STATISTICAL PROCESSING OF LARGE IMAGE SEQUENCES

(Accepted for publication in IEEE Trans. Image Processing)

F. Khellah[0]
Dept. Computer Science, Prince Sultan University, Riyadh, Saudi Arabia
P. Fieguth
Dept. Systems Design Engineering, University of Waterloo, Waterloo, Canada
M.J. Murray
Rutherford Appleton Laboratory, Chilton, Didcot, United Kingdom
M. Allen
Atmospheric, Oceanic and Planetary Physics, Oxford University, United Kingdom

***Abstract*—The dynamic estimation of large-scale stochastic image sequences, as frequently encountered in remote sensing, is important in a variety of scientific applications. However, the size of such images makes conventional dynamic estimation methods, for example the Kalman and related filters, impractical. In this paper we present an approach that emulates the Kalman filter, but with considerably reduced computational and storage requirements. Our approach is illustrated in the context of a 512×512 image sequence of ocean surface temperature.**

The static estimation step, the primary contribution here, uses a mixture of stationary models to accurately mimic the effect of a nonstationary prior, simplifying both computational complexity and modelling. Our approach provides an efficient, stable, positive-definite model which is consistent with the given correlation structure. Thus the methods of this paper may find application in modelling and single-frame estimation.

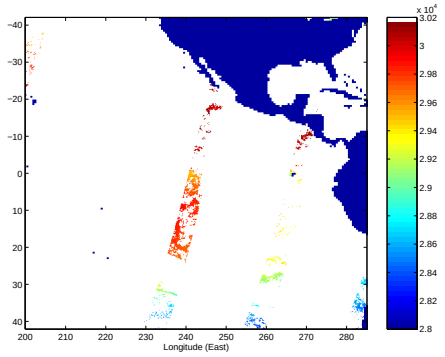


Fig. 1. ATSR SST observations for the month of October 1992: the measurements are accurate, but in sparse strips.

I. INTRODUCTION

There is a tremendous interest in the processing of image sequences. Although a great deal of this interest stems from excitement surrounding up-and-coming areas such as *multimedia* and *Internet video*, much of the fundamental research is being driven by the fact that storage, bandwidth, and computational power have, for the first time, really increased to the point where it is practical to process long sequences of large images.

Since video represents the overwhelming bulk of image-sequence data, much of the recent literature [5], [7], [16], [34], [35], [40], [47] is focused here; however image sequences from a camera or a broadcast source generally do not obey well-defined spatial or temporal statistics, so the processing of such signals proceeds heuristically or nonparametrically, and certainly locally, since image pixels are normally not meaningfully correlated at long ranges.

Instead, the interests of this paper are considerably more specialized and focus on the problem of dynamic estimation; that is, the spatio-temporal statistical processing of random fields, typically remotely-sensed over time. In particular, our research is motivated by a need to estimate ocean sea-surface temperature (SST) from sparse measurements, such as those shown in Figure 1, taken from the Along Track Scanning Radiometer (ATSR) [43], [44]. Ideally, we would like to produce dense images of SST, smoothed over time, but preserving the local features which are present in the measurements. Although motivated by this context, the methods developed in this paper are not SST or ocean-data specific, and the approach can be generalized and applied to any image-sequence problem whose spatial statistics can be parametrized (as will be described in Section II).

Generally, solutions to the image sequence processing problem fall into the generic prediction-update structure as shown in Figure 2, in which a sequence of observed images $\mathbf{y}(t)$ is processed, predicted estimates $\hat{\mathbf{x}}(t|t-1)$ are inferred from an estimated motion field $\mathbf{m}_t$, and the updated estimates $\hat{\mathbf{x}}(t|t)$ are driven by a residual field $\nu(t)$ which encodes the information present in $\mathbf{y}(t)$ which is not contained in $\hat{\mathbf{x}}(t|t-1)$.

Conceptually the statistical filtering problem is straightforward, and was solved decades ago in the form of the well-known Kalman Filter [29], whose structure closely parallels Figure 2:

Prediction Step:

$$\underline{\hat{x}}(t+1|t) = A\underline{\hat{x}}(t|t) \quad (1)$$

$$\tilde{P}(t+1|t) = A\tilde{P}(t|t)A^T + Q \quad (2)$$

where A captures the system[1] dynamics (diffusive, advective etc.) and Q the process or driving noise.

Update Step:

$$K(t) = \tilde{P}(t|t-1)C(t)^T\left(C(t)\tilde{P}(t|t-1)C(t)^T + R(t)\right)^{-1} \quad (3)$$

[0]The support of the Natural Science & Engineering Research Council of Canada is acknowledged. Contact author is P. Fieguth. EDICS Category 2-SEQP

[1]For simplicity of notation and presentation we assume the model parameters A, Q to be fixed over time; the results of this paper apply equally well to the time-varying case.

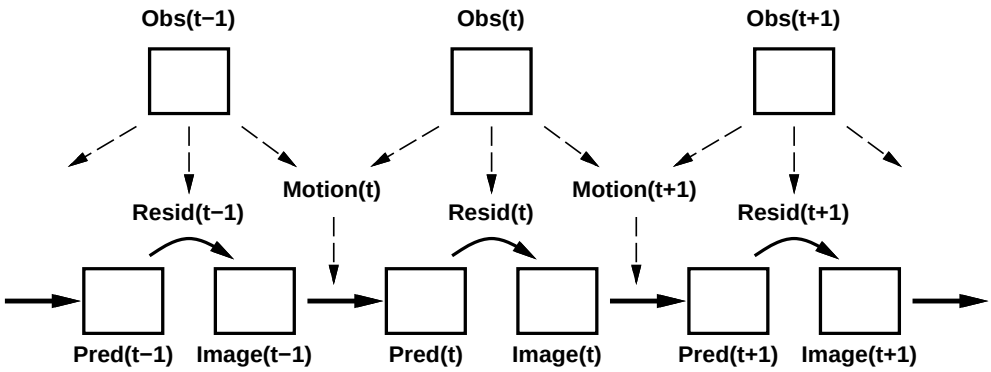


Fig. 2. Standard prediction-update process: An image sequence can be used to infer motion information and a residual, which determine the time-to-time prediction and update. A wide variety of problems can be cast into this framework.

$$\hat{\underline{x}}(t|t) = \hat{\underline{x}}(t|t-1) + K(t)[\underline{y}(t) - C(t)\hat{\underline{x}}(t|t-1)] \quad (4)$$

$$\tilde{P}(t|t) = [I - K(t)C(t)]\tilde{P}(t|t-1) \quad (5)$$

where measurements $\underline{y}(t) = C(t)\underline{x}(t) + \underline{v}(t)$ having a measurement error covariance of $R(t)$ are incorporated in order to improve the estimates. The need to generate and invert enormous matrices (e.g., in the computation of the gain $K(t)$) makes the brute-force application of the above equations to even modestly-sized images computationally infeasible for the indefinite future. Superficially, it would appear that there are two predominant challenges:

1) The storage of large matrices $A, C(t), K(t), \tilde{P}(t|t)$ etc;
2) The computation of the update step, involving a large matrix inversion.

In fact, the problem is much more difficult than this: because the location of measurements will change from time to time, the statistics of the problem will become nonstationary, therefore stationarity assumptions cannot be made. Next, whatever implicit or sparse representation is chosen for the 2D covariance $\tilde{P}(t|t)$ must guarantee positive definiteness, which is challenging. Finally, whatever representation is chosen for $\tilde{P}$ must be simultaneously compatible with an efficient approach to prediction and estimation.

That is, we require a representation of the problem which can accommodate spatial nonstationarities, is guaranteed to be positive definite, can be predicted over time, and is compatible with an efficient approach to solving the update step.

A variety of alternatives have, of course, been proposed. The reduced update Kalman filter (RUKF) was proposed in 1977 [49] as an efficient approach to the estimation of two-dimensional images. A more recently proposed three-dimensional version [34], [35] would, in principle, be able to perform dynamic estimation. The proposed RUKF gains its efficiency by requiring the estimator to be local. In addition, the underlying random field is assumed to be stationary and hence, the filter will reach steady-state after a few iterations. Although these assumptions are reasonable for processing images, locality leads to undesirable artifacts in sparsely-measured domains, and the long correlation length of remotely-sensed fields and the non-stationarity of their statistics pose difficulties.

Another alternative would be to assume a fixed prior over time $\tilde{P}_P = \tilde{P}(t+1|t)$. If the gain is also fixed, $K = K(t)$, then the estimator reduces to the steady-state Kalman filter [29], which is computationally very fast as one need not propagate or recompute the huge error covariances at each filter iteration. On the other hand, the steady-state filter is incompatible with the SST problem (and virtually any other scientific image-sequence problem) due to the time-varying measurement locations. Even if the gain is allowed to vary, a fixed prior precludes the computation of meaningful estimation error statistics, and causes the estimator to forget or discard past measurements and estimates too rapidly.

The 2D domain can be modelled as a collection of piecewise independent regions, where the different regions obey different statistics. Such an approach easily produces positive-definite, nonstationary covariances, however the region boundaries appear as obvious artifacts in the estimates.

Rather than trying to store or approximate the matrix $\tilde{P}(t+1|t)$ itself, an alternative is to build a statistical *model* for it. One approach, which has been successful in representing nonstationary covariances for extremely large problems, is based on a hierarchical model [13], [38], [21], however the prediction of such models is very difficult [30], [31].

