_w, that is critical in constructing solutions to optimal estimation problems.

Specifically, the problem of interest here is the computation of the minimum variance linear estimate of x based on y, as well as a statistical characterization of the error $\tilde{x} = x - \hat{x}$. There are numerous ways in which to represent the solution to this problem, but the one that is most convenient for our discussion is that given by the normal equations for this least squares problem:

$$(P_x^{-1} + C^T R^{-1} C)\hat{\boldsymbol{x}} = C^T R^{-1} \boldsymbol{y} \tag{8}$$

This problem formulation and the normal equation solution are well known in many disciplines including geophysical problems of optimal interpolation[9], where,as we have indicated, approximations or suboptimal solutions have generally been required in order to deal with the issues we have described, To understand some of these, consider the formal explicit solution, $\hat{x}$, to (8) and the resulting error covariance $P_{\tilde{x}}$:

$$\hat{\boldsymbol{x}} = L\boldsymbol{y} = P_{\tilde{x}} C^T R^{-1} \boldsymbol{y} \tag{9}$$

$$P_{\tilde{x}}^{-1} = P_x^{-1} + C^T R^{-1} C \tag{10}$$

Note that if P_x has a sparse or banded structure, indicative of local correlations, this structure is *not* generally preserved either in the estimation gain matrix L (10) or in the estimator error covariance $P_{\tilde{x}}$. Thus simple, local, smoothing algorithms (e.g., local least squares, local interpolation) while efficient computationally, generally represent a suboptimal approximation to (10) even in situations in which they appear to be best matched, i.e., when the field to be interpolated has local correlations. Moreover, a very important point is that the statistical structure of the resulting estimation error field, $P_{\tilde{x}}$, is *not* local, despite locality in P_x. Furthermore the calculation of $P_{\tilde{x}}$ is generally prohibitively complex (since, in particular, the inversion of the prior covariance P_x is extremely demanding). Thus the use of simple local algorithms generally involves a compromise in statistical consistency, in the explicit and faithful use of prior statistical models and information, in the calculation of accurate error statistics, and in the ability to account for correlations at many scales.

The situation looks much different, however, if we examine the normal equations (8) directly. If we begin with an implicit model for x as in (3) – or equivalently with a decomposition of P_x^{-1} as in (4) with M and P_w^{-1} having sparse or local structure – then from (10) we see that this structure is maintained in $P_{\tilde{x}}^{-1}$ and in the normal equations. In particular, since the measurements are point measurements of components of x with uncorrelated errors – so that C is a selection matrix and R diagonal – then $C^T R^{-1} C$ is also diagonal, so that $P_{\tilde{x}}^{-1} = P_x^{-1} + C^T R^{-1} C$ maintains the same structure as P_x^{-1}.

The significance of these observations is considerable. For example, for the time-recursive state space model (5) with local measurements, i.e.,

$$y(t) = Hx(t) + v(t) \tag{11}$$

we have from (4) and (7) that P_x^{-1} is block tridiagonal, a structure that is shared by $P_{\tilde{x}}^{-1}$. As a consequence, the normal equations can be solved in an extremely efficient fashion, namely Gaussian elimination – also known as the Kalman filter[1] – followed by back-substitution – known as the Rauch-Tung-Striebel (RTS) smoothing algorithm[23]. Furthermore in the process of performing these calculations we directly compute the diagonal elements of $P_{\tilde{x}}$ – i.e., the estimation error covariance matrices for $x(t)$ for each value of t. Moreover, perhaps less widely known, these calculations also yield a model for $\tilde{x}$ without any additional work. In particular since $P_{\tilde{x}}^{-1}$ has the same structure as P_x^{-1}, we might hope to model $\tilde{x}$ as

$$\tilde{M}\tilde{x} = \tilde{w} \tag{12}$$

where $\tilde{w}$ is block diagonal and $\tilde{M}$ has the same structure as M in (7) – i.e., so that $\tilde{x}$ has a time-recursive model as in (5). Such a model does in fact exist, and its parameters are directly and very simply computable from the original model (5) parameters and from the error covariances computed by the Kalman filter and RTS smoother.

Furthermore, since we have a model (12) for the estimation errors in this time-recursive statistical estimation problem, we can use the measurement residuals

$$\tilde{y} = y - C\hat{x} = C\tilde{x} + v \tag{13}$$

to detect statistically significant deviations from the assumed statistics. In addition, the recursive Kalman filter algorithm allows whitening of the data y and thus the efficient computation of like-

lihood functions, leading to statistically optimal methods for estimating parameters of the model (e.g., parameters embedded in M, P_w, C, and R).

