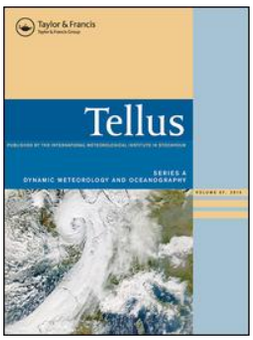

Tellus A: Dynamic Meteorology and Oceanography

ISSN: (Print) 1600-0870 (Online) Journal homepage: http://www.tandfonline.com/loi/zela20

Data assimilation for re-analyses: potential gains from full use of post-analysis-time observations

Martin Juckes & Bryan Lawrence

To cite this article: Martin Juckes & Bryan Lawrence (2006) Data assimilation for re-analyses: potential gains from full use of post-analysis-time observations, Tellus A: Dynamic Meteorology and Oceanography, 58:2, 171-178, DOI: 10.1111/j.1600-0870.2006.00167.x

To link to this article: https://doi.org/10.1111/j.1600-0870.2006.00167.x

Tellus (2006), 58A, 171–178
Printed in Singapore. All rights reserved

TELLUS

Data assimilation for re-analyses: potential gains from full use of post-analysis-time observations

By MARTIN JUCKES* and BRYAN LAWRENCE, *British Atmospheric Data Centre, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, OX11 0QX, UK*

(Manuscript received 14 March 2005; in final form 30 September 2005)

ABSTRACT

In recent years a number of operational meteorological centres have completed multidecadal reanalyses of their observation records using a version of their operational analysis systems. These operational systems aim to approximate the best possible analysis of the atmospheric state at a given time using all observations made prior to that time, and require major resources to produce. Re-analyses are made with the same real-time systems because they can be done as marginal activities on the back of operational efforts. In this paper, we examine some of the salient differences between the use of optimal real-time analyses and optimal retrospective analyses in the context of a simple linear system. In this case, the optimal real-time analysis could be obtained by the Kalman filter. When observations are available both before and after the analysis time the additional information can, in principle, be exploited to improve on the Kalman filter analysis. For linear systems the optimal retrospective analysis is given by the Kalman smoother, which combines a forward and backward Kalman filter. Results comparing these methods are presented which demonstrate the importance of using all the available data for a retrospective analysis. While using all future data is not yet tractable for retrospective meteorological analyses, such techniques are of use for more limited re-analysis.

1. Introduction

Data assimilation systems are used in earth system science for two main reasons: to provide initial conditions for forecast systems, and to synthesize available data to produce best estimates of some part of, or all of, the system state. In recent years a great effort has been put into producing reanalyses (e.g. Simmons and Gibson, 2000; Kalnay et al., 1996): retrospective analyses of recent observations using improved analysis systems. The use of a forecast analysis system in a reanalysis does not make full use of the observations (e.g. Cohn et al., 1994) because forecast systems naturally take limited account of observations after the analysis time (each cycle of the current ECMWF (The European Centre for Medium-Range Weather Forecasts) system, for instance, produces two analyses 6 h apart, with the last admitted observations coming 3 h after the later analysis[1]). In data processing terminology, the optimal approach would use a "smoother", which exploits the whole time series of observations, rather than a 'filter', which deals with an ongoing stream of observations. Cohn et al. (1994) emphasize the ability of the smoother to propagate information both upstream and downstream from an observation. They also show that while, as is well known, addition of information in a fully optimal analysis scheme will always improve the analysis, the same is not true of suboptimal schemes of the kind widely used in the context of atmospheric and oceanographic data assimilation.

Here we explicitly address the issue of the importance of post-analysis-time observations on the quality of an analysis by comparing the performance of a Kalman filter (Kalman and Bucy, 1961) and a Kalman smoother on a simple random walk. Despite its simplicity, this allows us to investigate the effects that various inaccuracies in the analysis system will have in the context of varying system stability. While we use the simple random walk system to demonstrate the importance of future data to an analysis, further work (Juckes, 2005) shows an alternative technique applied to ozone re-analysis which exploits future data in practice.