One promising and widely-used approach is to employ state reduction [6], [9], [42], [23], via principal components or subsampling, such that the Kalman filter is applied to $L\underline{x}$, where L is an $M \times N$ state reduction operator, where $M \ll N$. The drawback of this approach is that the state reduction forces a loss of information, so that detailed, local features cannot be preserved.

One final approach has been to represent a covariance inverse P^{-1} in terms of a sparse set of diagonal bands [2], [4], [9], [11], [12]; the number of diagonal bands determines the locality of the estimator. Although such approaches have been applied to remote sensing problems, we will see that for problems of practical interest, involving substantial correlation lengths, a very large number of bands, with corresponding computational and storage increases, are required to achieve accurate results.

The conclusion is that there currently exists no practical method for solving spatially irregular dynamic estimation problems on large (e.g., megapixel) domains. The *key* to our proposed approach is the following: rather than a *single* complicated, implicit, nonstationary representation, we will represent the nonstationary prior $\tilde{P}(t+1|t)$ as a spatially-weighted com-

bination or mixture of an *ensemble* of explicit, stationary priors, each of which has a simple representation, for which positive-definiteness is known, and which is compatible with efficient update methods!

The paper addresses the Prediction and Update steps in Sections II and III respectively, followed by experimental results in Section IV.

II. Prediction

In the prediction step, the system dynamics are used to predict two quantities:

1) Predicted values of the state estimates at the next time step, and
2) Predicted values of the error statistics at the next time step.

Of these, the former is straightforward, whereas the latter is much more difficult.

As will be seen in Section 3, a crucial requirement of our proposed update step is that the statistics be representable in a parameterized, analytic form. Although a variety of dynamic models satisfy this criterion, we will focus on the two-dimensional heat diffusion process:

$$\begin{aligned}\frac{\partial x(i,j,t)}{\partial t} &= a\cdot\frac{\partial^2 x(i,j,t)}{\partial i^2}+a\cdot\frac{\partial^2 x(i,j,t)}{\partial j^2}\\ &\quad -b\cdot x(i,j,t)+\gamma\cdot w(i,j,t) \qquad (6)\end{aligned}$$

where $x(i,j,t)$ represents a process of interest (e.g., temperature) at spatial position (i,j) and time t, and where $w()$ is unit-variance Gaussian white noise. The diffusion dynamics (6) have the following desirable attributes:

1) The process correlation function is exponential [14].
2) The dynamics admit a simple, local discretization.
3) Diffusion dynamics are found in a wide variety of applications.

The model (6) is, of course, an oversimplification of ocean dynamics, which are nonstationary and nonlinear. It is not our goal to propose an accurate ocean model, nor is this needed: our goal is a novel approach to the estimator, not to ocean dynamics; furthermore, in practice the model needs to be only good enough to regularize the space-time interpolation of measurements.

We discretize our model (6) onto an $m\times m$ grid to construct a system of difference equations

$$\underline{x}(t+1)=A\underline{x}(t)+\underline{w}(t) \qquad (7)$$

where $\underline{x}$ is a column vector of length $n=m^2$, representing the whole 2D process (stacked column by column). Under a Forward-Euler discretization scheme [18], [48] the dynamics matrix A is penta-diagonal and can be efficiently represented implicitly by a stationary convolutional kernel A_k:

$$A_k=\begin{bmatrix}0&\beta&0\\ \beta&\alpha&\beta\\ 0&\beta&0\end{bmatrix} \qquad (8)$$

Simplified by the sparse, implicit kernel representation of A in (8), the exact state prediction can be computed as a convolution, as is well known:

$$\underline{\hat{x}}(t+1|t)=A_k*\underline{\hat{x}}(t|t). \qquad (9)$$

The much greater challenge lies in the prediction of the error statistics. In principle, the estimation error is propagated through time as

$$\tilde{P}(t+1|t)=A\tilde{P}(t|t)A^T+Q. \qquad (10)$$

For large-scale dynamic estimation problems the exact, brute-force computation of (10) is impossible. Furthermore, from a storage perspective the updated estimation error $\tilde{P}(t|t)$ has a size of $\mathcal{O}(n^2)$ and is impossible to store explicitly. Finally, from a computational perspective the $\mathcal{O}(n^3)$ effort in the multiplications is infeasible for large 2-D dynamic estimation problems.

To be sure, the sparse structure of A leads to an obvious efficient approach to the matrix multiplication, however the storage of the resulting irregular and non-stationary $\tilde{P}(t+1|t)$ is still a problem. Our approach for the error prediction step is to parameterize the error covariances $\tilde{P}(t|t),\tilde{P}(t+1|t)$. Note that any positive semi-definite matrix P can be written as a product of standard deviations σ and correlation coefficients ρ:

$$\begin{aligned}P &= \{\underline{p}\,\underline{p}^T\}^{\frac{1}{2}}\odot\Phi \qquad (11)\\ &= \begin{bmatrix}\sigma^2_{(1,1)} & \sigma_{(1,1)}\sigma_{(2,1)} & \cdots & \sigma_{(1,1)}\sigma_{(n,n)}\\ \sigma_{(2,1)}\sigma_{(1,1)} & \sigma^2_{(2,1)} & & \vdots\\ \vdots & & \ddots & \\ \sigma_{(n,n)}\sigma_{(1,1)} & \cdots & & \sigma^2_{(n,n)}\end{bmatrix}\odot\\ &\quad\begin{bmatrix}1 & \rho_{(1,1)(2,1)} & \cdots & \rho_{(1,1)(n,n)}\\ \rho_{(2,1)(1,1)} & 1 & & \vdots\\ \vdots & & \ddots & \\ \rho_{(n,n)(1,1)} & \cdots & & 1\end{bmatrix}\end{aligned}$$

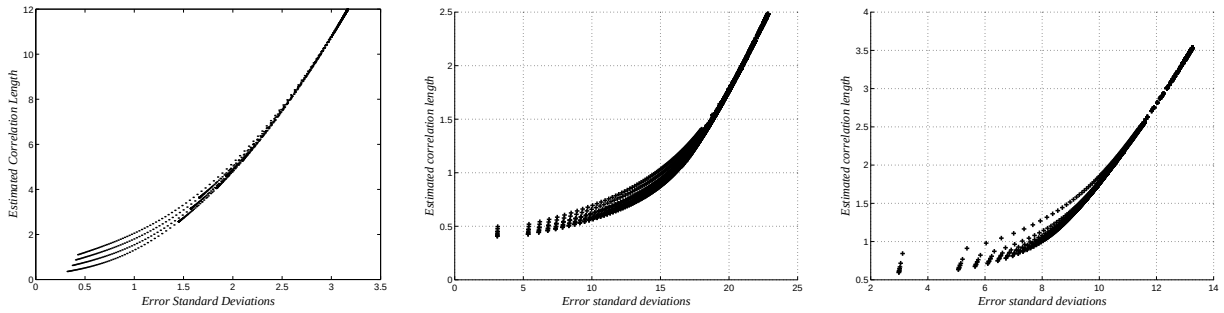
That is, P is explicitly expressed in terms of its variances $\underline{p}=\text{diag}(P)$ and its correlation coefficients Φ, where $\odot$ refers to element-by-element multiplication. So we wish to express the prediction (10) as

$$\begin{aligned}&\{\tilde{\underline{p}}(t|t)\,\tilde{\underline{p}}(t|t)^T\}^{\frac{1}{2}}\odot\Phi(t|t) \overset{\text{Prediction}}{\Longrightarrow}\\ &\qquad\{\tilde{\underline{p}}(t+1|t)\,\tilde{\underline{p}}(t+1|t)^T\}^{\frac{1}{2}}\odot\Phi(t+1|t). \qquad (12)\end{aligned}$$

This obviously raises new issues: how is the prediction in (12) to be affected, and how to choose a positive definite Φ?

The evolution of $\tilde{\underline{p}}$ is at least intuitive: the error variance decreases near measurements, and increases with every prediction step. The form of Φ is much less obvious. Naively, one might (we did, initially) assume that the correlation-coefficient structure might change relatively little with measurements, and that the most obvious positive definite structures, either the Lyapunov solution

$$\tilde{P}_p = A\tilde{P}_pA^T+Q \qquad (13)$$



(a) (b) (c)

Fig. 3. The empirical relationship between correlation length l and error standard deviations $\sqrt{\tilde{p}}$. For the static stationary-exponential case the relationship is parabolic (that is, l is linear in $\tilde{p}$), but nonstationarities (a) spread the relationship. Even for dynamic, simulated diffusion processes the relationship is still very nearly parabolic, illustrated for correlations lengths of (b) 4.0 and (c) 12.0.

or the Riccati solution[2]

$$\tilde{P}_p \quad = \quad A\tilde{P}_p A^T + Q - AK(t)C\tilde{P}_p A^T, \tag{14}$$

might suffice. Empirically, however, not only are the Lyapunov and Riccati solutions at the large and small extremes in terms of the estimation error variances, they *also* lie at the long and short extremes on the continuum of correlation lengths! Analytically, for a simple two-element problem

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad P = \begin{bmatrix} a & b \\ b & a \end{bmatrix}, \tag{15}$$

$$\text{Measurements } \underline{y} = I\underline{x} + \underline{v}, \quad \text{cov}(\underline{v}) = \begin{bmatrix} 1/\delta & 0 \\ 0 & 1/\delta \end{bmatrix} \tag{16}$$

it is easy to derive that $\tilde{a}$, the estimation error variance, cannot exceed a and is a decreasing function of δ, as expected, but *also* that $\tilde{b}$ drops *faster* than $\tilde{a}$. That is, that the correlation coefficient $\tilde{\rho} = \tilde{b}/\tilde{a}$ of the estimation error is *also* a decreasing function of δ.