The critical question, then, is whether we can find analogous classes of models for phenomena that vary in space rather than time, i.e., models that have a similar set of properties and that also allow us to capture rich classes of spatial phenomena including those with multiple correlation scales. One class of such models that has been widely proposed used is the class of Markov random fields (MRF's). As discussed in [11], such fields have models as in (3) in which M is an elliptic (symmetric, positive definite) partial difference operator and where $P_w = M$. In this case $P_x^{-1} = M$, emphasizing the correspondence between models and inverse covariances. Furthermore such models can capture multiple correlation scales. Moreover $\tilde{M} = P_{\tilde{x}}^{-1}$ in (10) is also an operator of the same structure as M so that subsequent data assimilation stages, in which the error statistics at one stage form the prior model for the next, face structurally identical estimation problems. The normal equations in this case correspond to an elliptic partial differential equation and the error covariance to the inverse of an elliptic operator. Consequently the required computations for estimation, error covariance calculation, anomaly detections, and likelihood evaluation are not simple and can in fact be prohibitively complex except in the case of stationary models and data (so that Fourier techniques can be applied).

The next section describes an alternative to MRF's for the modeling of random fields that overcomes these difficulties through the use of scale-recursive models, permitting the realization of the full set of advantages found for the time-recursive state model (5).

III. MULTIRESOLUTION MODEL APPROACH

The multiscale models of interest in this paper and originally introduced in [3, 13] are scale-recursive models defined on index sets that are organized as multilevel trees. A simple example of such a tree for a 2-D random field is illustrated in Figure 2. Here each level of the tree corresponds to a different scale of resolution in the representation of the random field, with coarser scales toward the top of the tree, and where the components of x correspond to variables defined at the various nodes of the tree. This modeling framework is more flexible than the figure might suggest however, because it is applicable to higher dimensional trees or to asymmetric and unusually shaped trees. This flexibility can be used to match the particular multiscale structure of the phenomenon being modeled or to capture local differences in scale structure (e.g., if the field has finer scale details in particular regions). For the purposes of this paper the quadtree structure of Figure 2 will suffice.

The specific model class of interest here is inspired by the successes of the time-recursive model (5). In particular, if s denotes any node on the tree and $s\bar{\gamma}$ its parent, then the components of x at these nodes are related by a coarse-to-fine recursion:

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s) \tag{14}$$

where $w(s)$ is a white noise process with identity covariance. Moreover, the general measurement model associated with this framework also allows measurements at multiple scales:

$$y(s) = C(s)x(s) + v(s) \tag{15}$$

where $v(s)$ is white, with covariance $R(s)$. In the applications considered here the measurements are all at the finest scale – i.e., at a sparse and irregular subset of nodes at the lowest level on the tree – and we will focus principally on the estimates at this finest scale as well. However, the statistical algorithm for the model (14),(15) can handle data at multiple resolutions and produces estimates (and error statistics) at all scales.

Optimal estimation, error model characterization, data whitening and likelihood calculation have extremely efficient realizations for this class of multiscale models. These efficiencies are a result of the structure of the tree and the model (14),(15) which leads to a divide-and-conquer structure for statistical analysis: conditioned on any node on the tree, each of the subtrees connected to this node are conditionally independent (for example, conditioned on the top node in Figure 2, each of the four distinct subtrees below this node are conditionally independent).

Thus for any node s the processing of the data in the subtree beneath it can be decomposed into independent processing of the data in each of the descendant subtrees. For example, as illustrated in Figure 3, optimal estimation of x (i.e., the collection of all $x(s)$'s) based on y (all $y(s)$'s) can be implemented as two sweeps on the tree. The fine-to-coarse sweep generalizes the Kalman filter and results in the calculation at each node s of the best linear estimate of $x(s)$ based on all of the data in the subtree below s; next a coarse-to-fine sweep generalizes the RTS algorithm and produces the best estimate and error variances at every node based on all of the data.

The resulting algorithm, which is described in detail in the appendix, involves only local calculations following the structure of the tree. Thus calculations for each node are performed once each on the upward and downward sweeps. Furthermore, if N denotes the number of nodes at the finest scale of the tree, i.e., the number of pixels at the finest scale of resolution, then the total number of nodes on the tree is $\frac{4}{3}N$. Thus the total complexity of the algorithm is proportional to N, resulting in constant complexity per grid point independent of the size of the grid. Moreover, these same calculations yield a model for the error $\tilde{x}(s)$ which has a multiscale form[14], so that subsequent data assimilation stages can be carried out in exactly the same fashion. Specifically,

$$\tilde{x}^s(s) \;=\; P(s\mid s)P_s^{-1}A(s)P_{s\bar{\gamma}}P^{-1}(s\bar{\gamma}\mid s)\tilde{x}^s(s\bar{\gamma}) + \breve{w}(s) \tag{16}$$

where P_s is the prior covariance at node s and $P(s\mid\sigma)$ represents the error covariance of $x(s)$ given all observations in the subtree below node σ. As detailed in the appendix, these error covariances are also calculated via scale-recursive generalizations of the Kalman filter Riccati equation and RTS error covariance calculation. Once again these algorithms have constant per pixel complexity. Furthermore, a closely related algorithm based on the Kalman filter allows us to whiten the data and compute likelihoods in an equally efficient fashion[13].

