Recursive Multiscale Estimation of Space-Time Random Fields

Terrence T. Ho[†], *Paul W. Fieguth*[‡], *Alan S. Willsky*[†]

† Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

‡ Dept. of Systems Design Engineering
University of Waterloo

ABSTRACT

We recently developed a multiscale recursive estimation procedure for the estimation of large-scale dynamic systems. The procedure propagates multiscale models for the estimation errors more efficiently than the Kalman filter's propagation of the error covariances, with a resulting computational complexity of $\mathcal{O}(N)$ and $\mathcal{O}\left(N^{3/2}\right)$, where N is the number of variables estimated, for 1-D and 2-D dynamic systems, respectively. To further reduce the computational cost, we introduce in this paper a new class of reduced-order spatially-interpolated multiscale models, and demonstrate their use in remote sensing.

1. Introduction

The statistical estimation of two-dimensional dynamic systems is of great interest in many applications, such as ocean-height[3] or ocean-temperature (Fig. 1) mapping, where methods for assimilating data with dynamics have hitherto been limited. Even systems governed by simple dynamics such as diffusion present a challenge, due to the large state-dimensions (compared to one-dimensional problems) and to the time-varying nonstationary statistics (compared to most static problems).

In our past work [4, 5] we adapted a multiscale stochastic modeling and estimation methodology to the estimation of 1-D and small 2-D dynamic systems. Instead of propagating error covariances over time, like the Kalman filter, the multiscale recursive estimation algorithm propagates *models* for the estimation error, not covariances, and does so efficiently.

This paper addresses the challenges encountered in larger problems than studied earlier. Specifically, we consider a class of reduced-order models and interpolation methods for estimation of 2-D processes that further reduce the computational cost, and present preliminary dynamic ocean-temperature estimation results.

Research of this paper supported in part by the Natural Science & Engineering Research Council of Canada, and by the Office of Naval Research under Grant N0014-91-J-1004.

Fig. 1. Mean ocean temperature and superimposed measurements.

2. Dynamic Estimation

Consider a dynamic system,

$$\boldsymbol{z}(t+1) = \boldsymbol{A}_d\,\boldsymbol{z}(t) + \boldsymbol{w}_d(t). \tag{1}$$

The form of least-squares time-recursive estimators (which includes the optimal Kalman filter) consists of two stages. One, the update stage,

$$\hat{\boldsymbol{z}}(t|t) = \hat{\boldsymbol{z}}(t|t-1) + \hat{\boldsymbol{\chi}}(t|t-1), \tag{2}$$

takes the measurements into account, where $\hat{\boldsymbol{\chi}}(t|t-1)$ is the estimate of the prediction error $\boldsymbol{\chi}(t|t-1)$. The measurement update step is essentially a static estimation problem, for which efficient methods are well-studied. Much more troubling is the prediction stage:

$$\hat{\boldsymbol{z}}(t+1|t) = \boldsymbol{A}_d\,\hat{\boldsymbol{z}}(t|t), \tag{3}$$

which accounts for time. The spatial mixing, introduced by all but the most trivial dynamics, destroys the particular statistical structure which one might wish to assume (via sparse matrices, preconditioning, or multiscale methods) to gain efficiency in the update stage.

This paper investigates ways of propagating multiscale models through the prediction step for fast estimation for large problems.

0-7803-5467-2/99/ $ 10.00 © 1999 IEEE

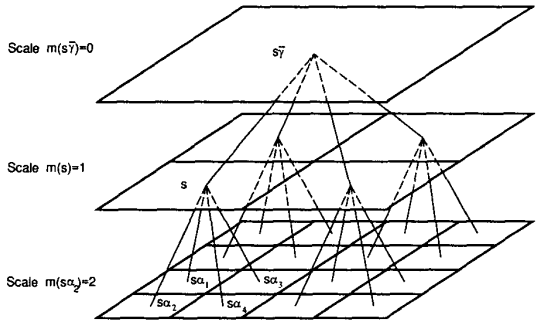


Fig. 2. An example multiscale quad-tree for modeling two-dimensional fields.

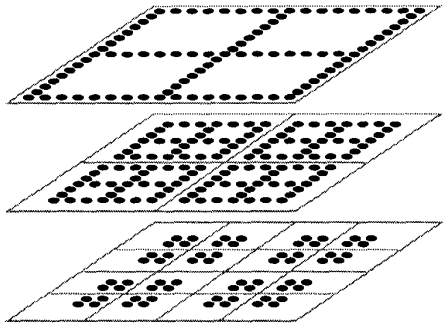


Fig. 3. The non-redundant linear functionals for modeling a 17×17 2-D MRF on a three-level tree.

