INTRODUCTION TO THE KALMAN FILTER

Robert N Miller
College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, Oregon, U.S.A.

Summary: The Kalman filter is a statistically optimized data assimilation method. From its formulation in its present form in the early 1960's, it has seen widespread application in the engineering community. Since it is a filter, as opposed to a smoother, the result of its application takes into account only observations taken at the time at which the analysis is being conducted and at earlier times. In this respect, it is suitable for forecast systems, but less so for retrospective analyses in which analysis is desired for past times, and the analysis at a given time may be improved by consideration of future data, i.e., data gathered at subsequent times.

The Kalman filter is easily formulated as a weak-constraint method, i.e., the assumption that the model and forcing fields are perfectly known is unnecessary and does not make implementation any more convenient. All weak constraint methods are computationally demanding, and the choice between the Kalman filter and other weak-constraint techniques is often a tactical one, depending on such considerations as programming convenience.

The Kalman filter depends on prior estimates of error statistics; as such, it can be viewed as an exercise in Bayesian statistics: at a given time t_0, the "present," given a first guess $\mathbf{u}_0$ for the system state as the mean of a prior estimate of the probability density of the solution space, and the fact that a measurement has occurred at time t_0, what is the posterior conditional probability density, conditioned upon the fact that the observation has taken place with the given results? Under the assumptions that the dynamics are linear and the initial distribution and the distributions of the observation and forcing errors are Gaussian, the distribution of the filtered solution will remain Gaussian, and therefore completely specified by its mean and covariance.

Formulated in these statistical terms, the Kalman filter allows *a posteriori* analysis in terms of formal statistical hypothesis testing. One can then establish formal confidence limits on whether the filter is actually functioning in an optimal fashion, which it will be if the prior assumptions are in fact valid.

Here we derive the Kalman filter as part of the explicit solution to a weak constraint variational problem for a linear system. This relatively unconventional approach is intended to show the close relation between the Kalman filter and other variational methods such as 4DVAR.

Following the derivation, we give an example of application of the Kalman filter to a problem in ocean data assimilation, and then go on to discuss *a posteriori* statistical validation.

1. INTRODUCTION

The Kalman filter is an optimized method. It shares this attribute with the variational methods such as 4DVAR. Here I shall derive the Kalman filter directly from a simple minimization procedure for a discrete data assimilation system.

1.1 Data assimilation: smoothers and filters

In this section, we shall derive a very general data assimilation scheme based on a very simple discrete assimilation cycle. Let $\mathbf{u}_0$ be our best guess for the state of the system at time t_0. Assume the state vector $\mathbf{u}_0$ to have dimension N. The assumed prediction model is linear, and given by:

$$\mathbf{u}_1 = L\mathbf{u}_0 + \tau \tag{1}$$

The subscript "1" denotes quantities evaluated at time t_1, and τ is the forcing. Now, at this point in the assimilation cycle, a vector $\mathbf{y}$ of m observations becomes available. We do not assume that $\mathbf{y}$ is a direct observation of the entire state vector $\mathbf{u}_1$. We only assume that the quantity measured is linearly related to $\mathbf{u}_1$, i.e., if our model were perfect and had perfect initial conditions, and there were no observation errors, we would have $\mathbf{y} = H\mathbf{u}_1$, where H is a given $m \times N$ matrix. For example, if $\mathbf{y}$ consisted of the observation of the j^{th} component of $\mathbf{u}_1$, then H would consist of a single row vector with a 1 in the j^{th} position and zeros elsewhere. We may look at $H\mathbf{u}_1$ as our prior estimate of $\mathbf{y}$, based on past measurements and on our knowledge of the system.

The updating operation consists of finding optimal corrections $\mathbf{v}_0$ and $\mathbf{v}_1$ to $\mathbf{u}_0$ and $\mathbf{u}_1$ based on $\mathbf{y}$. The easiest assumption to make is that L and τ are known exactly, and all of the errors in the system are due to mis-specification of $\mathbf{u}_0$. The optimized solution will be of the form:

$$\mathbf{u}_1 + \mathbf{v}_1 = L(\mathbf{u}_0 + \mathbf{v}_0) + \tau \tag{2}$$

i.e.,

$$\mathbf{v}_1 = L\mathbf{v}_0 \tag{3}$$

We choose $\mathbf{v}_0$ to minimize the cost function J, given by:

$$J = \mathbf{v}_0^T P_0^{-1} \mathbf{v}_0 + (\mathbf{z} - H\mathbf{v}_1)^T R^{-1} (\mathbf{z} - H\mathbf{v}_1) + \lambda(\mathbf{v}_1 - L\mathbf{v}_0) \tag{4}$$

where $\mathbf{z} = \mathbf{y} - H\mathbf{u}_1$ and the row vector of Lagrange multipliers λ enforces the strong constraint:

$$\mathbf{v}_1 = L\mathbf{v}_0 \tag{5}$$

Now look at the first variation δJ of J:

$$\delta J = J(\mathbf{v}_0 + \delta_0, \mathbf{v}_1 + \delta_1) - J(\mathbf{v}_0, \mathbf{v}_1) \tag{6}$$

$$= 2\mathbf{v}_0^T P_0^{-1} \delta_0 - 2(\mathbf{z} - H\mathbf{v}_1)^T R^{-1} H\delta_1 + \lambda(\delta_1 - L\delta_0) + O(\delta^2) \tag{7}$$

$$= (2\mathbf{v}_0^T P^{-1} - \lambda L)\delta_0 + 2(\lambda - (\mathbf{z} - H\mathbf{v}_1)^T R^{-1} H)\delta_1 \tag{8}$$

The expressions in parentheses in (8) represent the gradient of the cost function in the $\mathbf{v}_0$ and $\mathbf{v}_1$ directions. We can always choose λ to make the second expression vanish by the strong constraint

(5). This value of λ can then be substituted into the first expression which may look a bit more familiar as a transpose:

$$\nabla J|_{\mathbf{v}_0} = 2P^{-1}\mathbf{v}_0 - L^\star\lambda^T \tag{9}$$

where the expression on the left hand side is the gradient of J in the $\mathbf{v}_0$ directions, $L^\star$ is the adjoint of L and λ must be transposed because of our convention of choosing λ as a row vector. So application of the adjoint operator $L^\star$ on the Lagrange multiplier λ allows us to calculate the gradient of the cost function. We can then use this gradient in a descent method to find the optimal choice of $\mathbf{v}_0$. This is the adjoint method 4DVAR.

Now let us derive a weak constraint system, i.e., one in which the model is assumed to be imperfect. Let Q be the covariance matrix of the model error. If we set $Q = 0$, we should recover the strong constraint method. For the strong constraint system, $Q = 0$ so it does no harm to write:

$$\mathbf{v}_1 - L\mathbf{v}_0 = Q\lambda^T \tag{10}$$

So

$$\lambda = (\mathbf{v}_1 - L\mathbf{v}_0)^T Q^{-1} \tag{11}$$

When the norm of Q is small, the norm of Q^{-1} must be large so it is at least plausible that we should recover the strong constraint problem smoothly in the limit as $Q \to 0$; for details, see Bennett (1992).

The weak constraint cost function then becomes:

$$J = \mathbf{v}_0^T P_0^{-1}\mathbf{v}_0 + (\mathbf{z} - H\mathbf{v}_1)^T R^{-1}(\mathbf{z} - H\mathbf{v}_1) + (\mathbf{v}_1 - L\mathbf{v}_0)^T Q^{-1}(\mathbf{v}_1 - L\mathbf{v}_0) \tag{12}$$

The rightmost expression in (12) is a positive definite bilinear form on pairs of vectors of state dimension. We may write:

$$(\mathbf{v}_1 - L\mathbf{v}_0)^T Q^{-1}(\mathbf{v}_1 - L\mathbf{v}_0) + \mathbf{v}_0^T P^{-1}\mathbf{v}_0 \equiv << \mathbf{v}_0, \mathbf{v}_1; \mathbf{v}_0, \mathbf{v}_1 >> \tag{13}$$

Following our previous procedure, the first variation of the new cost function is given by:

$$\begin{aligned}
\delta J &= J(\mathbf{v}_0 + \delta_0, \mathbf{v}_1 + \delta_1) - J(\mathbf{v}_0, \mathbf{v}_1) && (14)\\
&= 2\mathbf{v}_0^T P_0^{-1}\delta_0 - 2(\mathbf{z} - H\mathbf{v}_1)^T R^{-1} H\delta_1 + 2\lambda(\delta_1 - L\delta_0) + O(\delta^2) && (15)\\
&= 2(\mathbf{v}_0^T P^{-1} - \lambda L)\delta_0 + 2(\lambda - (\mathbf{z} - H\mathbf{v}_1)^T R^{-1} H)\delta_1 && (16)
\end{aligned}$$

Setting the gradients equal to zero leads to:

$$\begin{aligned}
L^\star\lambda^T - P^{-1}\mathbf{v}_0 &= 0 && (17)\\
\lambda - (\mathbf{z} - H\mathbf{v}_1)^T R^{-1} H &= 0 && (18)
\end{aligned}$$

Equations (17), (18) and (10) make up the *Euler-Lagrange equations* for this variational problem. Note that these equations are coupled; this is the essential practical difference between this system and the strong constraint system, in which (5) could be used to solve explicitly for λ in (8).