In addition to the computational efficiencies admitted by these multiscale models, they also can be used to capture the statistical structure of rich classes of phenomena. For example, in [12] it is shown that multiscale models as in (14),(15) can be constructed to represent phenomena frequently modeled using MRF's. Of more direct importance here is the fact that this multiscale framework is directly suited to capturing phenomena that display a multitude of correlation scales. Of particular interest here is the class of so-called $1/f$ models[30], i.e., processes that display $1/f^{\mu}$-like spectra over a significant range of frequencies. Such models are closely related to fractals and fractional Brownian motions[31] and are widely used to describe a vast array of natural phenomena. In particular, such models are frequently used to describe the spatial structure of the ocean surface.

For example, Figure 4 (from [8]) shows a typical power spectrum for the ocean surface, modeled as a $1/f^\mu$-process with different values of μ over different wavenumber intervals.

Phenomena with $1/f^\mu$-like spectra display so-called self-similar scaling properties in that the variability of such a phenomenon scales geometrically with the spatial resolution at which the variations are measured. Such scaling rules are captured very simply in our multiscale model through the imposition of a scaling relationship in the gain $B(s)$ in (14). For example, if we let $m(s)$ denote the scale of a node s on the tree of Figure 2 – i.e., the level corresponding to that node, with $m(s) = 0$ at the coarsest scale and $m(s)$ increasing as we move to finer scales, then the choice

$$A(s) = 1 \qquad B(s) = B_o 2^{(1-\mu)m(s)/2} \cdot I \qquad (17)$$

displays the same scaling behavior as that implied by a $1/f^\mu$ spectrum.[31] Changes in scaling laws, corresponding for example to the changes in slope in Figure 4, can be captured simply by changing the value of μ over different ranges of scale. Local changes in scaling structure can also be easily accommodated by local modifications of $B(s)$.

IV. RESULTS

A. Multiscale Model Selection

The experimental results of this section are based upon one year of TOPEX/POSEIDON altimetric data[10]. In addition to subtracting the geoidal reference field[21, 22], the usual corrections are applied to the data: ionospheric[17], tidal[24], orbital[18], and atmospheric pressure loading.

With the multiscale framework outlined in the previous section, a first task is the determination of the order of the model (i.e., the dimension of $x(s)$) and the specific model parameters (e.g., the $A(s)$, $B(s)$, $C(s)$ values of (14),(15)). For the purposes of this paper, $x(s)$ is a scalar representing the ocean height at the particular scale and position corresponding to node s. The determination of the scalars $A(s), B(s)$ is then made by choosing these parameters to match the observed spectral characteristics of TOPEX/POSEIDON data. As we have mentioned previously, it is possible to use a more sophisticated statistical procedure to determine optimal estimates for these parameters, and such an investigation, using the properties of our multiresolution models to advantage, is currently underway. However, performing spectral matching is a widely-used method (e.g., [8]), and for the purposes of this paper this simpler fashion is sufficient.

Figure 5 shows a periodogram determined from TOPEX/POSEIDON data. The top spectrum of Figure 5 falls as $1/f^2$, which from (17) leads to choices of the multiscale model parameters of the form

$$A(s) = 1 \qquad B(s) = B_o 2^{-m(s)/2} \qquad (18)$$

This then gives us the correct spectral roll-off or slope for the spectrum of our multiscale process. The corresponding offset or power level of our multiscale model can then be set by choosing B_o to match the TOPEX/POSEIDON spectrum. The resulting value, $B_o = 35cm$, corresponds to the power spectrum shown in the bottom half of Figure 5. The resulting multiscale model

corresponding to (14) is

$$x(s) = x(s\bar{\gamma}) + 35 \cdot 2^{-m(s)/2} w(s) \tag{19}$$

That is, the aggregate surface height of the ocean at some position and scale equals the aggregate height of its parent node, i.e., at the same spatial position but at a coarser scale, plus a perturbation offset whose variance decreases geometrically with scale.