Why should this be so? The errors which remain after measurements have been taken should not be highly correlated, as a strongly correlated component should have been reflected in the measurements. Thus, it is the short-range errors, proportionately less observed, which remain. Thus we conclude that Φ *is* strongly sensitive to the measurement structure, and neither the Lyapunov nor the Riccati correlations are particularly good choices for Φ.

This is challenging and frustrating because the 2D covariance extension problem [26], [36] – how to infer a valid Φ from a sparse subset of correlations – is unsolved. We propose to circumvent the issue by modelling Φ in a *parameterized* form which is known to be positive definite. In our particular context, diffusion dynamics imply an exponential correlation structure [14], which is known to be positive definite when sampled in two dimensions.[3] Consequently, we propose to model the error structures $\Phi(t|t), \Phi(t+1|t)$ by an exponential model

$$\rho_{(i,j)(i+\tau_i,j+\tau_j)} = e^{-\left\{\frac{|\tau_i|+|\tau_j|}{l_{(i,j)(i+\tau_i,j+\tau_j)}}\right\}} \tag{17}$$

[2]Clearly the Riccati solution depends on the number, quality, and distribution of measurements. If we refer to "the Riccati solution," we mean the extreme case of dense measurements, with the measurement error variance set to the value expected in the problem context.

[3]There are other analytic positive-definite functions, of course: Gaussian, spherical, logistic, Bessel etc.

where l represents the correlation length.

Given (17), our former problem — how to implement the prediction of (12) — reduces to the prediction of the diagonal elements $\underline{p}$ and the correlation lengths $L(t+1|t)$. Our approach to both of these follows from the construction of relationships between the error variances and correlation lengths:

$$\begin{aligned} \underline{\tilde{p}}(t|t) \quad &\Longrightarrow \quad L(t|t) \\ \underline{\tilde{p}}(t+1|t) \quad &\Longrightarrow \quad L(t+1|t) \end{aligned} \tag{18}$$

Such a relationship is not obvious, yet it is implied by the discussion following (16). Indeed, for the static problem of (16) (and, in fact, for all stationary exponential cases) the correlation length is *linearly proportional* to the variance:

$$\frac{\tilde{p}}{p_o} = \frac{\tilde{l}}{l_o} \tag{19}$$

and justifies the premise of a $\tilde{p} \Leftrightarrow L$ relationship, where p_o, l_o are the process (Lyapunov) variance and correlation length, respectively.

For *non*stationary problems, however, this linear relationship no longer applies, at least in part because of ambiguities in even defining "correlation length" in such a case, yet the empirical $\tilde{p} \Leftrightarrow L$ relationship remains close to linear for all but small $\tilde{p}$, as is shown in Figure 3(a), in which posterior nonstationarities are introduced by measuring one pixel out of a state of five elements.

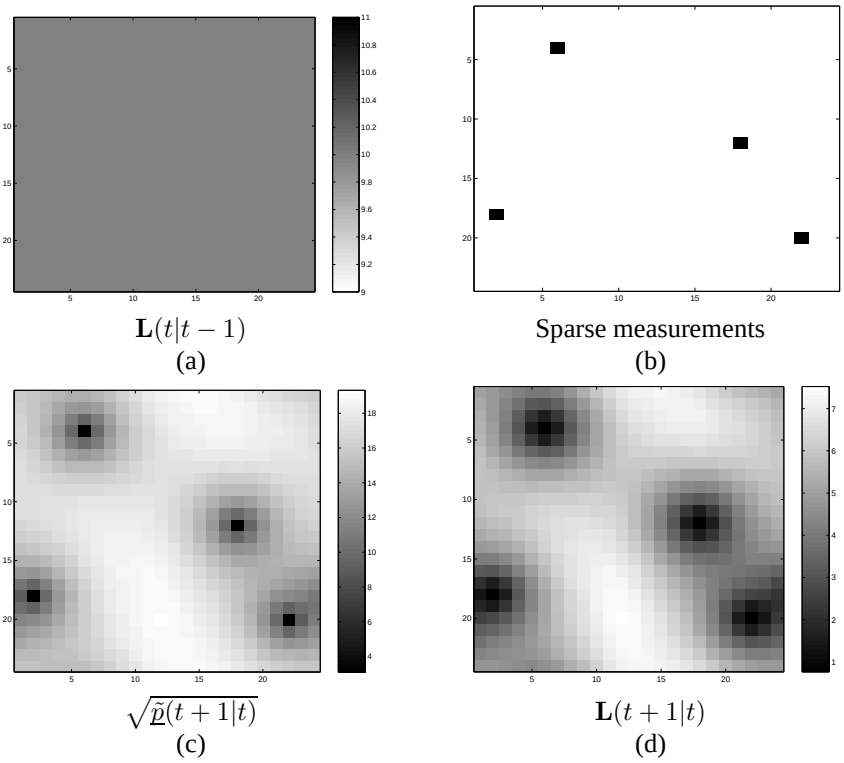
If we introduce dynamics and run a Kalman filter for a small-sized problem, this relationship remains, as seen in Figure 3(b,c). Note that this empirical relationship does not imply that the correlation length depends on the error variances only, rather that both are highly-similar functions of other factors: the quality and number of measurements at a given position, process covariance, the underlying dynamics etc.

In general the relationship may be assumed linear, or inferred experimentally over the range $[l_{small}\ ,\ l_{large}]$ of possible correlation lengths:

$$l_{small} = l_{Riccati}, \quad l_{large} = l_{Lyapunov}. \tag{20}$$

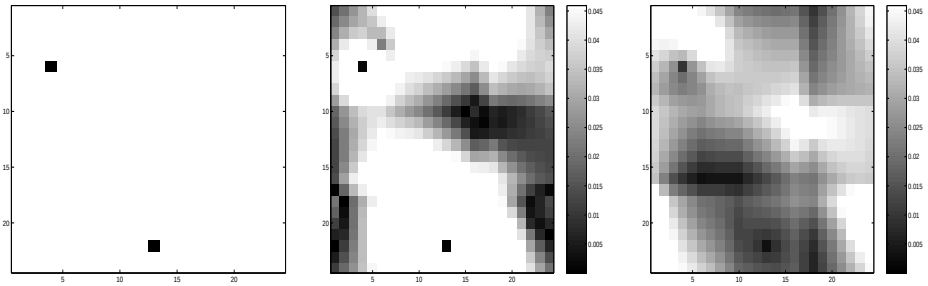
We complete the error prediction step by considering the propagation of the updated error variances

$$\underline{\tilde{p}}(t+1|t) = \text{diag}(A\tilde{P}(t|t)A^T + Q). \tag{21}$$



$\mathbf{L}(t|t-1)$ (a) Sparse measurements (b) $\sqrt{\tilde{p}(t+1|t)}$ (c) $\mathbf{L}(t+1|t)$ (d)

Fig. 4. Prediction: (a) The initial stationary prior model, which is exactly updated with (b) four point measurements. After twenty prediction steps, (c) the error standard deviations $\sqrt{\tilde{p}(20|19)}$ and (d) the estimation error correlation length $L(20|19)$ are both clearly nonstationary.



(a) (b) (c)

Fig. 5. Fixed correlation-length prior models: Suppose we wish to fuse two new measurements (a) with the nonstationary prior model from Figure 4(c,d). We use two stationary prior models: (b) a short correlation length $l_1 = 0.6$, and (c) a longer correlation $l_2 = 9$. The panels plot the degree of approximation (23), where dark shades represent higher quality. Note that although both (b),(c) contain regions of signifi cant error, an excellent estimated image could be found through the selection of appropriate subsets of (b) and (c).

Because of the sparseness of A, this product can be undertaken exactly. Recall that the discretized temporal dynamics matrix A is represented implicitly; consequently finding any particular predicted error variance amounts to modulating the kernel values A_k by the updated error variances and joint statistics. Because of the mixing effect of the dynamics, it is important to recognize that computing just the diagonal elements of the predicted error *will* require some of the off-diagonal elements of the updated error covariance $\tilde{P}(t|t)$. These elements may be available directly from the algorithm performing the update step, or computed from the empirical relationship between $\tilde{p}$ and l. This computation is not demanding as the number of necessary cross-covariances is very small in comparison to the size of the covariance matrix.

The prediction step is therefore complete: given the updated estimation error variances $\underline{\tilde{p}}(t|t)$, the associated correlation lengths $L(t|t)$ are inferred, leading to the computation of $\underline{\tilde{p}}(t+1|t)$ and $L(t+1|t)$, from which the predicted error covariance is finally computed as

$$\hat{P}(t+1|t) = \{\underline{\tilde{p}}(t+1|t)\,\underline{\tilde{p}}(t+1|t)^T\}^{\frac{1}{2}} \odot \Phi(t+1|t) \quad (22)$$

The process is illustrated in Figure 4: we start with a stationary covariance model, introduce four sparse measurements and perform an exact update, and then apply twenty prediction steps.