We note from (16) that λ is a linear function of H. We may decouple the system by writing $\alpha = H$ and $\lambda = (\mathbf{z} - H\mathbf{v}_1)^T R^{-1}\alpha$. Now write: $\mathbf{v}_0 = PL^\star\alpha^T R^{-1}(\mathbf{z} - H\mathbf{v}_1) \equiv r_0 R^{-1}(\mathbf{z} - H\mathbf{v}_1)$ where $r_0 = PL^\star H^T$ and $r_1 = Lr_0 + QH^T = (LPL^\star + Q)H^T$ The sequence $(r_0^{(j)}, r_1^{(j)})$, where the superscript $(\cdot)^j$ denotes the j^{th} column of the matrix is the *representer* of the j^{th} observation. Now look for solutions to the variational problem of the form:

$$\mathbf{v}_0 = r_0\mathbf{b} \tag{19}$$

$$\mathbf{v}_1 = r_1\mathbf{b} = (LPL^\star + Q)H^T\mathbf{b} \tag{20}$$

by construction:

$$r_1\mathbf{b} - Lr_0\mathbf{b} = (r_1 - Lr_0)\mathbf{b} = QH^T\mathbf{b} \tag{21}$$

So according to (10), $\mathbf{b}$ must satisfy:

$$H^T\mathbf{b} = H^T R^{-1}(\mathbf{z} - H(LPL^\star + Q)H^T\mathbf{b}) \tag{22}$$

hence:

$$\left[H(LPL^\star + Q)H^T + R\right]\mathbf{b} = \mathbf{z} \tag{23}$$

and:

$$\mathbf{v}_1 = (LPL^\star + Q)H^T\left[H(LPL^\star + Q)H^T + R\right]^{-1}\mathbf{z} \tag{24}$$

Recall that $\mathbf{v}_1$ is the correction to the first guess $\mathbf{u}_1$, and $\mathbf{z} = y - H\mathbf{u}_1$ so the optimal solution can be written:

$$\hat{\mathbf{u}}_0 = \mathbf{u}_0 + r_0\mathbf{b} \tag{25}$$

$$= \mathbf{u}_0 + PL^\star H^T\left[H(LPL^\star + Q)H^T + R\right]^{-1}(\mathbf{y} - H\mathbf{u}_1) \tag{26}$$

and:

$$\hat{\mathbf{u}}_1 = \mathbf{u}_1 + (LPL^\star + Q)H^T\left[H(LPL^\star + Q)H^T + R\right]^{-1}(\mathbf{y} - H\mathbf{u}_1) \tag{27}$$

Equation (27) is the definition of the Kalman filter. The matrix:

$$K = (LPL^\star + Q)H^T\left[H(LPL^\star + Q)H^T + R\right]^{-1}$$

is often referred to as the *Kalman gain matrix.*

We can now explain why the sequence (r_0, r_1) is referred to as the representer. Consider an arbitrary pair of vectors $(\mathbf{x}_0, \mathbf{x}_1)$ of dimension N. We have:

$$<< r_0^T, r_1^T; \mathbf{x}_0, \mathbf{x}_1 >> = r_0^T P^{-1}\mathbf{x}_0 + (r_1 - Lr_0)^T Q^{-1}(\mathbf{x}_1 - L\mathbf{x}_0) \tag{28}$$

$$= HL\mathbf{x}_0 + H(\mathbf{x}_1 - L\mathbf{x}_0) \tag{29}$$

$$= H\mathbf{x}_1 \tag{30}$$

where, with some abuse of notation, the bilinear form on the left hand side of (28) is calculated for individual columns of r_0^T and r_1^T. Any linear function f of a vector $\mathbf{v}$ can be represented as a dot product with some fixed vector $\mathbf{u}$, so that $f(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$. We might then call $\mathbf{u}$ the representer of f. In this way, the measurement model, i.e., the linear function which relates an arbitrary sequence of state vectors $\mathbf{x}_0, \mathbf{x}_1$ to the measurement of state $\mathbf{x}_1$ at time t_1 can be represented by the bilinear form $<< r_0, r_1; \mathbf{x}_0, \mathbf{x}_1 >>$.

If we wish to continue this process, the updated vector $\hat{\mathbf{u}}_1$ must become our first guess for the next step. We must therefore find an expression for its error covariance, corresponding to P_0.

We assume that the errors in our prediction model (1) are characterized by white noise, i.e., the system in nature is described by a Langevin equation:

$$\mathbf{u}_1^t = L\mathbf{u}_0^t + \tau + \mathbf{b}_1 \tag{31}$$

where the superscript $(\cdot)^t$ denotes the true system and $\mathbf{b}$ is a random white sequence with $E(\mathbf{b}_j \mathbf{b}_k^T) = Q\delta_{jk}$ where $E(\cdot)$ denotes expected value. We began with the assumption that P_0 was the error covariance matrix of the initial guess $\mathbf{u}_0$. We then have the expression for the forecast covariance P_1^f:

$$\begin{aligned} P_1^f &\equiv E((\mathbf{u}_1^t - \mathbf{u}_1)(\mathbf{u}_1^t - \mathbf{u}_1)^T) && (32)\\ &= E((L(\mathbf{u}_0^t - \mathbf{u}_0) + \mathbf{b}_1)(L(\mathbf{u}_0^t - \mathbf{u}_0) + \mathbf{b}_1)^T) && (33)\\ &= LPL^\star + Q && (34) \end{aligned}$$

