TECHNICAL MEMORANDUM

447

On the Equivalence Between Kalman Smoothing and Weak-Constraint Four-dimensional Variational Data Assimilation

M. Fisher

Research Department

Submitted to Q.J. Roy. Meteor. Soc.

December 17, 2004

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

Abstract

The fixed-interval Kalman smoother produces optimal estimates of the state of a system over a time interval, given observations over the interval, together with a prior estimate of the state and its error covariance at the beginning of the interval. At the end of the interval, the Kalman smoother estimate is identical to that produced by a Kalman filter, given the same observations and the same initial state and covariance matrix.

The dynamics of error propagation and the model error term of the covariance evolution equation act to reduce the dependence of the estimate on observations and prior states that are well separated in time. In particular, if the assimilation interval is sufficiently long, the estimate at the end of the interval is effectively independent of the state and covariance matrix specified at the beginning of the interval. In this case, the Kalman smoother provides estimates at the end of the interval that are identical to those of a Kalman filter that has been running indefinitely.

For a linear model, weak constraint four-dimensional variational data assimilation (4D-Var) is equivalent to a fixed-interval Kalman smoother. It follows that, if the assimilation interval is made sufficiently long, the 4D-Var analysis at the end of the assimilation interval will be identical to that produced by a Kalman filter that has been running indefinitely.

The equivalence between weak-constraint 4D-Var and a long-running Kalman filter is demonstrated for a simple analogue of the numerical weather prediction (NWP) problem. For this nonlinear system, 4D-Var analysis with a 10-day assimilation window produces analyses of the same quality as those of an extended Kalman filter. It is demonstrated that the current ECMWF operational 4D-Var system retains a memory of earlier observations and prior states over a period of between four and ten days, suggesting that weak constraint 4D-Var with an analysis interval in the range of four to ten days may provide a viable algorithm with which to implement an un-approximated Kalman filter. Whereas assimilation intervals of this length are unlikely to be computationally feasible for operational NWP in the near future, the ability to run an un-approximated Kalman filter should prove invaluable for assessing the performance of cheaper, but suboptimal alternatives.

1 Introduction

For an imperfect linear model, weak-constraint 4D-Var is equivalent to a fixed-interval Kalman smoother initialized with the same background state and covariance matrix at the beginning of the assimilation window. A proof of this equivalence is given by Ménard and Daley (1996), although as Ménard and Daley acknowledge, the result is classical and was demonstrated for the discrete case by Rauch, Tung and Striebel (1965), and for the continuous case by Bryson and Frazier (1963). Further proofs of the equivalence are provided by Li and Navon (2001).

The fixed-interval Kalman smoother produces optimal estimates of the state of a system at each time during the interval, based on observations throughout the interval. By contrast, at any given time during the interval, the Kalman filter uses observations only in the past. At the end of the analysis interval, both the fixed interval Kalman smoother and weak-constraint 4D-Var produce identical estimates to those of the Kalman filter, given the same initial state and covariance matrix.

For an imperfect model, the covariance matrix of model error acts in the Kalman smoother as a “forgetting factor”. It reduces the influence of past and future observations on the estimate, in recognition of the fact that such information cannot be propagated accurately by an imperfect model (Ménard and Daley, 1996). Furthermore, observational error typically contains a component that projects onto the model’s growing structures, limiting the usefulness of observations in the distant past in determining the current state.

If the assimilation window is sufficiently long that the analyzed state at a given time during the analysis interval is effectively independent of the initial conditions, then the analysis will be identical to that provided by a

Kalman smoother whose analysis interval stretches into the distant past. Consequently, at the end of the interval, the analysis will be identical to that of a Kalman filter that has been running indefinitely.

Since weak-constraint 4D-Var is equivalent to a Kalman smoother, it follows that 4D-Var gives an estimate at the end of the window that, for a sufficiently long assimilation window and a linear model, is identical to that of a Kalman filter that has been running indefinitely. That is, we may regard 4D-Var with a suitably long assimilation window as an algorithm for solving the Kalman filter equations.

For low-dimensional systems, the Kalman filter is a computationally efficient algorithm. The algorithm is non-iterative, and optimal estimates for all times during the assimilation interval are available after a single pass through the data. The Kalman smoother is similarly efficient, but requires two passes through the data.

The computational advantages of the Kalman filter, and of the Kalman smoother, are achieved by maintaining and propagating an explicit representation of the error covariance matrix for the state vector. The computational cost of handling this matrix increases rapidly with the dimension of the state vector. In particular, the number of integrations of the linear model required to propagate the covariance matrix from one time step to the next is equal to the dimension of the state vector. For large systems, the Kalman filter and the Kalman smoother become impractical unless severe approximations are made.