The prior variance P_s of $x(s)$ at each node of the tree can be determined from a recursion obtainable directly from (14):

$$P_s = E\left[x(s)x^T(s)\right] = A(s)P_{s\bar{\gamma}}A^T(s) + B(s)B^T(s) \tag{20}$$

The recursion is initialized with the prior variance P_o of $x(0)$ at the root node of the tree. Roughly speaking this variance can be thought of as specifying the prior level of uncertainly in the aggregate mean height of the ocean. In this paper, in order to avoid biasing our estimate of overall ocean height, we have set P_o is set to be very large ($\approx 10^5$).

The measurement model is straightforward, since our observations are direct measurements of the $\{x(s)\}$ on the finest scale of the tree, i.e., C is a selection matrix. If the finest scale occurs at scale $m(s) = M$, then

$$C(s) = \begin{cases} 0 & m(s) < M \quad \text{or } x(s) \text{ does not correspond to a TP observation.} \\ 1 & m(s) = M \quad \text{and } x(s) \text{ corresponds to a TP observation point.} \end{cases} \tag{21}$$

The final parameter that needs to be specified is the measurement noise variance $R(s)$. In particular, in this study we have accounted for two significant sources of error in the measurement data:

- The error in estimating the distance from the satellite to the ocean surface, assumed to be 5cm white Gaussian noise.

- The error in the geoid model, which manifests itself as an error in the geoid-corrected TOPEX/POSEIDON data.

The highest quality geoid models currently available[21, 19], are quite effective at capturing large scale and moderate scale geoid fluctuations, but are less accurate in regions of sharp local changes. Such a result is not surprising: geoid models are typically constructed as truncated spherical harmonic expansions, which can exhibit larger errors and Gibbs-like phenomena near abrupt changes. Furthermore, navigation errors in the satellite lead to errors in registering satellite measurements with points on the earth, and thus in areas of steep geoid gradient, such registration errors lead to greater uncertainty in the geoid reference field than in other regions in which the geoid is smoothly varying. As a result, altimetric measurements in the vicinity of steep geoid slopes are determined relative to a poor geoid reference and therefore represent a less accurate assessment of the ocean surface height. Consequently we have used the following measurement noise variance model:

$$R(s) = (5\text{cm})^2 + \phi(\text{Geoid Slope}) \tag{22}$$

where $\phi()$ is an increasing function (to be detailed in the next section).

Finally, it is important to make a comment about one of the consequences of using a simple scalar version of our multiscale model. In particular, since the spatial position of the multiscale tree on the ocean is somewhat arbitrary; that is, there is no particularly natural orientation for the multiscale tree, we will want to make sure that the estimates produced by our algorithm are insensitive to the precise positioning of the tree. However, consider a node s t a relatively coarse scale on the tree. Since the state at each node is a scalar, the correlation between the four children of node s, each of which still represent relatively coarse representations of ocean height, is captured by only one degree of freedom. In particular, the finer scale decompositions of each of these four descendants proceed completely independently, and as a consequence artifacts may appear along coarse tree boundaries due to inadequate correlation. There are several ways in which to avoid these artifacts. One of these involves using higher-order models (i.e., $x(s)$ becomes a vector at each node), as outlined, for example, in [13]. A second approach involving trees in which the nodes at a given scale represent aggregate values over overlapping regions is currently under development[6]. In this paper we use a simpler method that again is adequate for our purposes. Specifically, we compute ocean surface estimates for each of ten tree positions (each shifted with respect to the others) and average the results. It should be made clear that this is not at all like spatial low-pass filtering or interpolation, as strong nonstationarities, such as in the quality of the data as measured by $R(s)$, are maintained.

B. Gridding Results

Given a collection of observations and the multiscale model as defined in the previous section, the multiscale estimation algorithm (detailed in the Appendix) permits rapid computation of multiscale estimates, estimation error variances, and measurement residuals. Each of these computations will be explored in the ocean altimetry context below.

Multiscale Estimates

A sample map of ocean surface estimates, taken from the finest scale of the tree, is shown in Figure 6. This map is based upon a single repeat cycle, or ten days, of data (about 20,000 data points). The 250,000 estimates and associated estimation covariance information were computed in less than one minute on a Sun Sparc-10 (the map is based on ten trees of estimates, each tree requiring 5 seconds of computation time). Although Figure 6 shows estimates on one scale only, the one minute of computer time produces estimates and error variances on all scales of the tree.