So we may write the gain matrix K in terms of the forecast error covariance P_1^f:

$$K = P_1^f H^T (H P_1^f H^T + R)^{-1} \tag{35}$$

The error covariance P_1^a, where the superscript $(\cdot)^a$ denotes "analysis," i.e., the state estimate containing current information, of the updated state estimate is given by:

$$\begin{aligned} P_1^a &\equiv E((\mathbf{u}_1^t - \hat{\mathbf{u}}_1)(\mathbf{u}_1^t - \hat{\mathbf{u}}_1)^T) && (36)\\ &= (I - KH)P_1^f (I - KH)^T + KRK^T && (37)\\ &= (I - KH)P_1^f - P_1^f H^T K^T + KHP_1^f H^T K^T + KRK^T && (38)\\ &= (I - KH)P_1^f && (39) \end{aligned}$$

We may now repeat the cycle, defining a new cost function with the same form as (12), but with P_1^a in place of P_0 and $\hat{\mathbf{u}}_1$ in place of $\mathbf{u}_0$. It is worth noting here that it is often advantageous to use the form (38) rather than the simpler (39) for stability reasons, since it is more likely to remain positive definite in the course of large calculations which will inevitably be contaminated by roundoff errors. One intuitive explanation for this is that in the case of localized accurate observations in a large-scale calculation, some of the estimated error variances, i.e., the diagonal elements of P_1^a may become very much smaller than others. This means that in the course of computation of (39), very small numbers may be obtained as differences of numbers orders of magnitude larger. This

is the a common recipe for computational difficulty due to roundoff error. A detailed treatment of computational instabilities one may encounter in application of the Kalman filter, along with suggested remedies may be found in Bierman (1977).

Because we are interested in prediction, the correction to the first guess $\mathbf{u}_0$ at time t_0 according to (26) is not computed. We may summarize the Kalman filter in the following equations: State prediction:

$$\mathbf{u}_1 = L\mathbf{u}_0 + \tau \tag{40}$$

Covariance evolution:

$$P_1^f = LP_0L^T + Q \tag{41}$$

State update:

$$\hat{\mathbf{u}}_1 = \mathbf{u}_1 + K(\mathbf{y} - H\mathbf{u}_1) \tag{42}$$

where the *Kalman gain* K is given by:

$$K = P_1^f H^T (HP_1^f H^T + R)^{-1} \tag{43}$$

Covariance update:

$$P_1^a = (I - KH)P_1^f \tag{44}$$

So the Kalman filter consists of explicitly computing the solution to a weak constraint variational problem for the state at the end of the interval. Previous states are left unchanged; this is the definition of a filter. The solution is therefore optimal only at the final time. A new cost function is defined for the next assimilation cycle, based on the updated error covariance P_1^a. Only current observations are used directly in the calculation of the updated state. Previous observations impact the current state through the filtered value and the covariance matrix.

2. PRACTICAL CONSIDERATIONS

How do you actually do this? In all cases of interest, the bulk of the computational resources are spent on the covariance update:

$$P_1^f = LP_0L^\star + Q \tag{45}$$

$$= L(LP_0)^\star + Q \tag{46}$$

In the straightforward implementation of the Kalman filter, this is done by applying the model to the analysis covariance column by column, finally replacing P_0 by LP_0. The evolution operator L is then applied to the result row by row, and finally Q is added to the result. This means that for a system with state dimension N, $2N$ prediction steps are necessary for a single step of the covariance evolution. Typical models of the ocean and atmosphere can have $N \approx 10^5$, which makes direct

application of the Kalman filter a daunting task. Practical computations of the Kalman filter have been performed with systems with thousands of state variables (see, e.g., Miller et al., 1995). For larger systems, given state of the art computing resources, some compromises must be made.

An example of the application of the Kalman filter to a problem in ocean modeling was presented by Miller et al., 1995. In that work, the Kalman filter was applied to a linear shallow water model of the tropical Pacific. Details of the model itself are given by Cane and Patton (1984). Observations consisted of tide gauge data taken at island stations and temperature profiles taken from volunteer observing ships. These latter platforms are merchant ships which carry instruments and whose crews are trained to deploy them. Here we focus on the tide gauge data.

The model domain was the tropical Pacific basin, shown here in figure 1.