Weak constraint 4D-Var is an iterative algorithm. The computational cost of the algorithm results primarily from repeated integrations of the model and its adjoint, and from applications of the forward and adjoint observation operators. The number of integrations and applications performed depends on the number of iterations required to minimize the analysis cost function. This number is largely determined by the condition number of the Hessian matrix of the analysis cost function, and is fairly independent of the dimension of the control vector. Thus, the computational cost of 4D-Var increases far less rapidly with the dimension of the state vector than does the cost of the Kalman smoother. For state-vector dimensions typical of current operational NWP systems, 4D-Var is by far the more efficient algorithm.

Strict equivalence between 4D-Var and the Kalman smoother holds only for the case of a linear model. For nonlinear models, it is necessary in the Kalman smoother to linearize the model about some state. Various options exist, some of which are discussed by Jazwinski (1970). In principle, linearization is not required in order to minimize the 4D-Var cost function. However, an incremental formulation (Courtier *et al.*, 1994) is commonly used, and this performs repeated linearizations about successively better estimates of the optimal states for the analysis interval. It corresponds to a global iteration method for the nonlinear smoothing problem, as defined by Jazwinski (1970). Below, we compare weak constraint 4D-Var and the extended Kalman filter applied to a nonlinear model. At least for this case, both methods produce analyses of similar quality.

For a linear model and linear observation operators, the 4D-Var cost function is quadratic. A single, global minimum is guaranteed. For a nonlinear model or nonlinear observation operators, the cost function may have multiple minima. In this case, the global minimum may be difficult to locate. We note that Swanson and Vautard (1999) found no difficulties in locating the global minimum, for analysis intervals up to five days long, in strong-constraint 4D-Var analyses with a three-level quasi-geostrophic model. They found that the method of Pirés *et al.* (1996) allowed them to extend the analysis interval to ten days. We describe in section 2.3 an alternative to the method of Pirés *et al.* (1996), and demonstrate it for a simple nonlinear system.

The outline of this paper is as follows. In section 2, we introduce a simple analogue of the NWP problem, and use it to demonstrate the limited memory of the extended Kalman filter and the convergence of weak-constraint 4D-Var to the Kalman smoother for sufficiently-long assimilation intervals. In section 3, we estimate the memory of the current ECMWF 4D-Var analysis system (Courtier *et al.*, 1994), and show that this is similar to that of the simple analogue. Finally, in section 4, we conclude with a discussion of the implications of the results presented in earlier sections.

2 Assimilation experiments with a simple nonlinear system

2.1 An Assimilation System for the Lorenz (1995) Model

The results presented in this this section were generated using an assimilation system developed for pedagogical purposes by Leutbecher (2003), and based on the model of Lorenz (1995) and Lorenz and Emanuel (1998), hereafter denoted L95. The model is a simple, but surprisingly realistic analogue of mid-latitude atmospheric dynamics. It is defined by a set of forty coupled ordinary differential equations:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F \tag{1}$$

where the domain is cyclic, so that $x_{-1} = x_{39}$, $x_0 = x_{40}$, and $x_{41} = x_1$.

All the experiments presented in this paper follow Lorenz and Emanuel (1998) in setting $F = 8$ and taking a unit time interval to represent 5 days. For these values, the model exhibits realistic phase propagation and chaotic nonlinear behaviour. There are 13 positive Lyapunov exponents, the largest of which corresponds to a doubling time of 2.1 days, broadly commensurate (given the simplicity of the L95 model) with the error doubling times diagnosed for a modern NWP model by Simmons and Hollingsworth (2002) and by Simmons *et al.* (1995). Time integration was performed using a fourth-order Runge-Kutta method.

Leutbecher (2003) developed an extended Kalman filter (EKF) for the L95 model. The equations are standard, but are given here for completeness:

$$\begin{aligned}
\mathbf{x}_k^a &= \mathbf{x}_k^b + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^b\right) && \text{(2a)}\\
\mathbf{x}_{k+1}^b &= \mathscr{M}_{t_k\to t_{k+1}}(\mathbf{x}_k^a) && \text{(2b)}\\
\mathbf{K}_k &= \mathbf{P}_k^b\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{R}_k + \mathbf{H}_k\mathbf{P}_k^b\mathbf{H}_k^{\mathrm{T}}\right)^{-1} && \text{(2c)}\\
\mathbf{P}_k^a &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_k^b && \text{(2d)}\\
\mathbf{P}_{k+1}^b &= \mathbf{M}_{t_k\to t_{k+1}}\mathbf{P}_k^a\mathbf{M}_{t_k\to t_{k+1}}^{\mathrm{T}} + \mathbf{Q}_{k+1} && \text{(2e)}
\end{aligned}$$

Here, $\mathscr{M}_{t_k\to t_{k+1}}(\mathbf{x}_k^a)$ represents the integration of the nonlinear model from time t_k to time t_{k+1} with initial state $\mathbf{x}_k^a$. The matrix $\mathbf{M}_{t_k\to t_{k+1}}$ represents the operator corresponding to the tangent linear model, integrated over the same period, and linearized about a trajectory provided by a nonlinear model integration with initial state $\mathbf{x}_k^b$.