The ocean height variations shown in the figure are consistent with the known large-scale oceanographic behavior of the region (that is, a predominant gradient in the north-south direction with surface height offset on the order of one meter[10]). Moreover, the estimates such as those shown in the figure offer far higher resolution than has heretofore been available (e.g., [2]). It is this very leap in resolution that makes the quantitative assessment of our results difficult - we have come across no other altimetric maps of sufficient resolution to compare with our plots. For example, Figure 7 [10] shows an ocean altimetric map for the same region of the ocean and the same period of time as we have considered. The figure, typical of the methods used by oceanographers, is based upon gridding followed by spatial filtering. Thus the thorough validation of the enhanced resolution results provided by our method will require alternate methods such as integration with global circulation models, a problem that remains for the future. Nevertheless, the ability to produce such estimates efficiently is itself of significance.

11

Multiscale Error Variances

Estimation error variances corresponding to Figure 6 are shown in Figure 8. These values are based on the same ten day set of measurements as for the estimates just discussed; the distribution of measurement dropouts along the satellite tracks in this data set can be inferred from Figure 1. As before, the results are computed as the average over ten multiscale trees, still within the same one minute of computer time in which the estimates were computed.

Because of the spatially varying uncertainty in our measurements due to geoid model error, the occurrence of data dropouts, and the irregular pattern of data collection, we would expect that the uncertainty pattern in the optimal estimate of our ocean height map would be highly variable and would, to some extent, reflect these features. In particular, observe that the regions of lowest uncertainty (the lightly shaded regions in the figure) correspond to the points at which we have satellite measurements; a careful inspection of the figure will also reveal occasional darker breaks along these lines, corresponding to data dropouts. In addition, because of the spatially-varying noise model, the measurements near the Aleutian and Hawaiian chains (which induce a significant geoid gradient) are modeled as being noisier, resulting in elevated covariance values. The large region of uncertainty at the top of the figure is due to the Alaskan land mass, over which no oceanographic measurements are taken.

Specific off-diagonal terms in the error covariance matrix may also be computed using (16) with equal computational ease (as compared to other approaches which would require the impractical calculation of the full error covariance matrix, containing $\approx 10^{10}$ elements). For example, by computing error covariances between a large ensemble of tree nodes (here 50,000 pairs of nodes, randomly positioned in longitude) one can determine averaged correlation coefficients in our estimation error as a function of longitudinal separation, as shown in Figure 9.

Multiscale Model Heterogeneities

One of the drawbacks with certain accelerated methods, such as those based on FFTs, is the need for stationarity or uniformity of the phenomenon being modeled. In contrast, the multiscale framework employed in this paper allows us to incorporate nonstationarities without sacrificing computational efficiency.

Consider, for example, the Kuroshio current in the northwest Pacific off the coast of Japan. Due to the strength of this current, the gradient of the ocean surface in the neighborhood of the Kuroshio is approximately four times larger[28] than in relatively quiescent regions (the Pacific northeast, for example). To compensate for this effect, one can modify (18) by increasing those process noise values on those multiscale tree nodes which overlap part of the Kuroshio. Such a process noise is highly nonstationary, and by (20) implies a nonstationary prior covariance model. Since the adjustments to the process noise as discussed above remain compatible with the multiscale framework of (14), (15), not only does our approach remain efficient in the face of such heterogeneities, but the increase in computational burden over the homogeneous case is essentially nil.

Figures 10, 11 show estimates and error variances respectively for the northwest region of the Pacific, using a heterogeneous process noise model as detailed above. The distribution of error variances show the combined effects of irregular spatial sampling by the satellite, loss of satellite measurements over land (Japan), increased prior uncertainty over the Kuroshio, and nonstationary

geoid-model error. For purposes of comparison, Figure 12 shows the differences in the altimetry estimates produced by models with and without Kuroshio compensation.

Calculation of Measurement Residuals

The examination of measurement residuals, the differences between measurements and the corresponding estimates, serves to test the validity of our multiscale models. In particular, by normalizing these residuals with respect to their expected standard deviations one can isolate statistically significant outliers. Such an approach may be used to argue the inclusion of the geoid slope dependent term in the measurement error (22). Figure 13 shows the distribution of statistically large residuals, calculated using a simple measurement noise model

$$R(s) = (5\text{cm})^2 \tag{23}$$

that is, a noise model which does not take any geoid model errors into account. Figure 13 also plots the geoid gradient; the correlation between significant residuals and steep geoid slope is unambiguous, and argues in favor of a geoid slope-corrected measurement noise model. As an additional comparison, the same locations of large residuals are shown superimposed on a plot of ocean bathymetry contours (the shape of the ocean bottom) in the bottom half of the figure. To the extent that bathymetry features are responsible for locally steep slopes in the geoid, the residual-bathymetry correlation does not come as a surprise. Such residual-geoid correlation immediately motivates the development of adaptive models to estimate the geoid or locate unknown bathymetric features; the development of such models is underway.