III. UPDATE

Typically the most challenging aspect of the update step is the matrix inversion in (3). Here we propose to use a multiscale

approach which efficiently solves the update step and produces $\underline{\tilde{p}}(t|t)$, as needed for prediction. The method, which has previously been applied in a wide variety of image processing [13], [46], [20], [33], [38] and remote sensing contexts [21], [41], models a two-dimensional field on a quad-tree. The hierarchical nature of the model admits extremely efficient estimators, albeit introducing new challenges in terms of finding accurate and appropriate tree-models. The details of the multiscale approach are not important; the methods of this paper are in no way multiscale-specific, and the multiscale method is treated as a black box. Interested readers will find the multiscale algorithm described in detail in [21], [38] and code is available on-line.[4]

Many efficient estimators, including multiscale, apply most easily to *static* problems. Challenges arise from *dynamic* estimation [30], [31] because the error statistics normally become non-stationary: in Figure 4 the correlation length varies spatially from short (near measured pixels) to large (far away). Thus even our proposed parameterization has problems, in that (17) is guaranteed to be positive definite only for a *fixed* (i.e., stationary) correlation length, *not* a space-varying one. One possibility is to update the non-stationary prior model (for example, that from Figure 4 (c,d)), but approximated by a stationary prior. We can measure the error in the approximation relative to the expected size of the estimation error $\sqrt{\tilde{p}(l)}$, paying a greater attention paid to those pixels where the estimator should be doing better:

$$\text{Approximation Error} = \frac{|\hat{x} - \hat{x}_{\text{approx}}|}{\sqrt{\tilde{p}}}, \tag{23}$$

where $\hat{x}$ is the best or exact state estimate, computed using the correct nonstationary prior model, and where $\hat{x}_{\text{approx}}$ is an approximate estimate, using a prior based on fixed (stationary) correlation lengths.

For example, consider the non-stationary prior model from Figure 4 (c,d), but approximated with a stationary prior. Figure 5 (b,c) plots the approximation (23) in the estimates $\hat{x}(l)$, computed using fixed correlation length l. Clearly the estimates computed using a short correlation length, Figure 5(b), are more accurate in those parts of the prior, Figure 4 (d), having a short correlation length, and similarly Figure 5(c) for longer lengths.

The key insight is that although both Figure 5(b),(c) contain regions of significant error, an excellent set of estimates could have been found through the selection of appropriate *subsets* of (b) and (c). To extend this idea, can we interpolate the two sets of estimates to acquire even better estimates for intermediate correlation lengths?

In other words, given two stationary static estimates $\underline{\hat{x}}(l_1)$, $\underline{\hat{x}}(l_2)$ and associated estimation error variances $\underline{\tilde{p}}(l_1), \underline{\tilde{p}}(l_2)$, how can we interpolate them effectively such that the validity or accuracy of the interpolated estimates is high over the whole domain? In addition, the estimates and error variances should be smooth, without blocky artifacts at artificial boundaries.[5]

[4]Several versions of the multiscale estimator are available at *ocho.uwaterloo.ca*.

[5]As would, for example, be introduced by a decoupled piecewise-constant correlation prior.

Let us generalize the problem by approximating the estimates $\hat{x}$ and error variances $\tilde{p}$ by interpolating K stationary prior models

$$\hat{\hat{x}}(i,j) = \sum_{k=1}^{K} \hat{x}(i,j,l_k)\alpha_x(l_k, l_{ij}) \tag{24}$$

$$\hat{\tilde{p}}(i,j) = \sum_{k=1}^{K} \tilde{p}(i,j,l_k)\alpha_p(l_k, l_{ij}) \tag{25}$$

where l_{ij} is the nominal, space-varying correlation length of the prior model.

There are two issues here: how to compute the best interpolating weights α_x, α_p, and how to choose the optimum set of values l_k for the interpolants?

A. Optimum Weight Computation

Suppose we are given the K interpolating lengths

$$\{l_{small} = l_1, l_2, l_3, \cdots, l_K = l_{large}\}$$

which span the entire expected range $[l_{small}, l_{large}]$ of correlation lengths. Then the weights associated with any fixed length $l_{small} \le l \le l_{large}$ can be estimated using least-squares.

Suppose we first consider a stationary estimation problem with a fixed prior correlation length of l. Then the exact estimates and error variances may be written as

$$\widehat{\underline{x}}(l) = \sum_{k=1}^{K} \underline{\hat{x}}(l_k)\alpha_x(l_k, l) + \underline{e}_x \tag{26}$$

$$\underline{\tilde{p}}(l) = \sum_{k=1}^{K} \underline{\tilde{p}}(l_k)\alpha_p(l_k, l) + \underline{e}_p \tag{27}$$

where $\underline{e}_x, \underline{e}_p$ represent the errors in the interpolation. To keep the interpolation unbiased the sum of the weights is forced to equal one,

$$\alpha_x(l_1, l) = 1 - \sum_{k=2}^{K} \alpha_x(l_k, l), \quad \alpha_p(l_1, l) = 1 - \sum_{k=2}^{K} \alpha_p(l_k, l) \tag{28}$$

such that the interpolation becomes

$$\widehat{\underline{x}}(l) = \underline{\hat{x}}(l_1) + \sum_{k=2}^{K} (\underline{\hat{x}}(l_k) - \underline{\hat{x}}(l_1))\alpha_x(l_k, l) + \underline{e}_x \tag{29}$$

$$= \underline{\hat{x}}(l_1) + H_x \underline{\alpha}_x(l) + \underline{e}_x \tag{30}$$

$$\underline{\tilde{p}}(l) = \underline{\tilde{p}}(l_1) + \sum_{k=2}^{K} (\underline{\tilde{p}}(l_k) - \underline{\tilde{p}}(l_1))\alpha_p(l_k, l) + \underline{e}_p \tag{31}$$

$$= \underline{\tilde{p}}(l_1) + H_p \underline{\alpha}_p(l) + \underline{e}_p \tag{32}$$

In order to estimate α_x, α_p we need the exact estimates $\widehat{\underline{x}}(l), \underline{\tilde{p}}(l)$. Based on our chosen correlation function (17) we can generate random sample paths for a small-sized 2D process, from which measurements are taken and exact estimates are computed. The squared interpolation error $\| \underline{e}_x \|^2$ is minimized by solving a least-squares problem, whose well-known solution is

$$\widehat{\underline{\alpha}}_x(l) = ({H_x}^T H_x)^{-1} {H_x}^T (\widehat{\underline{x}}(l) - \widehat{\underline{x}}(l_1)). \tag{33}$$

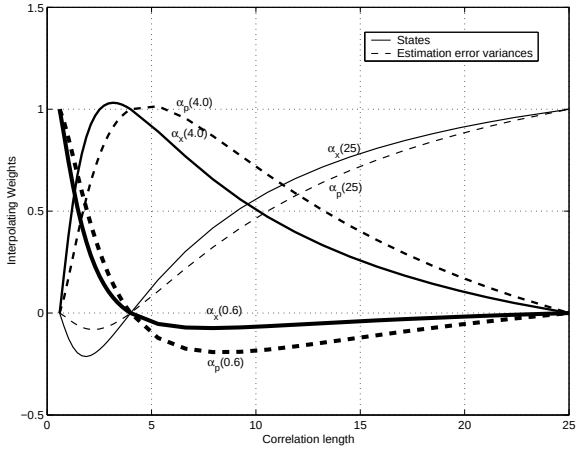


Fig. 6. The general shape of the interpolating weights for the states α_x (solid lines) and the estimation error variances α_p (dashed lines). The weights are constrained to sum to one.

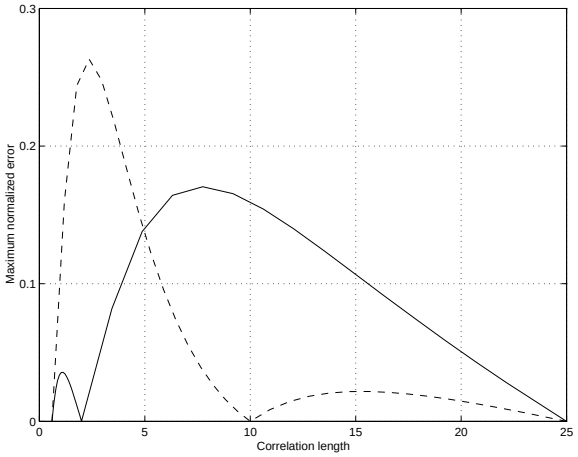


Fig. 7. Illustration of the effect of the intermediate correlation length position on the maximum error in the 2D domain: solid, $l_2 = 2$; dashed, $l_2 = 10$.

and similarly for $\underline{\hat{\alpha}}_p(l)$. Once the estimated weights $\{\hat{\alpha_x}(l_i, l)\}, \{\hat{\alpha_p}(l_i, l)\}$ have been computed as a function of l, we can express the state estimates $\underline{\widehat{\hat{x}}}(l)$ and the error variance estimates $\underline{\widehat{\tilde{p}}}(l)$ at any correlation length l.

Figure 6 shows one example of $K = 3$ interpolating weights $[l_{small} = 0.6\ , l_2 = 0.4,\ l_{large} = 25]$. As expected, the weights sum to one, and the interpolation is exact at correlation lengths 0.6, 4.0, 25; further, the weights decay in value as we move away from the corresponding correlation length. The figure does raise the question how the choice of intermediate length l_2, or the inclusion of additional lengths, affects the accuracy of the interpolation, discussed next.