Observations of a true state were made every 6 hours at the first 3 points in each set of 5 (i.e. the observed points were 1, 2, 3, 6, 7, 8, 11, 12, 13, etc.). Random perturbations were added to the observations, consistent with the assumed covariance matrices of observation error, $\mathbf{R}_k = \sigma_o^2\mathbf{I}$. The standard deviation of observation error, σ_o, was set to 0.15 times the climatological standard deviation of the model state, $\sigma_{\mathrm{clim}} \approx 3.64$.

"True" states were provided by integrating the L95 model, so that the nonlinear model used in equation 2b was perfect. Despite this, the EKF was found to diverge unless a small model error covariance term was included in the covariance evolution equation. For the experiments described in this paper, we set $\mathbf{Q}_{k+1} = \sigma_q^2\mathbf{I}$. Figure 1 shows the mean background error variance normalised by σ_{clim} for a number of 230-day integrations of the EKF with a range of values of $\sigma_q/\sigma_{\mathrm{clim}}$. For values smaller than about 7×10^{-4}, the filter diverges, whereas for large values of σ_q, the filter becomes inaccurate. The experiments presented in this paper used $\sigma_q = 0.005\sigma_{\mathrm{clim}}$.

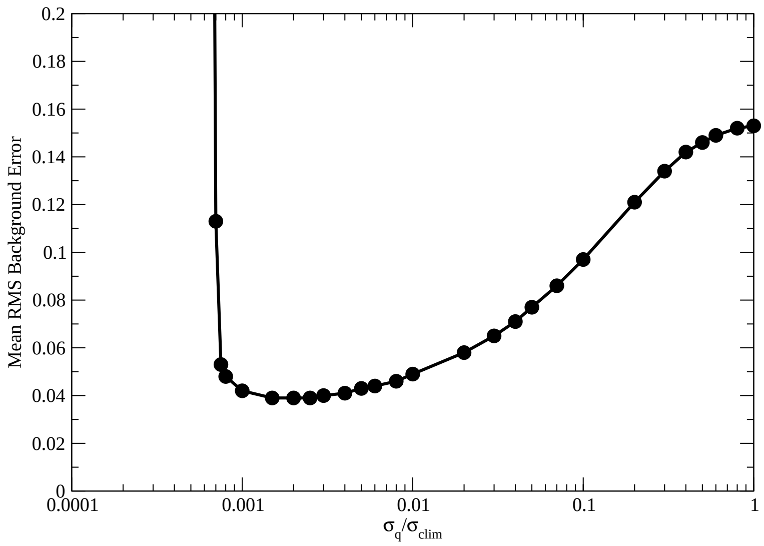

Figure 1: Mean background error standard deviation for a 230-day integration of the L95 EKF, as a function of the assumed standard deviation of model error. The units for both axes are relative to the climatological standard deviation of the model state.

2.2 The Limited Memory of the Extended Kalman Filter

To demonstrate the limited memory of the EKF, three sets of 230-day integrations were performed. For each set, “truth” was provided by the last 230 days of a 5365-day integration of the L95 model from a different initial state, and different random perturbations were added to the observations. For each integration, the covariance matrix of background error was frozen at its initial value until the integration reached some time $t = 230 - T$ (days), after which the covariance matrix was allowed to evolve.

Freezing of the covariance matrix was achieved by replacing it by a constant matrix at the start of each time step, and effectively converted the EKF into an “Optimal-Interpolation” (OI) analysis system for the first $230 - T$ days. The constant background error covariance matrix was constructed by running the OI system over a long period, calculating the mean covariance matrix of background error, and using this new covariance matrix in a new run of the OI system. This process was repeated until no further improvement in the quality of the OI analyses could be obtained.

Figure 2 shows the mean analysis error of the final analysis (at day 230) for integrations that were switched from OI to EKF at different times $t = 230 - T$. For each set, integrations started from the same background state, and used the same observations. For $T = 0$, the covariance matrix was frozen throughout. That is, the analysis system was OI. The dashed lines in figure 2 correspond to the opposite extreme of an analysis system that was run as an EKF throughout the 230 days. Despite some random variation in the quality of the final analyses, due to the small sample size, it is clear that the quality of the final analysis increases as T is increased. It is also clear that an asymptotic level of analysis error variance, equivalent to that of the EKF, is reached once T exceeds about 10 days. That is, the quality of the final analysis is independent of the background state, error covariance and observations more than 10 days earlier.