Figure 14 plots root mean square estimation residual magnitudes as a function of geoid slope. This figure not only convincingly demonstrates the dependence of the residuals on the geoid gradient, but also gives a quantitative form for the geoid-slope dependent term in the measurement noise model (function $\phi()$ of (22)) used for the other results in this section. Such a heterogeneous set of measurement noise variances may be used with no appreciable increase in computational burden (just as before, with the heterogeneous process noise model for the Kuroshio).

V. CONCLUSIONS

We have demonstrated the application of a highly efficient multiscale estimation framework to the problem of ocean altimetry estimation based on irregularly sampled satellite measurements. A number of significant difficulties which have led to significant suboptimalities and approximations in many other estimation algorithms are resolved by our approach: the multiscale framework presented in this paper possesses the efficiency to deal with truly enormous, possibly nonstationary, problems, computing both estimates and error variances with relative computational ease. Furthermore the concept of scale is made explicit, permitting the explicit characterization of phenomena possessing interactions across a number of scales.

Although throughout this paper the ocean altimetry application has been used as a vehicle for demonstrating the use of the multiscale framework in such a modeling context, the success of the application motivates many further possible applications as well as extensions within the current context. With respect to the latter, we can point to several problems of considerable scientific interest, including the following:

13

- The distribution of measurement residuals (Figure 13) demonstrates clearly the presence of geoid error as well as suggesting a way in which to correct for it and thus provide local corrections to our estimate of the geoid. In particular, it is possible that *joint* estimation of the geoid and ocean height may simultaneously improve estimates of both of these quantities.

- The precise shape of power spectrum of the ocean remains a matter of current scientific interest. Multiscale likelihood methods[13] provide an efficient and statistically rigorous machinery for examining problems of identifying the statistical structure of random fields, and their use should be of value in assessing the nature of the ocean spectrum.

In addition there are a number of extensions of our multiscale modeling framework under investigation including the development of higher-order methods for estimating both surface height and surface gradients (a problem of independent interest in surface reconstruction problems in computer vision), the development of multiresolution methods simultaneously in space and time, and the development of models involving overlapping regions at each scale, as mentioned in the Section IV.

APPENDIX - ESTIMATOR DETAILS

This appendix describes the algorithm that implements our multiscale estimation scheme. The description below is complete but terse; interested readers are referred to [3, 4, 13] for a more thorough development.

The multiscale smoother is basically the same as the Rauch-Tung-Striebel smoother operating in one dimension (along scale), with the addition of a merge operation, that combines the information of multiple child nodes into one parent node (upwards pass), and a split operation, which distributes information from a parent node to its multiple child nodes (downwards pass).

A certain amount of notation is required in order to describe the relative positions of state nodes on a tree; Figure 15 shows the various relations:

s is an abstract index for identifying nodes on the tree
$\bar{\gamma}$ is the raising operator; i.e., $s\bar{\gamma}$ is the parent of s
$\bar{\delta}$ is the sibling operator; i.e., $s\bar{\delta}$ is the sibling node next to s
α is the lowering operator; i.e., $s\alpha_n$ is the nth child of s
q is the order of the tree; i.e., the number of descendants of each parent

Note that operators can be cascaded, e.g., $s\bar{\gamma}\bar{\delta}\alpha_2$. The terms "upwards" and "downwards" are used with respect to the tree of Figure 15; that is "upwards" implies a movement towards coarser scales, and "downwards" towards finer scales.

The tree process and observation relations are described as follows:

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s) \tag{24}$$
$$y(s) = C(s)x(s) + v(s) \tag{25}$$

where the process noise satisfies

$$w(s) \sim \mathcal{N}(0, I) \qquad E\left[w(s)w(t)^T\right] = I\delta_{s,t} \tag{26}$$

and with a prior covariance at the root node

$$x_o = x(0) \sim \mathcal{N}(0, P_o) \tag{27}$$

From [4], corresponding to any choice of downwards model in Equation 24, we have the following upward model:

$$
\begin{align}
x(s\bar{\gamma}) &= F(s)x(s) + \bar{w}(s) \tag{28} \\
y(s) &= C(s)x(s) + v(s) \tag{29} \\
F(s) &= P_{s\bar{\gamma}}A^T(s)P_s^{-1} \tag{30} \\
E\left[\bar{w}(s)\bar{w}^T(s)\right] &= P_{s\bar{\gamma}}\left(I - A^T(s)P_s^{-1}A(s)P_{s\bar{\gamma}}\right) \tag{31} \\
&= Q(s) \tag{32}
\end{align}
$$

P_s is the *prior* variance of the state $x(s)$. To make the estimator equations more compact, additional notation is required at this point:

$$
\begin{align}
Y_s &= \{y(\sigma) \mid \sigma \text{ is a descendant of } s\} \tag{33} \\
\hat{x}(\sigma \mid s) &= E\left[x(\sigma) \mid \sigma \in Y_s \cup y(s)\right] \tag{34} \\
\hat{x}(\sigma \mid s+) &= E\left[x(\sigma) \mid \sigma \in Y_s\right] \tag{35} \\
P(\sigma \mid s) &= \text{Cov}\left[x(\sigma) - \hat{x}(\sigma \mid s)\right] \tag{36} \\
P(\sigma \mid s+) &= \text{Cov}\left[x(\sigma) - \hat{x}(\sigma \mid s+)\right] \tag{37}
\end{align}
$$

The algorithm now proceeds in three steps, outlined below.