3. Multiscale Estimation

The multiscale recursive estimator in [5] achieves its computational efficiency by propagating a *model* for the estimation errors χ through time without explicitly calculating or storing the second-order statistics $\boldsymbol{P}_\chi(t|t)$. The error fields are modeled on tree structures; one particularly common and useful tree is the regular quadtree, shown in Fig. 2, on which each node has $q = 4$ children:

$$\boldsymbol{x}(s;\cdot|\cdot) = \boldsymbol{A}(s;\cdot|\cdot)\boldsymbol{x}(s\bar{\gamma};\cdot|\cdot) + \boldsymbol{B}(s;\cdot|\cdot)\boldsymbol{w}(s;\cdot|\cdot). \quad (4)$$

$\boldsymbol{x}(s;\cdot|\cdot)$ is the state at s, which indexes the nodes of the tree, and $\boldsymbol{w}(s;\cdot|\cdot)$ is a white noise process uncorrelated with $\boldsymbol{x}(0;\cdot|\cdot)$. From (4) it follows that the $q+1$ subtrees connected to node s are conditionally decorrelated by the state $\boldsymbol{x}(s;\cdot|\cdot)$, which makes possible an efficient scale-recursive estimation algorithm on the tree.

We are interested in defining the multiscale state at s as a linear function of the process $\boldsymbol{\chi}$ of interest:

$$\boldsymbol{x}(s;\cdot|\cdot) = \boldsymbol{L}(s;\cdot|\cdot)\boldsymbol{\chi}(\cdot|\cdot). \quad (5)$$

The choice of $\boldsymbol{L}(s;\cdot|\cdot)$ is not arbitrary. The linear functionals must satisfy the conditional decorrelation property and reproduce the desired statistical behavior of the modeled process. For example, the class of first-order Markov random fields can be modeled exactly by letting $\boldsymbol{x}(s;\cdot|\cdot)$ contain subregion boundary pixels, as shown in Fig. 3.

Given (5), the *a priori* model parameters $\boldsymbol{A}(s;t|t-1)$ and $\boldsymbol{B}(s;t|t-1)$ are determined from the joint statistics between $\boldsymbol{x}(s;t|t-1)$ and $\boldsymbol{x}(s\bar{\gamma};t|t-1)$, which themselves follow from $\boldsymbol{P}_\chi(t|t-1)$. The multiscale estimation algorithm[2] solves the update step, and allows the *updated* error model parameters $\boldsymbol{A}(s;t|t)$ and $\boldsymbol{B}(s;t|t)$ to be determined, however the real challenge arises in obtaining a model for the *predicted* error.

4. Multiscale Prediction

Except for the most trivial cases, the mixing introduced by the dynamics from time t to $t+1$ implies that the statistics of $\boldsymbol{x}(s;t|t)$ and $\boldsymbol{x}(s;t+1|t)$ are related in a non-obvious manner. Specifically, in order to find the predicted error model $\boldsymbol{A}(s;t+1|t)$ and $\boldsymbol{B}(s;t+1|t))$ we need the joint statistics between

$$\boldsymbol{x}(s;t+1|t) = \boldsymbol{L}(s;t+1|t)\boldsymbol{A}_d\boldsymbol{\chi}(t|t) \quad (6)$$

$$\boldsymbol{x}(s\bar{\gamma};t+1|t) = \boldsymbol{L}(s\bar{\gamma};t+1|t)\boldsymbol{A}_d\boldsymbol{\chi}(t|t) \quad (7)$$

which is much more involved than the joint statistics between just $\boldsymbol{x}(s;t|t)$ and $\boldsymbol{x}(s\bar{\gamma};t|t)$ because of the temporal dynamics $\boldsymbol{A}_d$. The details of the specific additional joint statistics which are required depend on the choice of $\boldsymbol{L}(s)$ and on the details of the dynamics. In principle the joint statistics can be computed for *any* two nodes $\boldsymbol{x}(s_1;t|t)$, $\boldsymbol{x}(s_2;t|t)$; obviously the number of such computed statistics must be limited, since finding *all* of these is equivalent to the brute-force computation of the whole posterior covariance — precisely that which we want to avoid.