Further demonstration of the convergence of the analysis to the EKF solution is provided by examining the covariance matrix of background error estimated by the analysis system for the final (day 230) analysis. Figure 3 shows the estimated standard deviation of background error, and the corresponding correlation matrix, for three values of T, for the set of integrations corresponding to the uppermost curve in figure 2. For $T = 0$, the matrix is simply the initial OI covariance matrix, whereas for $T = 230$ days it is the EKF covariance matrix.

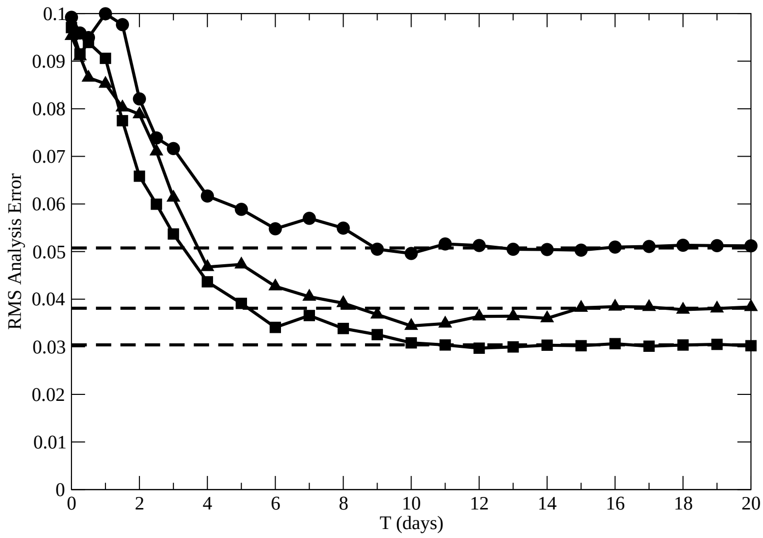

Figure 2: Mean standard deviation of analysis error (relative to $\sigma_{\rm clim}$) for the final analyses of three sets of 230-day integrations of the L95 EKF, correponding to different realizations of the "truth", and to different random perturbations of the observations. For each set, the background error covariance matrix was frozen during the period 0 *to* $(230-T)$ *days, and analysis errors are plotted as a function of* T*. Dashed lines shows the mean standard deviation of final analysis error for the EKF (i.e. for* $T = 230$ *days).*

The centre panel shows the covariance matrix for $T = 5$ days. It is already close to the EKF covariance matrix. For $T = 10$ days (not shown), it is indistinguishable from the matrix produced by the full EKF.

2.3 Weak Constraint 4D-Var

Weak-constraint 4D-Var analyses were run for the L95 system, and were compared with the EKF described above. To further demonstrate the independence of the estimate at the end of a long analysis interval on the background state and covariance specified at the beginning of the window, the usual background term was omitted from the analysis cost function.

An incremental formulation of the cost function was used (Courtier *et al.*, 1994). That is, the analysis cost function was expressed as a function of departures from a first guess, and the model was linearized about this first guess. The resulting cost function was quadratic, and could be minimized by solving the linear equation obtained by setting the gradient of the cost function to zero. This linear equation was solved using standard linear algebra methods.

The analysis cost function was:

$$\begin{aligned} J(\delta\mathbf{x}_1,\ldots,\delta\mathbf{x}_K) = {} & \frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^g + \delta\mathbf{x}_k)\right)^{\rm T}\mathbf{R}_k^{-1}\left(\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^g + \delta\mathbf{x}_k)\right) \\ & + \frac{1}{2}\sum_{k=1}^{K}\eta_{\mathbf{k}}{}^{\rm T}\mathbf{Q}_k^{-1}\eta_{\mathbf{k}} \end{aligned} \tag{3}$$

where $\eta_k = \delta\mathbf{x}_k - \mathbf{M}_{t_{k-1}\to t_k}\delta\mathbf{x}_{k-1}$.

Here, the subscript k denotes time step relative to the beginning of the analysis interval. The matrices $\mathbf{R}_k$, $\mathbf{H}_k$ and $\mathbf{Q}_k$ are the same as were used for the EKF, in equations 2c and 2e. The first guess is denoted $\mathbf{x}_k^g$. The linear