B. Optimum Correlation Lengths

There are two critical factors affecting the computation of the interpolating weights: the number of interpolating priors K and the particular choice of correlation lengths $\{l_i\}$.

We propose to optimize the set of weights by minimizing the worst-case fractional error (23):

$$\text{Minimize} \qquad \max_{l} \left(\max_{i,j} \frac{|\hat{x}(i,j,l) - \widehat{\hat{x}}(i,j,l)|}{\sqrt{\tilde{p}(i,j,l)}} \right). \quad (34)$$

The sensitivity of this criterion to a change in the correlation lengths $\{l_k\}$ is depicted in Figure 7. The figure illustrates the case of $K = 3$, where two different values are chosen for the intermediate length l_2. It is obvious that modest changes in l_2 can have a very substantial affect on the quality of the interpolation, and that the quality can vary significantly with l.

To solve for the optimal lengths, given K, we start with a heuristic initial guess for the intermediate correlation lengths $l_2, \ldots, l_{K-1}$, from which the optimization problem (34) is sufficiently straightforward to be solved using a standard direct search (MATLAB *fmins*). Two examples are depicted in Figure 11(a),(b), for $K = 3, 5$[6] respectively. Examining the two panels, we observe that the worst-case relative error for $K = 3$ in Figure 11(a) is 0.1 standard-deviations; as we increase the number of interpolants to $K = 5$ in (b), the relative error drops to just 0.02 standard-deviations. That is, with $K = 5$, our approximated update will induce errors no greater than one *fiftieth* of a 1σ error bar for any stationary process.

Table I generalizes the above figures to give the dependence of the relative error (34) as a function of the maximum correlation length and the number of interpolants K. Clearly, as the number of interpolating correlation lengths increases the relative error drops to insignificance. The maximum correlation length is normally determined by the context of the estimation problem; as l_{large} increases the interpolation problem becomes more challenging, so a larger value of K is required to maintain a fixed level of update accuracy.

Thus, to recap, we have derived weights $\alpha_x(l_k, l), \alpha_p(l_k, l)$ allowing us to interpolate a modest number of stationary estimates to any desired correlation length. Figure 11 and Table I both test stationary examples, in which the prior correlation l is spatially constant. The use of this interpolating approach for *non*stationary l is considered next.

IV. EXPERIMENTAL RESULTS

Two sets of examples are presented: synthetic toy-problems, where "truth" is known and where exact estimates can be computed by brute force, and a large-scale ocean sea-surface-temperature estimation, to illustrate the applicability of our approach to large problems of current interest.

Process correlation length	*Number of priors K*			
	3	5	7	9
3	0.0109	0.0013	0.0004	0.0002
8	0.0384	0.0070	0.0022	0.0009
17	0.0789	0.0145	0.0045	0.0019
25	0.0996	0.0183	0.0056	0.0024

TABLE I

MAXIMUM ERROR (34), MEASURED IN STANDARD DEVIATIONS, AS A FUNCTION OF K AND PROCESS CORRELATION LENGTH.

[6]To keep the interpolation simple, only three $\{l_i\}$ adjacent to l are used. Clearly the choice of three correlation lengths is arbitary, and the method extends trivially to any number.

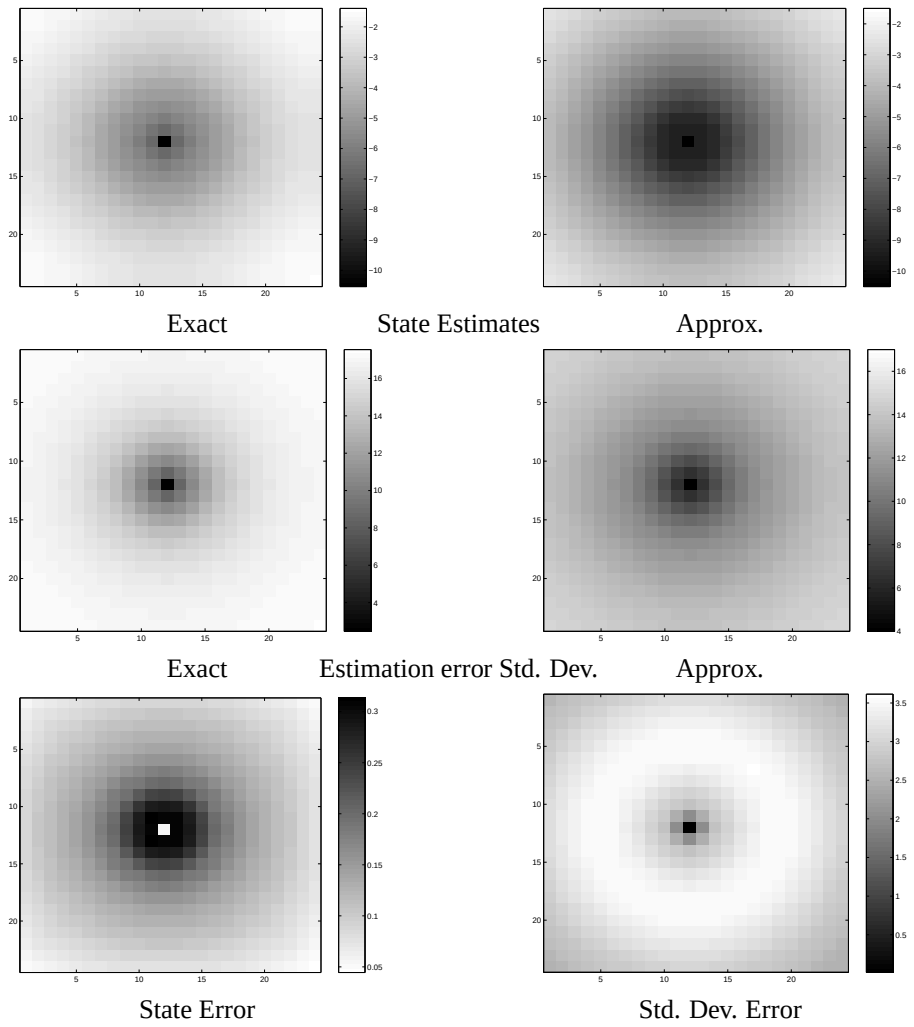
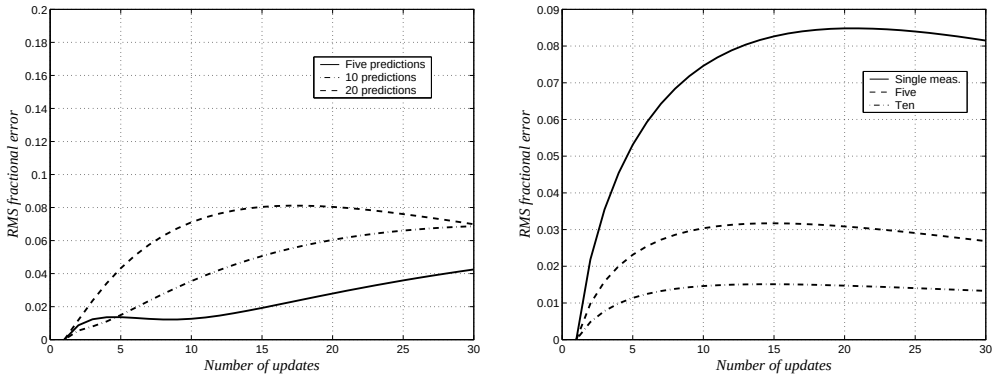
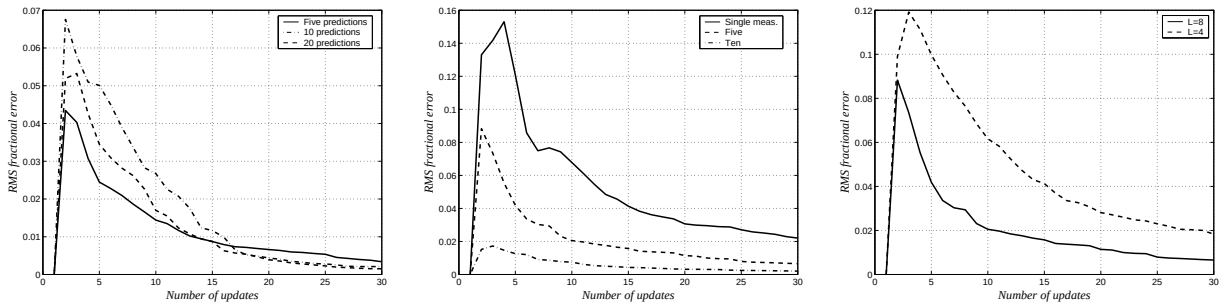


Fig. 8. A single measurement test, comparing the exact and approximate estimates and error standard deviations. The state error is the normalized difference (23); the error in standard deviation is plotted as the absolute difference.