1. Initialization
At each leaf node s, assign the following prior values:

$$
\begin{align}
\hat{x}(s \mid s+) &= 0 \tag{38} \\
P(s \mid s+) &= P_s \tag{39}
\end{align}
$$

2. Upward Sweep
The upward sweep operates very much like a Kalman filter operating along scale, with the addition of a merge step. The Kalman filter update step is performed at all nodes:

$$
\begin{align}
\hat{x}(s \mid s) &= \hat{x}(s \mid s+) + K(s)\left[y(s) - C(s)\hat{x}(s \mid s+)\right] \tag{40} \\
P(s \mid s) &= [I - K(s)C(s)]P(s \mid s+) \tag{41} \\
K(s) &= P(s \mid s+)C^T(s)V^{-1}(s) \tag{42} \\
V(s) &= C(s)P(s \mid s+)C^T(s) + R(s) \tag{43}
\end{align}
$$

The Kalman filter prediction step is applied at all nodes except for leaf nodes (which were initialized as outlined above):

$$\hat{x}(s \mid s\alpha_i) = F(s\alpha_i)\hat{x}(s\alpha_i \mid s\alpha_i) \tag{44}$$

$$P(s \mid s\alpha_i) = F(s\alpha_i)P(s\alpha_i \mid s\alpha_i)F^T(s\alpha_i) + Q(s\alpha_i) \tag{45}$$

Finally, at all nodes except leaf nodes, the merge step combines predicted estimates from offspring $(1 \ldots q)$ into a single prediction to be used in the update step:

$$\hat{x}(s \mid s+) = P(s \mid s+)\sum_{i=1}^{q}P^{-1}(s \mid s\alpha_i)\hat{x}(s \mid s\alpha_i) \tag{46}$$

$$P(s \mid s+) = \left[(1-q)P_s^{-1} + \sum_{i=1}^{q}P^{-1}(s \mid s\alpha_i)\right]^{-1} \tag{47}$$

3. Downward Sweep

The termination of the upward sweep gives the smoothed estimate $\hat{x}^s(0) = \hat{x}(0 \mid 0)$ at the root node. The remainder of the smoothed estimates are found by propagating information back down the tree:

$$\hat{x}^s(s) = \hat{x}(s \mid s) + J(s)\left[\hat{x}^s(s\bar{\gamma}) - \hat{x}(s\bar{\gamma} \mid s)\right] \tag{48}$$

$$P^s(s) = P(s \mid s) + J(s)\left[P^s(s\bar{\gamma}) - P(s\bar{\gamma} \mid s)\right]J^T(s) \tag{49}$$

$$J(s) = P(s \mid s)F^T(s)P^{-1}(s\bar{\gamma} \mid s) \tag{50}$$

The smoothed measurements are given by $\hat{x}^s(s)$; the corresponding estimation error variances are given by $P^s(s)$. Cross covariances are not computed explicitly, rather the means for their computation is implicit (although by no means obvious) in the above algorithm. The multiscale form of the smoothing error[14] is as follows:

$$\tilde{x}^s(s) = P(s \mid s)F^T(s)P^{-1}(s\bar{\gamma} \mid s)\tilde{x}^s(s\bar{\gamma}) + \breve{w}(s) \tag{51}$$

$$\tilde{x}^s(s) = x(s) - x^s(s) \tag{52}$$

where $\breve{w}(s)$ represents white noise.

References

[1] B. Anderson, J. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979

[2] D. Chelton, M. Schlax, D. Witter, J. Richman, "Geosat Altimeter Observations of the Surface Circulation of the Southern Ocean", *J. Geophys. Research* (95), pp.17877–17903, 1990

[3] K. Chou, "A Stochastic Modeling Approach to Multiscale Signal Processing", PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, 1991

[4] K. Chou, A. Willsky, A. Benveniste, "Multiscale Recursive Estimation, Data Fusion, and Regularization", *IEEE Trans. on Automatic Control*, to appear.