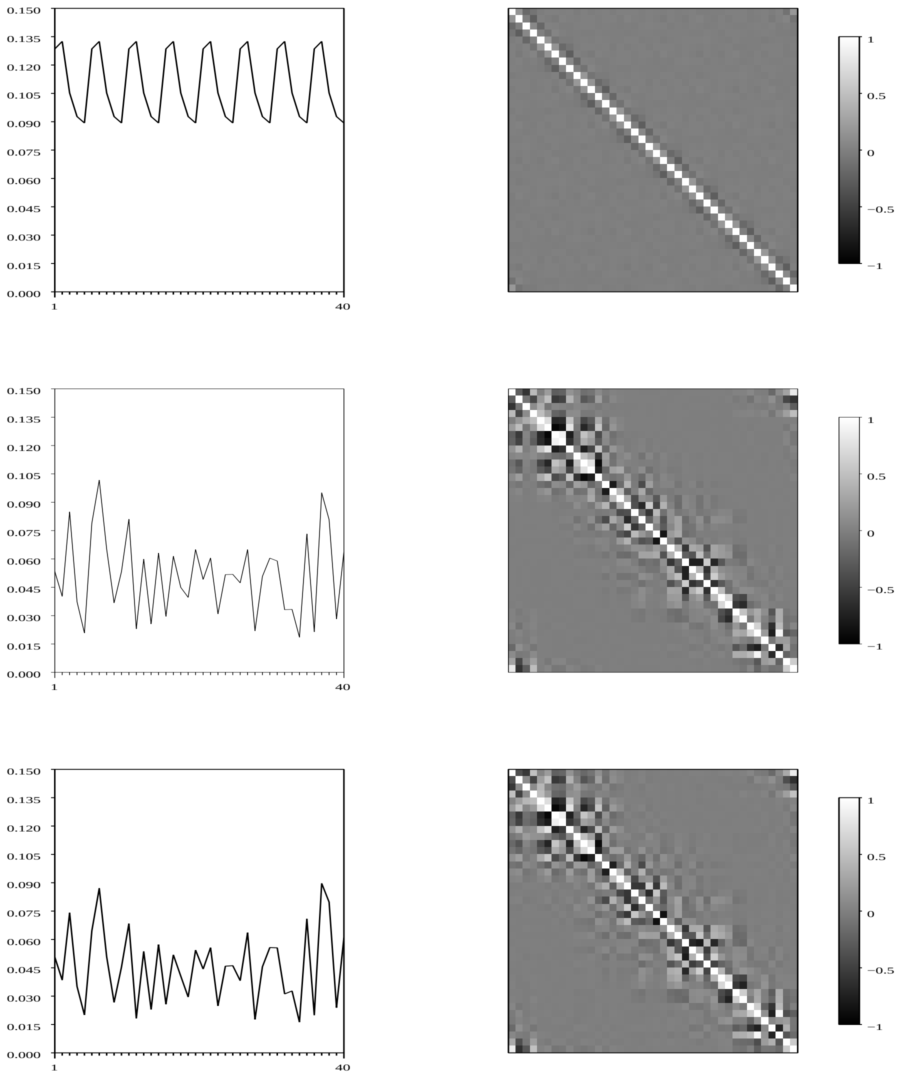

Figure 3: Background error variances and correlation matrices produced by the OI-EKF system for $T = 0$ (top), $T = 5$ days (middle), and $T = 230$ days (bottom).

model, $\mathbf{M}_{t_{k-1}\to t_k}$, was linearized about this first guess. Note that the first guess appears in equation 3 only as a linearization state. There is no constraint that the analysis should remain close to the guess.

The analysis was defined as

$$\mathbf{x}_k^a = \mathbf{x}_k^g + \delta\mathbf{x}_k^* \tag{4}$$

where $(\delta\mathbf{x}_1^* \ldots \delta\mathbf{x}_K^*)$ minimizes J.

For all the experiments presented here, an analysis was produced every six hours, regardless of the length of the analysis interval. This resulted in a sequence of overlapping analysis intervals. The first guess for any given analysis was taken from the analysed states of the overlapping portion of the preceding analysis, augmented with a six-hour integration of the nonlinear model to fill the part of the interval that did not overlap. The initial state for this integration was provided by the analysed state at the end of the preceding analysis interval.

Analyses were produced throughout the same 230-day period that the EKF was integrated. In all cases, the first analysis of the period used an analysis interval of six hours, and the length of the interval was increased by six hours for each cycle of analysis until the desired interval length was reached, after which the length of the interval was kept constant.

The lack of a background term in the 4D-Var cost function may result in under-determined analyses. That is, given increments $(\delta\mathbf{x}_1^* \ldots \delta\mathbf{x}_K^*)$ that minimize the cost function, an infinite family of minimizing solutions may be generated by adding linear combinations of the vectors that span the null-space of the Hessian matrix of the cost function. In practice, only the first analysis of the 230-day period was found to be under-determined. In this case, the solution algorithm resolved the indeterminacy by generating increments with a zero projection onto the null-space of the Hessian. Note that this solution corresponds to the limiting solution as background error variance tends to infinity for a cost function that includes a background term.

Figure 4 shows the mean RMS analysis error at the end of the analysis interval for a set of integrations with analysis intervals of different lengths. The RMS errors are indicated by dots joined by a thin line. For a selection of interval-lengths, the RMS error throughout the interval is shown by a shallow u-shaped curve. The RMS analysis errors were averaged over 100 analyses, each separated by 48 hours, and covering the last 200 days of the integrations. The corresponding mean RMS analysis errors for the EKF and for OI are indicated by horizontal dashed and dotted lines.