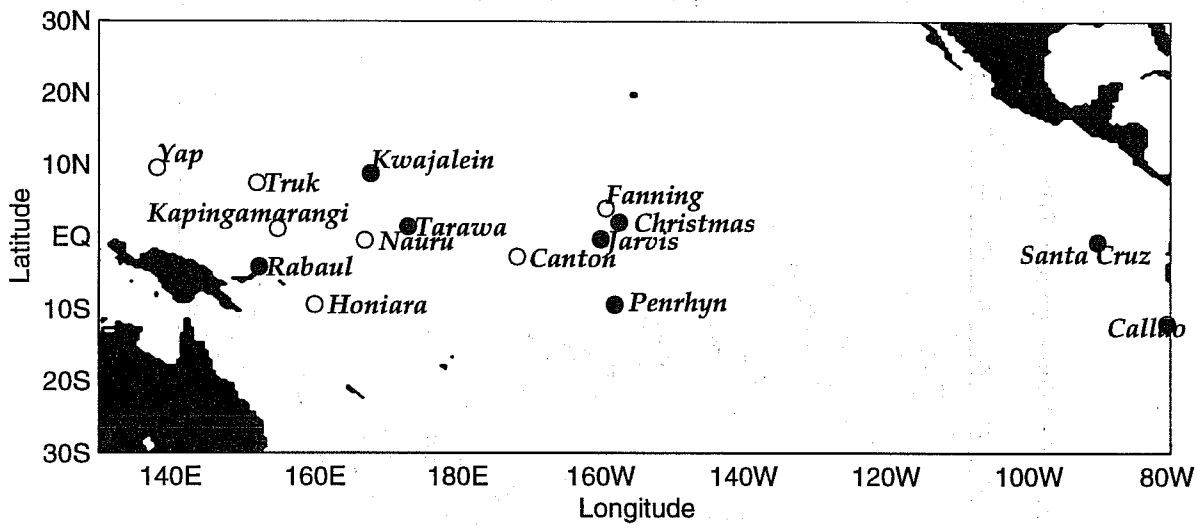

Figure 1: Model domain for Pacific Kalman filter experiments. Filled circles denote locations from which data were assimilated. Open circles denote locations from which data were available, but were held back from the assimilation process for the purpose of verification.

The model was further simplified by a long-wave approximation, so the state vector at each point on the grid consisted of the zonal advection velocity and the mass anomaly for each of two baroclinic modes. The meridional velocity is calculated diagnostically, consistent with the long-wave approximation. This system had approximately 4000 state variables.

The tropical ocean is an excellent test bed for advanced data assimilation methods. Linear models describe the large scale low frequency behavior of the tropical ocean adequately, so methods such as the Kalman filter can be applied directly without further approximation, and data are sparse, so it is necessary to make the best use of all the data available.

Construction of the system noise covariance Q proceeded from the assumption that the errors were dominated by errors in the wind forcing. Q was estimated by a large Monte-Carlo experiment, in which, for each trial, a random wind field with given covariance characteristics was generated and used to force the model for a single time step. For each of these trials, the model was initialized with the state vector set to zero. The covariances among the state variables were calculated for each

trial, and the results averaged afterwards.

Random forcing with arbitrary admissible covariance is easily generated. Assume the wind errors have covariance matrix C. Write the Cholesky decomposition of C: $C = BB^T$ where B is a lower triangular matrix. If a random vector $\mathbf{q}$ is generated with the dimension of the forcing τ and covariance given by $E(qq^T) = I$, where I is the identity matrix of appropriate dimension, i.e., the elements of q have unit variance and are uncorrelated, then a simple calculation will show that the vector Bq has covariance C. For details, see Miller et al. (1995).

In a dissipative system such as this one, if the observing array, described by the matrix H remains constant throughout the run, the error covariance reaches a steady equilibrium value, so it is possible to examine the observations and the model output and compare the actual variance of the data misfit with the prior estimate. Figure 2 shows the result of assimilation of data at Penrhyn:

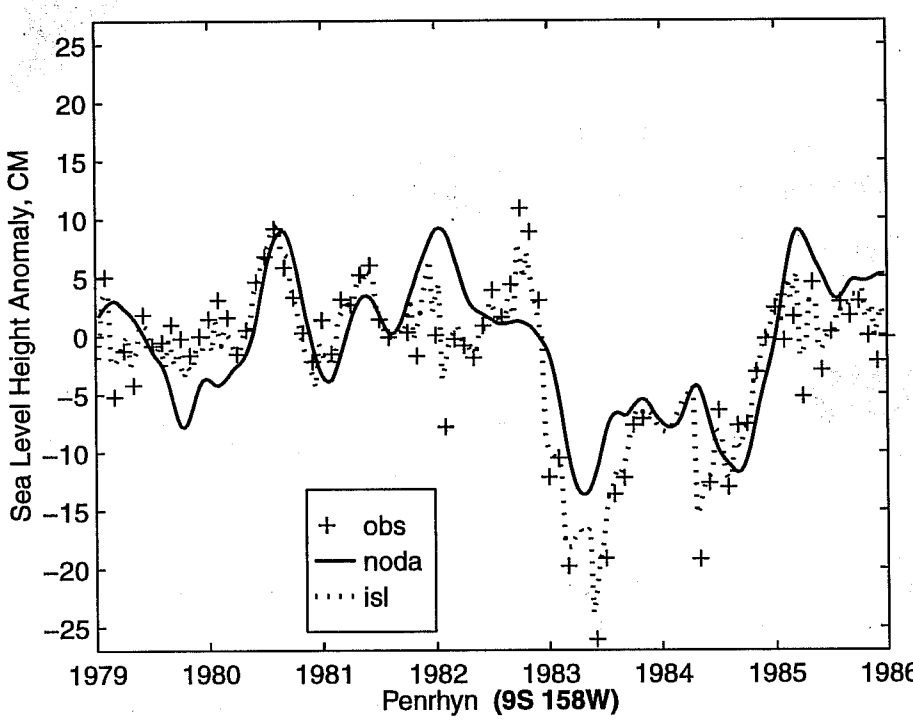

Figure 2: Results of Kalman filter experiment with shallow water model of the tropical Pacific. Comparison of raw model output, filter output and observations for Penrhyn Island. Data from the Penrhyn tide gauge participated in the assimilation

It is, of course, to be expected that the output of the filter is close to the observations; in fact, the agreement can be made arbitrarily good by the simple expedient of decreasing R or increasing Q; this is clearly not the intent of such an exercise. Results at Penrhyn are, in fact, quite good. The variance of the time series of data misfits was $34cm^2$, while the prior estimate was $38cm^2$. After assimilation, the variance of the data misfit was $2.0cm^2$, while the prior estimate was $2.8cm^2$. One can take some comfort from the fact that these numbers seem close, but closer investigation is required to evaluate how close those numbers should be.

One way to assess success or failure of a data assimilation experiment is to hold back data, and

see if the assimilation improves the estimate of those observations which the system didn't know about. Figure 3 shows the results for sea level anomalies at Truk, which did not participate in the assimilation.

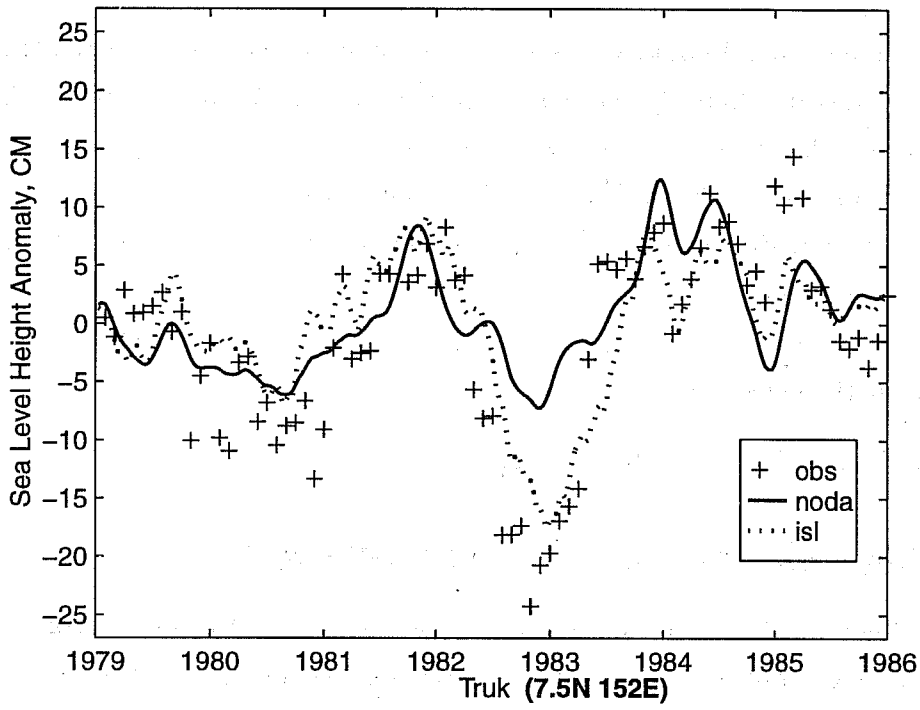

Figure 3: Results of Kalman filter experiment with shallow water model of the tropical Pacific. Comparison of raw model output, filter output and observations for Truk Island. Data from the Truk tide gauge did not participate in the assimilation

The assimilation clearly improves the analysis at Truk, and the prior estimates are reasonably reliable. The variance of the time series of data misfits in the run without assimilation was $41cm^2$ as was the prior estimate. The variance of the time series of data misfits following assimilation of data at the tide gauges indicated in figure 1 was $29cm^2$. The prior estimate was $26cm^2$. Not all of the comparisons of the model output with data were this reliable, and some were very far from the prior estimates. Comparisons such as these are useful in that they provide guidance, but without more detailed statistical analysis, we have no quantitative measure of how good our prior estimates are.