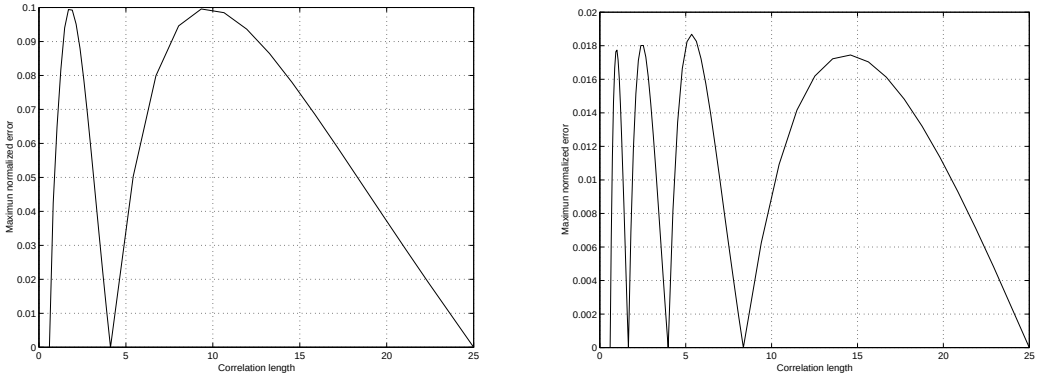
(a) (b)

Fig. 9. RMS fractional error (35) for two cases: (a) The effect of the number of prediction steps, five pixels observed; (b) the effect of the number of measurements, one prediction step per update.



(a) (b) (c)

Fig. 10. RMS of fractional error (35) for time-varying measurement positions varying (a) the number of prediction steps per update, (b) the number of measurements, and (c) the process correlation length.



(a) (b)

Fig. 11. Two examples of optimizing the intermediate correlation lengths l_i.

A. Synthetic: Time-invariant

Consider a two-dimensional 24×24 pixel diffusion problem with a process steady-state correlation length $l_{large} = 7$. We will approximate the estimation problem using the prediction / update methods developed in Sections II and III, with $K = 5$.

Suppose we have a single measurement, sampled at each time step with a measurement error variance of 4. Figure 8 depicts the exact estimation results obtained by the Kalman filter (left) and by our approximate approach (right) after twenty time steps.

The approximate estimates and error variances are free of error at the measured position and its neighbouring pixels, with the approximate filter tending to underestimate the estimation error variances for elements away from the measured position. The bottom panels in the figure show the fractional and absolute errors of the estimates and standard deviations, respectively. Because of the high accuracy of the update step, particularly with $K = 5$ interpolants, the bulk of the error can be attributed to the prediction step. It is important to note that the single-measurement test of Figure 8 is really a worst-case example: as more pixels are measured, the quality of the estimates depends to a lesser degree on the precision of the prediction and update steps.

Next, consider measuring more than one pixel. Since the estimation error variance $\tilde{\underline{p}}$ is a function of the number of measurements, in comparing update steps with different numbers of measurements a slight variation on (34) should be considered. We define the root-mean-square fractional error as

$$\mathrm{RMSFE}(l) = \frac{\sqrt{\frac{1}{n^2}\sum \frac{\left(\hat{\underline{x}}(l) - \hat{\tilde{\underline{x}}}(l)\right)^2}{\tilde{\underline{p}}(l)}}}{\sqrt{\frac{1}{n^2}\sum \frac{\hat{\underline{x}}(l)^2}{\tilde{\underline{p}}(l)}}} \tag{35}$$

where all of the divisions are point-wise. This measures the size of the errors $\hat{\underline{x}}(l) - \hat{\tilde{\underline{x}}}(l)$ relative to the estimates $\hat{\underline{x}}(l)$, modulated by the expected size of the estimation error $\sqrt{\tilde{\underline{p}}(l)}$.

In Figure 9(a) the RMSFE is plotted as a function of number of updates for various number of prediction steps (five, ten, twenty) between successive update steps. It is obvious that increasing the number of predictions magnifies the approximation effect imposed by our method, although the approximation error never exceeds 10% of the estimates.

A second test, plotting the RMSFE as a function of the number of measurements, is illustrated in Figure 9(b). As is consistent with expectation, an increase in the number of measurements leads to a reduction in the fractional error.

B. Synthetic: Time-varying

From the remote-sensing perspective, a more realistic problem has the measurement positions vary over time. The following experiment is the same as in the previous section, except that the number and positions of observations change randomly over time.

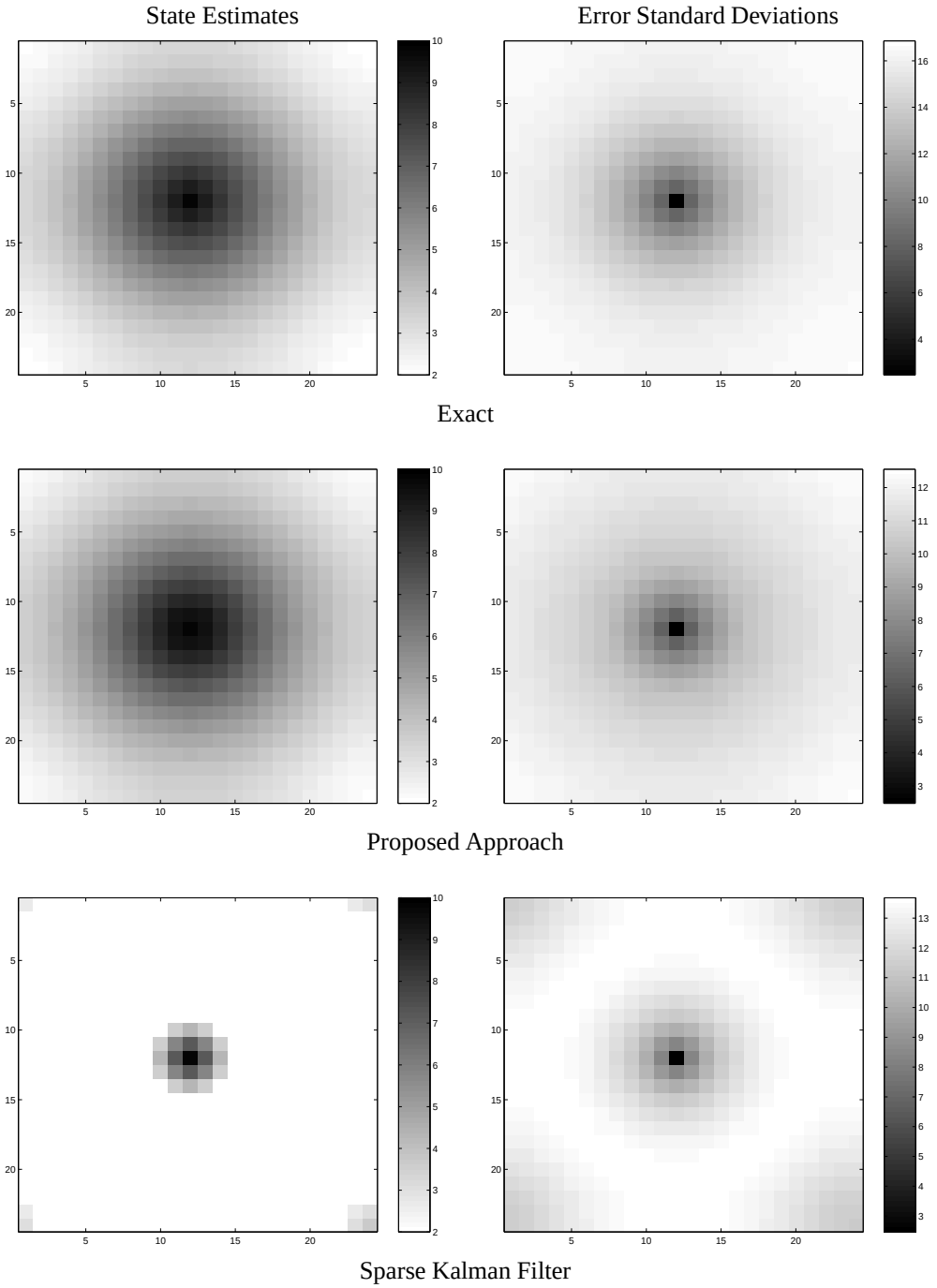


Fig. 12. Comparison of the exact Kalman filter with our proposed, parameterized form and with a sparse Kalman filter based on a third-order neighbourhood structure.

Figure 10 shows three tests, where the RMSFE (35) is plotted as a function of the number of update steps for varying numbers of prediction steps or measurements, and different process correlation lengths. As is expected, fewer prediction steps, more measurements, or a greater correlation length all contribute to a reduced RMSFE.

All three tests show a qualitatively similar behaviour in RMSFE: initially small, growing rapidly, and then decreasing over time. Because the process is started in steady-state, the prior model will have correlation length l_{large}, thus the first update step will produce exact estimates and the RMSFE begins at zero. As measurements are introduced the problem becomes nonstationary, and the resulting approximations cause the RMSFE to increase. However as the random process becomes more thoroughly measured, the error statistics become more uniform and the RMSFE decreases. Clearly the size and width of the peak will be a function of the time step, the rate of change of the random field, and the number and distribution of measurements.

C. Sparse Kalman Filter Comparison

As discussed in the Introduction, one common approach to solving dynamic estimation problems is the sparse Kalman filter [2], [8], [9], [11], [12]. The approach is based on the premises that the error covariance inverses can be approximated as banded, implying a Markov structure for the error field. Our key interests concern testing efficiency, accuracy, and stability.