Unsurprisingly, for short analysis intervals the lack of a background term in the analysis cost function produces very poor analyses. For an analysis interval of one day, the RMS analysis error of 4D-Var is more than three times that of OI. It is likely that the analysis error of 4D-Var could be reduced below that of OI for all analysis intervals if a suitable background term were incorporated into the analysis cost function.

The lack of a background term becomes less detrimental as the length of the analysis interval is increased. For an interval of two days, 4D-Var gives better analyses at the end of the interval than OI. The quality of the analyses improves monotonically as the analysis interval is increased. For a 10-day interval, the 4D-Var analyses at the end of the interval are essentially as good as those produced by the EKF. In the middle of the analysis interval, the 4D-Var analyses are significantly better than those of the EKF, illustrating the benefits of smoothing over filtering for retrospective analysis.

For a nonlinear model, the 4D-Var algorithm described above is not fully equivalent to the EKF, since different linearization states are used to define the tangent linear model. In the EKF, the analysis for a given time depends on a sequence of earlier analyses, each of which used a model linearized about a background state. The corresponding 4D-Var analysis depends on a sequence of analysis states generated using a model linearized about an earlier analysis. It is likely that the linearizations used by 4D-Var are more accurate than those used by the EKF. This may explain the slight advantage of 4D-Var over the EKF, apparent in figure 4, for analysis intervals longer than 10 days.

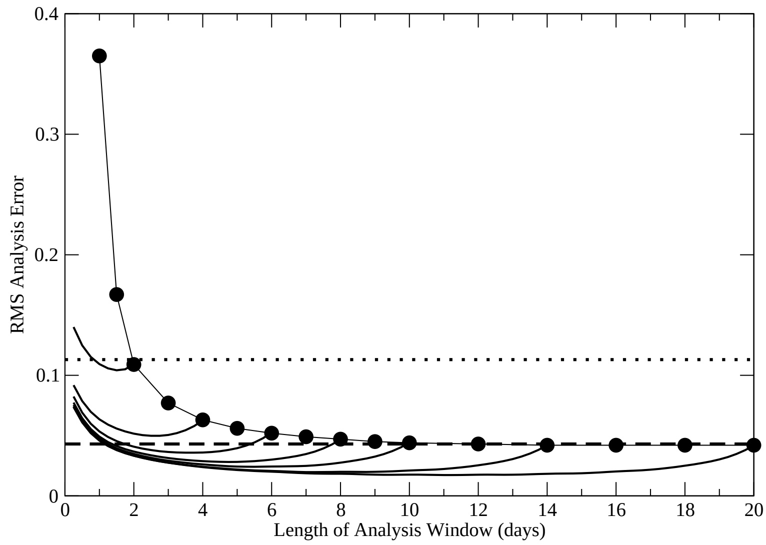

Figure 4: Mean RMS analysis error (relative to σ_{clim}) averaged over a 200-day interval for weak constraint 4D-Var analyses, as a function of the length of the analysis interval in days. RMS errors at the end of the analysis interval are shown by dots. For some interval-lengths, the RMS analysis error throughout the interval is also indicated. The dashed and dotted lines show RMS analysis errors for the EKF and OI analysis methods, respectively.

Although the 4D-Var algorithm presented above performs a single minimization for each analysis interval, it is usual to perform a sequence of minimizations. Each minimization produces a provisional analysis according to equation 4, which is used as the linearization state for the next minimization. This incremental 4D-Var algorithm may be regarded as a global iteration method for the nonlinear smoothing problem (Jazwinski, 1970).

3 The memory of the ECMWF analysis system

The simple system discussed above has a memory of approximately ten days. It is clearly of interest to determine the corresponding memory for a more realistic system. To do this, we consider two of the integrations of the ECMWF 4D-Var analysis system described by Kelly *et al.* (2004). In both cases, the analysis system was cycled for a period of one month. One integration assimilated all operationally-used observations. For the other integration, all satellite observations were withdrawn. Both integrations started with the same background state at the beginning of the month.

Figure 5 shows the evolution of the 12h and 24h root-mean-square forecast error for 500hPa geopotential in the southern hemisphere, verified against ECMWF operational analyses. It is clear that withdrawing all satellite data has a detrimental effect on forecast skill. However, withdrawing satellite data does not result in an instantaneous reduction in forecast skill, but in a steady decline over roughly seven days. That is, the analysis system retains a memory of the initial, accurate background state for approximately one week.