Even for dynamics that operate locally in space (e.g., diffusion), standard state assignments suggested by Fig. 2 can be difficult, since many physically proximate finest-scale states may distantly separated on the tree. However for nearest-neighbor dynamics, a recently developed non-redundant state assignment, shown in Fig. 3, can be highly effective, since all of the pixels adjacent to node s lie either on s, its parent, or one of its children, which *greatly* limits the additional joint statistics which are required. For models whose state dimensions grow linearly with $\sqrt{N}$ (as is the case in Fig. 3), the computational complexity becomes $\mathcal{O}(N^{3/2})$.

As an aside, as the predicted error statistics change over time, ideally so should the linear functionals $\boldsymbol{L}(s;\cdot|\cdot)$; however we impose a fixed set of $\boldsymbol{L}(s)$ for the current application, which does imply that the prediction errors are only approximately realized.

5. Reduced-Order 2-D Models

Given the degree of correlation between neighboring elements of typical 2-D random fields, a reduced-order state that models only a subsampled set of the boundary points may adequately capture the correlation structure of the field and lead to faster estimation. However for non-redundant models (as in Fig. 3), where each field pixel appears in only one tree state, a subsampled state implies that some elements of the field of interest do not appear anywhere on the tree. That is, no estimates or error statistics for these elements are directly available, and measurements of these elements cannot be placed on the tree. Although it is possible to ignore these missing elements for the purposes of static estimation, in performing the prediction step, estimates and error statistics will be required.

Let us denote the boundary points not represented in $\boldsymbol{x}(s)$ by $\boldsymbol{\zeta}(s)$. The simplest solution is to linearly predict the missing points from each state:

$$\boldsymbol{\zeta}(s) = \boldsymbol{M}(s)\boldsymbol{x}(s) + \boldsymbol{w}_{\zeta}(s\alpha_{\zeta}), \tag{8}$$

where

$$\boldsymbol{M}(s) = \boldsymbol{P}_{\zeta,x}(s)\boldsymbol{P}_{x}^{-1}(s), \tag{9}$$

$$\mathrm{cov}(\boldsymbol{w}_{\zeta}(s\alpha_{\zeta})) = \boldsymbol{P}_{\zeta}(s) - \boldsymbol{P}_{\zeta,x}(s)\boldsymbol{P}_{x}^{-1}(s)\boldsymbol{P}_{\zeta,x}^{T}(s) \tag{10}$$

Although fast, this approach is rather limited in that it ignores the 2-D structure of the field and, more significantly, it neglects to use nearby pixels from other state elements in the interpolation process.

A more sophisticated method is to interpolate $\boldsymbol{\zeta}(s)$ based on *all* nearby state elements; that is, a two-dimensional interpolation based on elements from the parent and all children:

$$\boldsymbol{\zeta}(s) = \boldsymbol{M}(s)\begin{bmatrix} \boldsymbol{x}(s\bar{\gamma}) \\ \boldsymbol{x}(s) \\ \boldsymbol{x}(s\alpha_1) \\ \vdots \\ \boldsymbol{x}(s\alpha_q) \end{bmatrix} + \boldsymbol{w}_{\zeta}(s\alpha_{\zeta}). \tag{11}$$

with the estimation equations following as in (9),(10). The advantage of this 2-D interpolation is that it leads to better estimates and realized error variances; on the other hand, one pays a computational penalty, especially during the prediction step, since a greater number of joint statistics need to be computed in order to compute the model parameters.

We have found 2-D interpolation for reduced-order models to be an effective tradeoff between computationally intensive full-state models and statistically approximate one-dimensional interpolation.