The computational demands for this method are dependent on the number of bands b_1, b_2 kept in the dynamics A and covariance inverses P^{-1} respectively (and on the number of terms p in the case of polynomial approximation for matrix inversion). The dynamics are often banded (penta-diagonal, in our case, allowing A to be represented exactly); the challenge normally stems from limiting b_2 and p. The computational complexity of the error prediction step is dominated by matrix multiplication, of order $\mathcal{O}(2 \cdot b_1 \cdot b_2 \cdot n)$. The complexity of the update step is determined by computing the matrix inversion $\mathcal{O}({b_2}^p \cdot n)$, the gain matrix $\mathcal{O}({b_2}^2 \cdot n)$, and the updated covariance $\mathcal{O}({b_2}^2 \cdot n)$. There are stability concerns, however, because not every banded covariance inverse is positive definite, as we shall see.

For our proposed approach, on the other hand, the error prediction step is extremely fast and has linear complexity $\mathcal{O}(n)$, since each pixel is predicted locally. The complexity of the update step will depend on the number of interpolants K and on the estimation algorithm chosen to implement this step. If the multiscale approach [13], [33], [38] is used, as proposed, then the update step has complexity $\mathcal{O}(K \cdot n^{\frac{3}{2}})$, and stability is guaranteed since each of the K subproblems is constructed to be positive definite. In this particular example, since we are interested in testing the proposed method of interpolating updates rather than the details of a multiscale implementation, each of the individual, stationary updates are computed exactly.

Figure 12 shows the comparison between the two approaches for the case of a 24×24 process. The matrix inversion of the sparse approach is computed exactly, so Figure 12 presents an optimistic level of performance for the sparse method, and the degree of approximation should be expected to be inferior in more realistic settings. We have chosen $b_1 = 5$, representing the dynamics exactly; the only approximation enters with the representation of the covariance inverse, with $b_2 = 12$ bands. The update step in our proposed method is based on $K = 5$ interpolants.

The results are depicted after 50 update steps, comparing the estimates and error variances for the exact, sparse, and proposed methods. The sparse state estimates and error statistics fail to propagate very far from the measured location and are quite poor, compared to the *much* more accurate results produced by our proposed method.

For the same experimental context from above, Figure 13 shows the accuracy, measured as RMS fractional error, of the two approaches. Given that *all* of the approximation in the sparse approach lies in the representation of the error covariance, controlled by b_2, six different values of b_2 are tested. Not only is the RMSFE *much* greater for the sparse approach than for our proposed one, but in about half of the sparse experiments the estimator becomes unstable (negative-definite covariances), whereas our proposed approach guarantees stability.

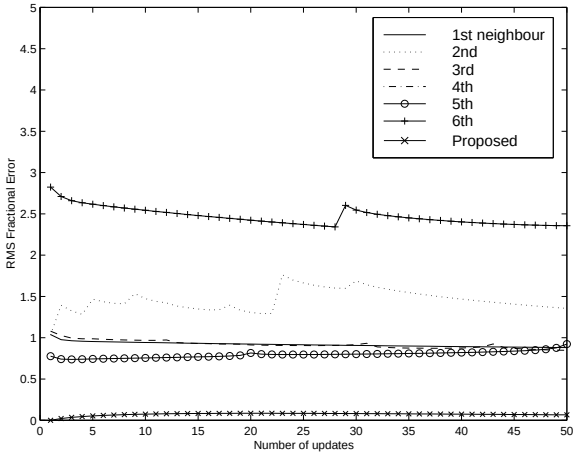


Fig. 13. RMS Fractional Error (35) for our proposed approach and sparse Kalman filters with varying numbers of bands: $b_2 = 5, 9, 13, 21, 25, 29$ corresponding to increasing neighbourhood orders. The erratic nature of the sparse-KF curves is attributed to stability problems —closeness to singularity or negative definiteness of state covariances.

D. Remote Sensing

Ultimately the main motivation of this research is to apply it in solving large-scale 2D dynamic estimation problems, for example the estimation of the sea surface temperature (SST) from a sequence of sparse 2-D satellite measurements (e.g., SSM/I[32], ATSR[43], [44], or AVHRR[15]). SST is a quantity of fundamental significance in climate monitoring, numerical weather prediction and oceanography, for example in the early detection of climate anomalies such as El Nino[44] or global warming. We are using SST observations from the Along-Track Scanning Radiometer (ATSR), whose daily global coverage is less than 20%, and many areas are subject to persistent cloudiness, so there are extensive regional gaps.

We estimate temperature anomalies (that is, mean-removed) based on mean-removed measurements and a diffusive model for the ocean dynamics. We set the diffusion parameters to approximate the empirical statistics, such that the process spatial correlation length is 20 pixels (5 degrees), and a corresponding temporal correlation length of 300 prediction steps (ten days); in practice an experienced ocean scientist would set these parameters (possibly nonstationarily), guided by additional insights. The model is initialized with a spatial prior model, isotropic exponential with a correlation length of 20 pixels and a standard deviation of 70 K. The update steps are twelve hours apart, such that night-time and day-time measurements are separately included, and fifteen prediction steps are applied between successive updates.

The resulting temperature estimates and corresponding estimation error standard deviations are shown in Figure 14. The estimated images are dense, smooth, and do not possess any blocky artifacts arising from the estimation process. The error standard deviations very clearly show the smooth evolution over time from certainty to uncertainty, as more recent measurements correspond to regions of greater certainty than past ones.

Clearly the model employed here is only an approximation, both in terms of the heuristic choices in diffusion parameters,

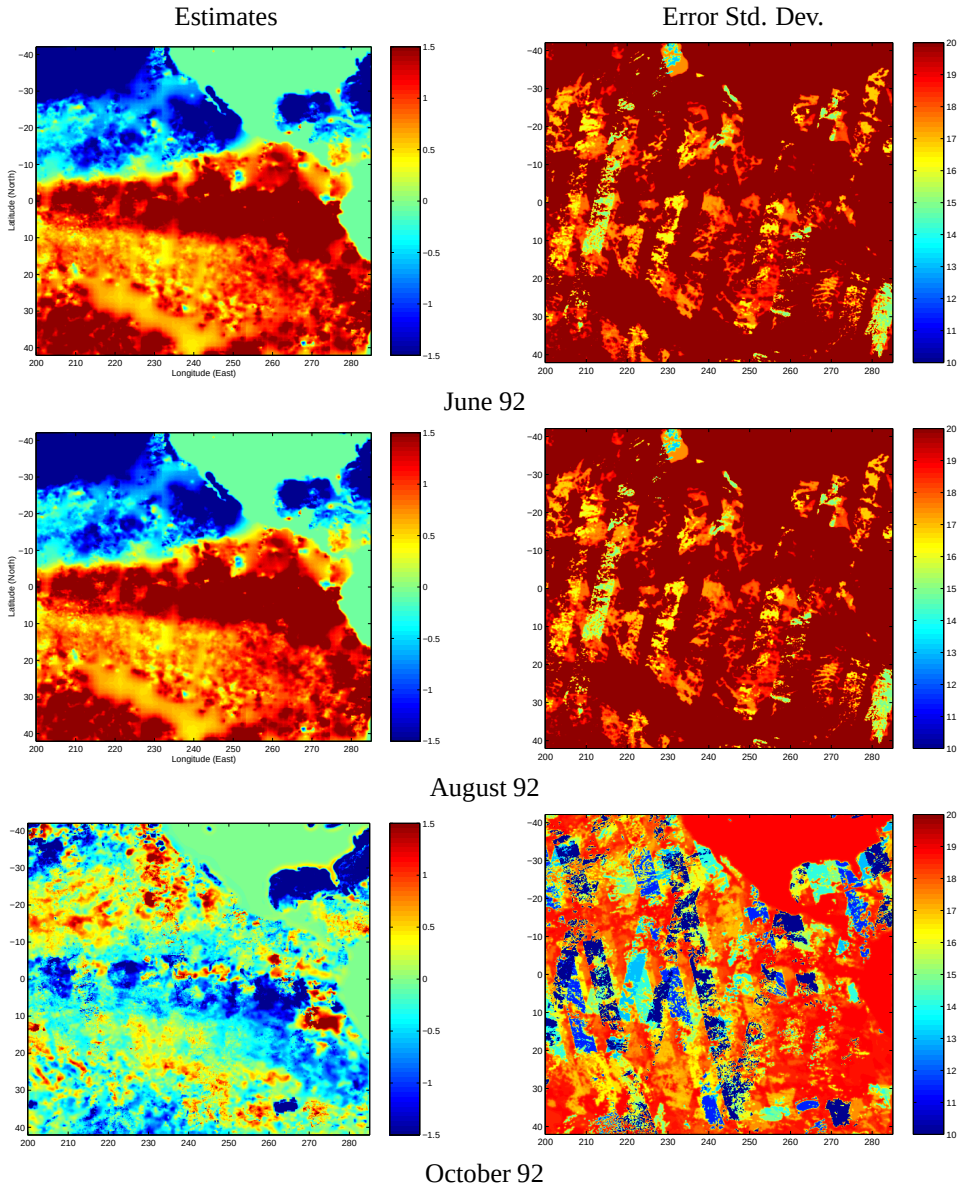


Fig. 14. Anomaly estimates and error standard deviations for the proposed estimator, applied to five months of ATSR sea surface temperature data, starting in June, 1992.

as well as limiting ourselves to a strictly diffusive model, which fails to consider any surface motion / flows. An advective term, based on known or inferred currents, would lead to improved estimates.

V. Conclusions

The dynamic estimation and filtering of large data sets over time is a substantial challenge of considerable interest in remote sensing. The ability to estimate periodic (hourly, daily, etc) snapshots is a critical step in the production of operational data products from sparse satellite data.