A further illustration of the memory of the ECMWF analysis system is given by figure 6. This shows the RMS spread of 500hPa geopotential, averaged over the globe, for an analysis ensemble. Each member of the ensemble was an independent run of the ECMWF 4D-Var analysis-forecast system over a 30-day period. The observations and sea-surface temperatures for each member were randomly perturbed. For the observations, the perturbations were drawn from a Gaussian distribution with covariance matrix equal to the prescribed covariance matrix of observation error used by the analysis. Perturbations to the sea-surface temperature were generated using the method devised by Vialard (2004). Further details of the analysis ensemble are discussed

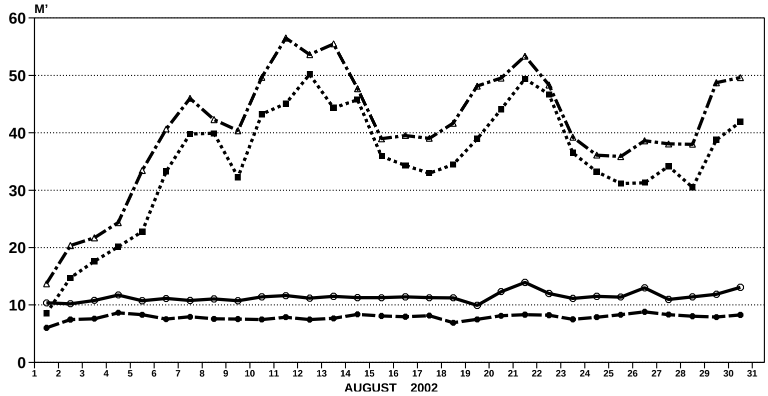

Figure 5: Evolution of RMS forecast error for 500hPa geopotential averaged over the area south of 20°S for two integrations of the ECMWF 4D-Var analysis system. The dashed and solid curves show 12h and 24h forecasts respectively from analyses that used satellite data. Dotted and dot-dashed curves show corresponding forecasts from analyses from which all satellite data were withdrawn.

by Žagar (2004), and by Fisher (2003).

The spread of the ensemble members increases over the first few days, reaching an initial plateau after about four days. After about ten days, the ensemble spread shows no clear trend, suggesting that memory of the common background state used in the first analysis has been lost.

4 Discussion

We have demonstrated that the extended Kalman filter has a limited memory. For a simple analogue of the NWP problem, the analyses produced by the EKF are essentially independent of the observations, states and covariances more than ten days earlier. We defer for future investigation the extent to which this timescale depends on the assumed statistics of model error, on the density and accuracy of observations, and on the characteristics of error growth in the model. However, we did demonstrate that a current operational 4D-Var data assimilation system has a memory for earlier observations and background states of between four and ten days.

For a linear model, the equivalence between the Kalman smoother and weak-constraint four-dimensional variational data assimilation implies that the 4D-Var analysis at the end of the analysis interval should tend towards the Kalman filter analysis as the length of the interval is increased. Although this equivalence does not hold exactly for the nonlinear case (due to differences in the linearization states for the tangent linear model), we showed that 4D-Var gave analyses of equal quality to those of the EKF for the L95 system.

Weak constraint 4D-Var with a ten-day analysis interval is likely to be twenty times more computationally expensive than a current 12-hour 4D-Var system. (Note, however, that we have not investigated how the numerical conditioning of the 4D-Var minimization problem is affected by the introduction of a weak constraint, by extending the analysis interval, or by the choice of analysis variable.) Such an analysis system may be too expensive for operational implementation on current computers. However, it is already feasible to run such analyses in an environment free from real-time constraints. These analyses are likely to prove invaluable in evaluating cheaper, suboptimal analysis methods.

In this paper, we removed the background term from the 4D-Var cost function. By restoring this term, weak-constraint 4D-Var analyses could be varied continuously between 3D-Var and the EKF by varying the length of

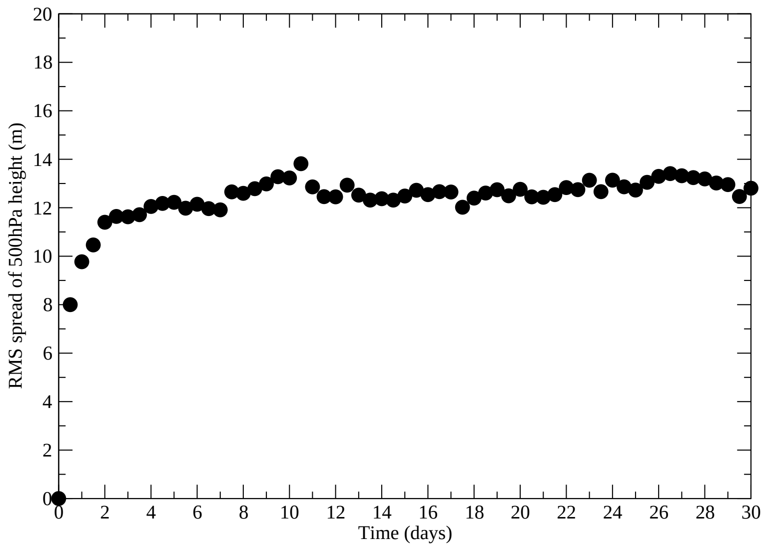

Figure 6: RMS spread of 500hPa geopotential, averaged over the globe, for an analysis ensemble. All members of the ensemble used the same background state at day 0.

the analysis window. Thus, the length of the analysis interval may be seen as a parameter that directly controls the tradeoff between computational cost and analysis quality. With a background term of good quality, it is possible that the analysis interval required to produce analyses close to those of the EKF may be considerably shorter than the four to ten days suggested above.