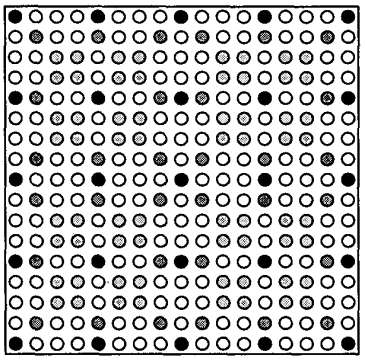
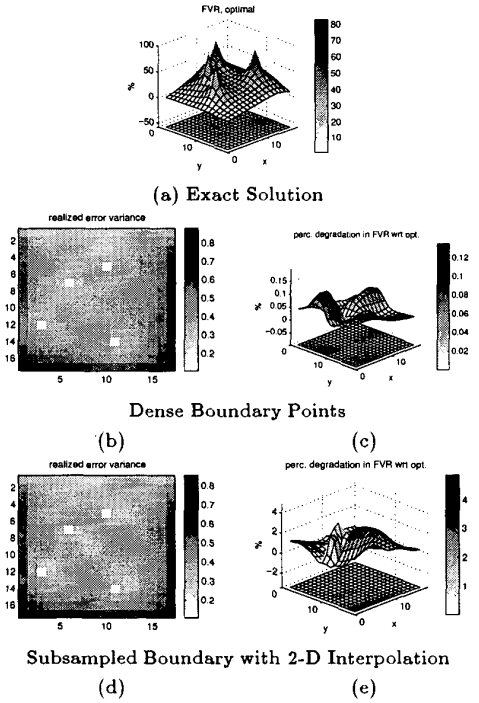


Fig. 4. Reduced-order non-redundant linear functionals for modeling a 17 × 17 2-D field. The hollow circles indicate elements of the field not modeled. Compare with dense-boundary functionals in Fig. 3



(a) Exact Solution

Dense Boundary Points

(b) (c)

Subsampled Boundary with 2-D Interpolation

(d) (e)

Fig. 5. Performance comparison of the optimal FVR of the exact solution (a) with suboptimal multiscale steady-state estimators; (b), (d) Realized updated error variances. (c), (e) Degradation in multiscale FVR, within about 2% of the optimum.

Fig. 6. Dynamic estimation of Pacific surface-temperature, three days apart.

6. Results

We will model our dynamics as a diffusion process,

$$\frac{\partial z}{\partial t} = \nabla^2 z - \beta \cdot z + \gamma \cdot \omega, \tag{12}$$

where z is the temperature distribution; ω is white Gaussian noise, and β and γ are constants. Fig. 5 shows a 17×17 pinned cooling sheet example with four point measurements. We compare the realized statistics and fractional variance reduction

$$\text{FVR} = \frac{\text{Var(process)} - \text{Var(updated error)}}{\text{Var(process)}} \tag{13}$$

of the exact estimator with two multiscale ones — the first using dense boundary points and the other subsampled boundaries with 2-D interpolation. The FVR of the full-order estimator is affected only minutely — $\approx 0.1\%$ — with respect to the exact solution. The approximate, reduced-order approach has errors on the order of only 1% to 2%.

Fig. 6 shows a small part of a more ambitious exercise — the dynamic estimation of the ocean surface temperature over six months. We model the ocean surface as diffusive (over sufficiently short time steps, here of one day); the plotted results are taken from the five-month point of the estimated sequence. We see that the non-redundant linear functionals work well for modeling estimation errors under a variety of general conditions, e.g., when and the number and locations of measurements are time-varying, implying that the problem is time-varying and does not attain steady-state.

Future extensions to the above results include running tests for substantially larger domains and applying the method to local dynamics other than diffusion.

7. References

[1] T. M. Chin, W. C. Karl, A. S. Willsky, "A Distributed and Iterative Method for Square Root Filtering in Space-Time Estimation", *Automatica* (31) #1, pp.67–82, 1995

[2] K. Chou, A. Willsky, A. Benveniste, "Multiscale Recursive Estimation, Data Fusion, and Regularization", *IEEE Trans. on Automatic Control* (39) #3, pp.464–478, 1994

[3] P. Fieguth, D. Menemenlis, T. Ho, A. Willsky, and C. Wunsch. "Mapping Mediterranean Altimeter Data with a Multiresolution Optimal Interpolation Algorithm". *Journal of Atmospheric and Oceanic Technology*, 15:535–546, April 1998.

[4] T. Ho, P. Fieguth, and A. Willsky. "Computationally Efficient Steady-State Multiscale Estimation ...". Submitted to *Automatica*.

[5] T. Ho, P. Fieguth, and A. Willsky. "Computationally Efficient Multiscale Estimation of Large-Scale Dynamic Systems". *ICIP'98*, Chicago, 1998.

[6] T. T. Ho. *Multiscale Modeling and Estimation of Large-Scale Dynamic Systems*. PhD Thesis, M.I.T., Cambridge, MA, September 1998.

[7] C. T. Mutlow and A. M. Závody and I. J. Barton and D. T. Llewellyn-Jones, "Sea surface temperature measurements ...", *J. of Geophys. Res.* (99), pp.22575–22588, 1994