[5] R. Daley, *Atmospheric Data Analysis*, Cambridge University Press, New York, 1991

[6] P. Fieguth, *Application of Multiscale Estimation to Large Scale Multidimensional Problems*, PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, *in preparation*

[7] L. Fu, E. Christensen, M. Lefebvre, Y. Menard, "TOPEX/POSEIDON mission overview," *J. Geophys. Res.*, 1994, in press

[8] P. Gaspar, C. Wunsch, "Estimates from Altimeter Data of Baryotropic Rossby Waves in the Northwestern Atlantic Ocean", *J. Physical Oceanography* (19)#12, pp.1821–1844, 1989

[9] M. Ghil, P. Malanotti-Rizzoli, "Data Assimilation in Meteorology and Oceanography", *Advances in Geophysics* (33), pp.141–266, 1991

[10] C. King, D. Stammer, C. Wunsch, "The CMPO/MIT TOPEX/POSEIDON Altimetric Data Set", *MIT CGCS Report #30*, 1994

[11] B. Levy, "Non-causal estimation for Markov random fields", In *Proc. of the Int'l Symposium MTNS-89, Vol 1* (M. Kaachoek, J. Schuppen, A. Ran ed.s), Birkhauser-Verlag, 1990

[12] M. Luettgen, W. Karl, A. Willsky, R. Tenney, "Multiscale Representations of Markov Random Fields", *IEEE Trans. Signal Processing* (41) #12, pp.3377–3396, 1993

[13] M. Luettgen, "Image Processing with Multiscale Stochastic Models", PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, 1993

[14] M. Luettgen, A. Willsky, "Smoothing Error Models", LIDS Technical Report LIDSP 2334, MIT, 1994

[15] P. Malanotte-Rizzoli, "Data Assimilation: Fundamentals, Global and Mediterranean Examples", *Ocean Processes in Climate Dynamics: Global and Mediterranean Examples* (P. Malanotte-Rizzoli, A. Robinson ed.s), NATO Asi Series (419), 1994

[16] P. Marth et al., "Prelaunch Performance of the NASA Altimeter for the TOPEX/POSEIDON Project", *IEEE Trans. on Geoscience and Remote Sensing* (31) #2, pp.315–332, 1993

[17] F. Monaldo, "Topex Ionospheric Height Correction Precision Estimated from Prelaunch Test Results", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.371–375, 1993

[18] R. Nerem et al., "Expected Orbit Determination Performance for the Topex/Poseidon Mission", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.333–354, 1993

[19] R. Nerem et al., "Gravity model development for Topex/Poseidon: Joint gravity model 1 and 2", *J. Geophys. Res*, 1994

[20] J. Nystuen, C. Andrade, "Tracking Mesoscale Ocean Features in the Caribbean Sea Using Geosat Altimetry", *J. Geophys. Research* (98)#C5, pp.8389–8394, 1993

[21] R. Rapp, Y. Wang, N. Pavlis, "The Ohio State 1991 Geopotential and Sea Surface Topography Harmonic Coefficient Models", Report #410, Dept. of Geodetic Science and Surveying, Ohio State University, 1991

[22] R. Rapp, "Geoid Undulation Accuracy", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.365–370, 1993

[23] H. Rauch, F. Tung, C. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems", *AIAA Journal*, (3) #8, 1965

[24] R. Ray, "Global Ocean Tide Models on the Eve of Topex/Poseidon", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.355–364

[25] B. Ripley, *Spatial Statistics*, Wiley, 1991

[26] T. Schlatter, G. Branstator, I. Ihiel, "Testing a Global Multivariate Statistical Objective Analysis Scheme with Observed Data", *Mon. Weather Review* (104), pp.765–783, 1976

[27] D. Stammer, C. Wunsch, "Preliminary assessment of the accuracy and precision of TOPEX/POSEIDON altimeter data with respect to the large scale ocean circulation", *J. Geophys. Res.*, 1994, in press

[28] *Kuroshio: Physical Aspects of the Japan Current* (H. Stommel, K. Yoshida ed.s), University of Washington Press, 1972

[29] R. Tokmakian, P. Challenor, "Observations in the Canary Basis in and Azores Frontal Region Using Geosat Data", *J. Geophys. Research* (98)#C3, pp.4761–4773, 1993

[30] G. Wornell, "Synthesis, Analysis, and Processing of Fractal Signals", PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, 1991

[31] G. Wornell, "Wavelet-Based Representation for the $1/f$ Family of Fractal Processes", *Proc. IEEE*, Sept. 1993

[32] C. Wunsch, "Sampling Characteristics of Satellite Orbits", *J. of Atmospheric and Oceanic Tech.* (6)#6, pp.891–907, 1989

List of Figures