Weak constraint 4D-Var requires the covariance matrix of model error to be specified. This is an alarming prospect, as estimates of the statistics of model error are difficult to obtain. In this respect, it is encouraging that for the L95 system, a marked improvement over optimal interpolation was achieved using a crude, diagonal model error covariance matrix. Moreover, similar analysis quality was obtained for a wide range of values of σ_q (see figure 1). This suggests that analysis quality may be relatively insensitive to the exact specification of the model error covariance matrix.

This paper does not address a number of issues that may affect the degree to which the potential of the Kalman smoother may be realized using the variational approach. The full NWP problem contains several strongly nonlinear processes with short time scales. It remains to be seen whether minimization of the analysis cost function remains practicable in the presence of such nonlinearities if the analysis interval is extended significantly. However, it is worth noting that the 4D-Var system for the L95 model described above used a cost function with a unique, global minimum. Effectively, the assimilating model was linear, and all nonlinearities were regarded as model errors. The same approach may be possible for a full NWP system, provided that the covariance matrix of model error is carefully constructed to define different time scales for different physical processes. Further attention to the construction of the covariance matrix of model error will be needed to account for model bias and serially-correlated model error.

Finally, we note that the ability to run a Kalman smoother may have significant implications for retrospective analysis, such as the ECMWF 40-year re-analysis (Simmons and Gibson, 2000). For this purpose, an analysis defined in the centre of the interval could take advantage of observations several days in the future as well as the past. This analysis should be significantly more accurate than the corresponding NWP analysis, that used only past observations. Furthermore, the Kalman smoother analyses would have greater dynamical and temporal consistency than is possible with current techniques, and this could have significant advantages for applications such as chemical transport modelling (Stohl and Cooper, 2004) and climate studies.

5 References

Bryson A. E. and Frazier M., 1963: Smoothing for Linear and Nonlinear Dynamic Systems. *Proc. Optimum Sys. Synthesis Conf., U.S. Air Force Tech. Rept. ASD-TDR-63-119*.

Courtier P., Thépaut J-N. and Hollingsworth A., 1994: A strategy for operational implementations of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367–1387.

Fisher M., 2003: Background error covariance modelling. *Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean,* 45–64.

Jazwinski A. H., 1970: Stochastic Processes and Filtering Theory. *Academic Press*.

Kelly G., McNally A., Thépaut J.-N., Szyndel M., 2004: OSEs of all main data types in the ECMWF operation system. *in Proc. Third WMO Workshop on the Impact of Various Observing Systemson NWP.* Alpback, Austria, 9-12 March 2004..

Leutbecher M., 2003: Adaptive observations, the Hessian metric and singular vectors. *Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean,* 393–416.

Li Z. and Navon I. M., 2001: Optimality of 4*D*-Var and its relationship with the Kalman filter and Kalman smoother *Q. J. R. Meteorol. Soc.,* **127,** 661–684.

Lorenz E. N., 1995: Predictability: A problem partly solved. *Proc. Seminar on Predictability, Vol. 1, ECMWF, Reading, Berkshire, UK,* 1-18.

Lorenz E. N. and Emmanuel K. A., 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.,* **55,** 399-414.

Ménard, R. and Daley, R., 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus,* **48A,** 221–237.

Pirés C., Vautard R. and Talegrand O., 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus,* **48A,** 96-121.

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

Simmons A. J. and Hollingsworth A., 2002: Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.,* **128,** 647–677.

Simmons A. J. and Gibson J. K., 2000: The ERA-40 Project Plan. *ECMWF ERA-40 Project Report Series* No. 1

Simmons A. J., Mureau R. and Petroliagis T., 1995: Error growth and estimates of predictability from the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.,* **121** 1739–1771.

Stohl A. and Cooper O. R., 2004: A Cautionary Note on the Use of Meteorological Analysis Fields for Quantifying Atmospheric Mixing. *J. Atmos. Sci.,* **16,** 1446–1453

Swanson and Vautard, 1999: Four-Dimensional Variational Assimilation and Predictability in a Quasi-Geostrophic Model. *Tellus,* **50A,** 369–390.

Vialard J., Vitart F., Balmaseda M. A., Stockdale T. N. and Anderson D. L. T., 2004: An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. *Submitted to Mon. Wea. Rev. (accepted pending minor revisions)*.

Žagar N., Andersson E. and Fisher M., 2004: Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. *submitted to Q. J. R. Meteorol. Soc.*