3. POSTERIOR VALIDATION OF KALMAN FILTER RESULTS

Having performed a Kalman filter experiment, it is necessary to decide whether it worked or not. Sometimes, as in the case of the tropical Pacific described in the previous section, one can withhold data and examine whether the results of assimilation matches the withheld data. Statistics such as RMS differences between the observations and the assimilation results can be computed, and compared to their prior estimates based on the estimated error covariance matrices. Sometimes these numbers match reasonably well, but it is difficult to determine how well one expects them to

match. Miller (1990) attempted to apply bootstrap methods to answer this question, but practical Kalman filters are suboptimal, and the time series of residuals are autocorrelated. The validity of bootstrap estimates of the variation of the variance of the data misfit from sample to sample is therefore questionable.

In fact, there are several objective tests one can apply to the output of the Kalman filter to verify proper operation. Probably the most important time series to examine is the sequence of differences between the observations and their forecast values, i.e., the sequence $\mathbf{y}_k - H\mathbf{u}_k^f$, where the subscript $(\cdot)_k$ denotes the k^{th} time step, and the superscript $(\cdot)^f$ denoted the forecast value, i.e., the estimate of the quantity at time t_k without assimilation of data at time t_k. This series is known as the *innovation sequence.*

A simple calculation shows that the covariance of the innovation sequence is given by $HP_k^fH^T+R$. If we now look at the sequence:

$$(\mathbf{y}_k - H\mathbf{u}_k^f)^T(HP_k^fH^T + R)^{-1}(\mathbf{y}_k - H\mathbf{u}_k^f) = \tag{47}$$

$$\sum_{i,j}(H(u_k^t - u_k^f) + b^o)_i(H(u_k^t - u_k^f) + b^o)_j(HP_k^fH^T + R)_{ij}^{-1}$$

Without loss of generality, we can consider the matrix $(HP_k^fH^T + R)$ to be diagonal; we know that this must be so in a coordinate system related to the original one by a simple orthogonal change of coordinates, and the quadratic form in (47) is invariant under orthogonal change of coordinates. In these coordinates, the components of the innovation are exactly the principal components, whose variances are the diagonal elements of $(HP_k^fH^T + R)$. A straightforward algebraic calculation then shows that:

$$(\mathbf{y}_k - H\mathbf{u}_k^f)^T(HP_k^fH^T + R)^{-1}(\mathbf{y}_k - H\mathbf{u}_k^f) = \sum_i(\mathbf{y}_k - H\mathbf{u}_k^f)_i^2/\lambda_i^2 \tag{48}$$

where λ_i^2 is the i^{th} diagonal element of $(HP_k^fH^T + R)$. This is clearly a χ^2 variable on $N-1$ degrees of freedom, so the validity of the error estimates from the Kalman filter are subject to classical hypothesis testing. In the engineering community, this χ^2 test is known as a "sanity check," (J. Thorp, personal communication).

The autocovariance properties of the innovation sequence can also be used to evaluate the performance of a Kalman filter. Kailath (1968) showed in general that the innovation sequence is white in a Kalman filter which is functioning optimally. The intuition for this comes from the fact that the innovation sequence is the connection between the Kalman filter and the real world. A white sequence contains no information. If the innovation sequence is not white, then there is real world information remaining in the innovation sequence which is not accounted for. An optimal filter would extract all of the available information, leaving behind only white noise. Daley (1992) described direct use of the statistics of the innovation sequence as a performance diagnostic for the Kalman filter.

If an implementation of the Kalman filter is found to have a consistent autoregressive structure in the innovation sequence, it is possible to modify the Kalman filter to include an explicit autoregressive model in the filter itself. This was done with a model of the tropical Pacific by Chan et al. (1996). They incorporated a first order autoregressive model of the innovations with a one-month lag into the Kalman filter of Miller and Cane (1989), and the result was a new filter which gave rise to a white innovation sequence. The origin of this one month autoregressive structure could be the observation error or the wind forcing error, or it could be a consequence of some simplification made in the derivation of the physical model. Whatever the most important defect of this data assimilation system turns out to be, it has the signature of a one-month autoregressive process.

4. DISCUSSION AND SUMMARY

The Kalman filter was derived as the solution to a weak-constraint variational problem in order to illustrate its explicit relation to the strong and weak constraint variational methods. The Kalman filter is usually presented as a discrete time algorithm in the oceanic and atmospheric modeling communities, even though a continuous time Kalman filter is easily derived in a similar fashion; see Bennett (1992). Variational data assimilation methods are often presented in continuous time and/or space form, in order to take advantage of the tools of the calculus of variations. The fact that the variational methods are often described in the setting of continuous time and/or space, and the Kalman filter is almost always described in discrete time can hinder understanding of the close relation between the commonly described methods.