In this paper we have proposed a highly efficient approximate variation of the Kalman filter. The static step, the primary contribution of this paper, uses a mixture of stationary models to accurately mimic the effect of a nonstationary prior. Not only is this effective in simplifying the computational complexity, but it also simplifies modeling. Given a spatial map of correlation length, an ad-hoc derived prior model may be negative definite or very close to singular, and the absence of an efficient solution to two-dimensional covariance extension precludes finding an exact prior. However our approach provides an efficient, stable, positive-definite model which is consistent with the given correlation structure. Thus the methods of this paper may find application in modeling, single-frame estimation, and dynamic estimation.

References

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

[2] A. Asif and J. Moura. Data assimilation in large time-varying multidimensional fields. *IEEE Transactions on Image Processing*, Vol. 8, No. 11:1593–1607, 1999.

[3] M. Basseville, A. Benvensite, K. Chou, S. Golden, R. Nikoukhan, and A. Willsky. Modeling and estimation of multiresolution stochastic processes. *IEEE Trans on Information Theory*, Vol. 38, No. 2:766–784, 1992.

[4] D. Boggs, M. Ghil, and G. Keppenne. A stabilized sparse-matrix u-d square-root implementation of a large-state extended kalman filter. *Proc. 2nd International Symposium on Assimilation of Observations*, Vol. 1:219–224, 1995.

[5] J. Brailean and A. Katsaggelos. Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences. *IEEE Trans. Image Processing*, Vol. 4, No. 9:1236–1251, 1995.

[6] M.A. Cane, A. Kaplan, R.N. Miller, B. Tang, E.C. Hackert, and A.J. Busalacchi. Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *J. Geophysical Research*, 101:22,599–22,617, 1996.

[7] D. Cano and M. Benard. 3d kalman filtering of image sequences. In *Image Sequence Processing and Dynamic Scene Analysis*, pages 563–579. Springer-Verlag, 1993.

[8] M. Chin, W. Karl, and A.S. Willsky. A distributed and iterative method for square root filtering in space-time estimation. *Automatica*, Vol. 31:67–82, 1995.

[9] M. Chin, A.J. Mariano, and E.P. Chassignet. Spatial regression and multi-scale approximations for sequential data assimilation in ocean models. *Journal of Geophysical Research*, Vol. 104:7991–8014, 1999.

[10] T. Chin. *Dynamical Estimation in Computational Vision*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1991.

[11] T. Chin. On Kalman filter solution of space-time interpolation. *IEEE Transactions on Image Processing*, Vol. 10, No. 4:663–666, 2001.

[12] T. Chin, W. Karl, and A. Willsky. Sequential filtering for multi-frame visual reconstruction. *Signal Processing*, Vol. 28:311–333, 1992.

[13] K. Chou, A. Willsky, and A. Benvensite. Multiscale recursive estimation, data fusion, and regularization. *IEEE Trans on Automatic Control*, Vol. 39, No. 3:464–477, 1994.

[14] J. Coleman. *Gaussian Spacetime Models: Markov Field Properties*. PhD thesis, University of California at Davis, David, CA, 1995.

[15] W.J. Emery. The advanced very high resolution radiometer (avhrr): A brief reference guide. *Photogrammetric Engineering and Remote Sensing*, Vol. 58, No. 8:1183–1188, 1992.

[16] S. Erturk. Image sequence stabilisation based on kalman filtering of frame positions. *Electronics Letters*, Vol. 37, No. 20:1217–1219, 2001.

[17] S. Omatu et. al. Estimation of nitrogen dioxide concentrations in the vicinity of a roadway by optimal filtering theory. *Automatica*, Vol. 24, No. 1:19–29, 1988.

[18] S. Farlow. *Partial Differential Equations for Scientists and Engineers*. Dover Publications, Inc., New York, 1993.

[19] P. Fieguth, M. Allen, and M. Murray. Hierarchical methods for global-scale estimation problems. *In Proceedings of Canadian Conference on Electrical Engineering, Waterloo, Canada*, pages 161–164, 1998.

[20] P. Fieguth, W. Karl, and A. Willsky. Efficient multiresolution counterparts to variational methods for surface reconstruction. *Computer Vision and Image Understanding*, Vol. 70, No. 2:157–176, 1998.

[21] P. Fieguth, W. Karl, A. Willsky, and C. Wunsch. Multiresolution optimal interpolation and statistical analysis of topex/poseidon satellite altimetry. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 33, No. 2:280–292, 1995.

[22] P. Fieguth, F. Khellah, M. Allen, and M. Murray. Large scale dynamic estimation of ocean surface temperature. *In Proceedings of IEEE 1999 International Geoscience and Remote Sensing Symposium*, Vol. 3:1826–1828, 1998.

[23] P. Fieguth, D. Menemenlis, and I. Fukumori. Mapping and pseudo-inverse algorithms for ocean data assimilation. *IEEE Trans. Geoscience and Remote Sensing*, 41:43–51, 2003.

[24] P. Fieguth and A. Willsky. Fractal estimation using models on multiscale trees. *IEEE trans. on Signal Processing*, Vol. 44, No. 5:1297–1300, 1996.

[25] A. Frakt. *Internal multiscale autoregressive processes, stochastic realization, and covariance extension*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1997.

[26] A. Frakt and A. Willsky. Computationally efficient stochastic realization for internal multiscale autoregressive models. *Multidimensional Systems and Signal Processing*, Vol. 12, No. 2:109–142, 2001.

[27] P. Gaspar and C. Wunsch. Estimates from altimeter data of baryotropic rossby waves in the northwestern atlantic ocean. *J. Physical Oceanography*, Vol. 19, No. 12:1821–1844, 1989.

[28] A. Gelb. *Applied Optimal Estimation*. MIT Press, Cambridge,MA, 1974.

[29] M. Grewal and A. Andrews. *Kalman Filtering*. Prentice-Hall, New Jersey, 1993.

[30] T. Ho. *Multiscale Modeling and Estimation of Large Scale Dynamic Systems*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1998.

[31] T. Ho, P. Fieguth, and A. Willsky. Computationally efficient steady-state multiscale estimation for 1-d diffusion processes. *Automatica*, Vol. 37:325–340, 2001.

[32] J. Hollinger, J.L. Peirce, and G.A. Poe. Ssm/i instrument evaluation. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 28, No. 5:781–790, 1990.

[33] W. Irving, P. Fieguth, and A. Willsky. An overlapping tree approach to multiscale stochastic modeling and estimation. *IEEE Trans. on Image Processing*, Vol. 6, No. 11:1517–1529, 1997.

[34] J. Kim and J. Woods. Spatio-temporal adapive 3-D Kalman filter for video. *IEEE Trans. on Image Process*, Vol. 6, No. 3:414–424, 1997.

[35] J. Kim and J.W. Woods. 3-d kalman filter for image motion estimation. *IEEE Trans. Image Processing*, Vol. 7, No. 1:42–52, 1998.

[36] H. Lev-Ari, S. Parker, and T. Kailath. Multidimensional. maximum-entropy covariance extension. *IEEE Trans. Information Theory*, Vol. 35:497–508, 1989.

[37] M. Luettgen. *Image Processing with Multiscale Stochastic Models*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1993.

[38] M. Luettgen, W. Karl, A. Willsky, and R. Tenney. Multiscale representations of markov random fields. *IEEE Trans. Signal Processing*, Vol. 41, No. 12:3377–3396, 1993.

[39] M. Luettgen and A. Willsky. Multiscale smoothing error models. *IEEE Trans. on Automatic Control*, Vol. 40, No. 1:173–175, 1995.

[40] L. Matties, T. Kanade, and R. Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, Vol. 3:209–236, 1989.

[41] D. Menemenlis, P. Fieguth, and A. Willsky. Adaption of fast optimal interpolation algorithm to the mapping of oceanographic data. *Journal of Geophysical Research*, Vol. 102, No. C5:10573–10583, 1997.

[42] D. Menemenlis, T. Webb, C. Wunsch, U. Send, and C. Hill. Basin-scale ocean circulation from combined altimetric, tomographic and model data. *Nature*, 385:618–621, 1997.

[43] M. Murray and M. Allen. Actual and potential information in dual-view radiometer observations of sea surface temperature from ATSR. *Journal of Geophysical Research*, Vol. 103, No. C4:8153–8166, 1998.

[44] C. Mutlow and A. Zavody. Sea surface temperature measurements by the along -track scanning radiometer on the ERS1 satellite: Early results. *Journal of Geophysical Research*, Vol. 99, No. C11:22575–22588, 1994.

[45] G. Noriega and S. Pasupathy. Application of kalman filtering to real-time processing of geophysical data. *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 30, No. 5:897–883, 1992.

[46] M. Schneider, P. Fieguth, W. Karl, and A. Willsky. Multiscale methods for segmentation and reconstruction of signals and images. *IEEE Trans. on Image Processing*, Vol. 9, No. 3:456–468, 2000.

[47] B. Suresh and B. Shenoi. New results in two-dimensional kalman filtering with application to image restoration. *IEEE Trans. on Circuits and Systems*, Vol. 28, No. 4:307–318, 1981.

[48] E. Twizell. *Computational Methods for Partial Differential Equations*. Wiley, New York, 1984.

[49] J. Woods and C. Radewan. Kalman filtering in two dimensions. *IEEE Trans. Inform. Theory*, Vol. IT-23:437–481, 1977.