The most common derivation of the Kalman filter is as a simple least squares method. This is the approach taken in Gelb (1974), and in the work of Ghil et al. (1981) in which the Kalman filter was first set in context of numerical weather prediction. All of these methods are formally free of assumptions about the underlying distributions – one can calculate minimum variance estimators for any reasonably well behaved distribution. Details regarding the underlying distributions are necessary if one wishes to assign confidence limits based on the prior estimates of error covariances.

A variety of strategies for reducing the extreme computational burden associated with the Kalman filter have been introduced recently. Fukumori (1995), Fukumori and Malanotte-Rizzoli (1995) and Hackert et al. (1996) have adopted the strategy of computing the covariance evolution with lower resolution than the underlying model. This reduces not only the number of actual arithmetic operations required, but also reduces the memory requirements to manageable size. All of these efforts were at least moderately successful.

A similar approach was adopted by Cane et al. (1996), who devised a reduced dynamical model to calculate the covariance evolution in a system very similar to that investigated by Miller et al. (1995) with the full Kalman filter and a low resolution model. The prediction model used by Cane et al. (1996) had higher resolution than that used by Miller et al. (1995), but the covariance evolution was performed on a subspace of the space of empirical orthogonal functions (EOF's) of a long time series of state vectors generated by a long model run. Less than 100 EOF's accounted for more than

95% of the variance in the original model run. The evolution operator which transformed the EOF amplitudes at one time into those at the next time step was determined empirically. Their results were essentially identical to those of Miller et al. (1995), at a much lower cost in computational resources. A survey of reductions of the Kalman filter was given by Cohn and Todling (1996).

Generalizations of the Kalman filter to nonlinear problems have attracted a fair amount of attention of late. The simplest generalization is the *extended* Kalman filter (see, e.g., Gelb, 1974), in which the linearized prediction operator at the current state vector is used to calculate the covariance evolution (34). This has been shown to be unsatisfactory in some cases. Examples in which the covariance evolution was modified by using ensemble calculations to account explicitly for the nonlinearity can be found in Evensen (1994) and Miller et al. (1994).

4. REFERENCES

Bennett, A. F., 1992: "*Inverse Methods in Physical Oceanography*," Cambridge University Press, New York, 346pp.

Bierman, G. J., 1977: "*Factorization Methods for Discrete Sequential Estimation*," Academic Press, New York. 241pp.

Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert and A. J. Busalacchi, 1996: Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter, *J. Geophys. Res., in press.*

Cane, M. A. and R. J. Patton, 1984: A numerical model for low-frequency equatorial dynamics. *J. Phys. Oceanogr.*, **14**, 1853-1863.

Chan, N. H., J. B. Kadane, R. N. Miller and W. Palma, 1996: Estimation of tropical sea level anomaly by an improved Kalman filter. *J. Phys. Oceanogr.*, **26**, 1286-1303.

Cohn, S. E. and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics, *J. Meteor. Soc. Japan*, *74*, 63-75.

Daley, R., 1992: The lagged innovation covariance: A performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 178-196.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*. **99**, 10143-10162.

Fukumori, I., 1995: Assimilation of TOPEX sea level measurements with a reduced-gravity shallow water model of the tropical Pacific ocean, *J. Geophys. Res.*, *100*, 25027-25039.

Fukumori, I. and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation; an example with an idealized Gulf Stream model. *J. Geophys. Res.*, **100**, 6777-6793.

Gelb, A., 1974: "*Applied Optimal Estimation*", MIT Press, Cambridge, MA. 374pp.

Ghil, M., S. E. Cohn, J. Tavantzis, K. Bube and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. in *Dynamical Meteorology: Data Assimilation Methods,* edited by L. Bengtsson, M. Ghil and E Kallen, Springer-Verlag, NY, pp139-224.

Hackert, E. C., R. N. Miller and A. J. Busalacchi, 1996: An optimized design for a moored instrument array in the tropical Atlantic ocean. *J. Geophys. Res.*, subjudice.

Kailath, T., 1968: An innovations control approach to least square estimation – Part I: Linear filtering in additive white noise. *IEEE Trans. Automat. Control*, **13**, 646-655.

Miller, R. N., 1990: Tropical data assimilation experiments with simulated data: The impact of the tropical ocean and global atmosphere thermal array for the ocean. *J. Geophys. Res.*, **95**, 11461-11482.

Miller, R. N., A. J. Busalacchi and E. C. Hackert, 1995: Sea surface topography fields of the tropical Pacific from data assimilation. *J. Geophys. Res.*, **100**, 13389-13425.

Miller, R. N., M. Ghil and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037-1056.

Miller, R. N. and M. A. Cane, 1989: A Kalman filter analysis of sea level height in the tropical Pacific. *J. Phys. Oceanogr.*, **19**, 773-790.