For linear systems with Gaussian noise the Kalman filter derives the optimal solution (minimising the squared error expectation) at a given time using observations up to and including that time, while the Kalman smoother gives the optimal using observations both before and after the analysis time. These approaches will be referred to as "real-time" and "retrospective" analyses below.

The Kalman filter and Kalman smoother are used here as they are effective means of finding fully optimal analyses for the low-

*Corresponding author.
e-mail: m.n.juckes@rl.ac.uk
[1]http://www.ecmwf.int/research/ifsdocs/CY28r1/Assimilation/Assimilation-15-03.html
DOI: 10.1111/j.1600-0870.2006.00167.x

order systems discussed here. The relationship between these and various suboptimal implementations of 4D-VAR used in atmospheric data assimilation are discussed in Li and Navon (2001).

The comparison is made not only for the idealised situation in which an unbiased model with random errors of known error covariance is available but also for biased models with inaccurately specified error covariances. A number of authors have addressed the issue of model bias in data assimilation (Griffith and Nichols, 1996; Dee and Da Silva, 1998; Bell et al., 2004) by seeking to estimate the bias as an additional model variable with some persistence. There does not appear to be much work, however, quantifying the impact of bias and inaccurate statistics on the analysis. This issue is addressed here in the context of a very simple model which permits analytic solution of the analysis error covariances.

2. Model system: the Langevin equation

We investigate an extremely simple system in order to show how the Kalman filter and smoother are affected by changes in relative amplitude of observational and modelling errors, and how instability and damping in the system affects the accuracy of an analysis. Our investigation will be based on the behaviour of a single linear mode described by the Langevin equation:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \alpha u + \gamma + \varsigma W, \tag{1}$$

where α, γ and ς are constants and W is a Wiener process. The standard Langevin equation has $\alpha < 0$ and $\gamma = 0$, but these restrictions are relaxed here so that results relevant to unstable modes and to models with bias can be obtained.

Appendix A shows that this equation can be integrated over a finite time interval Δt to give a random walk:

$$u(t + \Delta t) = \theta u(t) + \Gamma + \sqrt{\theta}\, e, \tag{2}$$

where $\theta = \exp[\alpha \Delta t]$, $\Gamma = \gamma \Delta t$ and e is a white noise process (Kloeden and Platen, 1992). The scaling of the noise term in eq. (2), which occurs as a result of the underlying Wiener process being integrated over the time interval Δt, could be absorbed into the definition of e. However, it is convenient to keep this scaling, as it results in greater symmetry between results obtained for forward and reverse assimilation algorithms and hence in greater simplicity for the results obtained for the Kalman smoother. As defined here, e is independent of the deterministic component of the system (i.e. θ). The variance of e, the model noise, is determined in Appendix A. For small Δt it is given by

$$\sigma_m^2 = \varsigma^2 \Delta t + O(\Delta t)^2.$$

Suppose that a realisation of the above random walk is "observed" at times $t = k\Delta t$:

$$u_{o:k} = u(k\Delta t) + e_o, \tag{3}$$

where $u_{o:k}$ is an observation taken at interval k, and e_o is the observational noise with variance σ_o^2.

θ can be varied to look at the different behaviour of unstable ($\theta > 1$), neutral ($\theta = 1$) and damped ($\theta < 1$) systems (we assume α real and hence $\theta > 0$). Equation (2) should provide insight into the behaviour of growing ($\theta > 1$) and decaying ($\theta < 1$) modes within a higher-order analysis system.

3. Perfectly characterised model

A perfectly characterised model is here understood to be a model which has random errors with known statistical properties. This is distinct from what is known in the meteorological literature as the "perfect model" situation in which the model is assumed to have no random component and no errors.

The standard forms of the Kalman filter and Kalman smoother can easily be applied to this system to obtain optimal analyses for a given set of observations. The equations and a complete description of the notation being used are given in Appendix B.

If observations are equally spaced and the model error variances do not vary in time, the analysis error will also be independent of time. The error variance estimated by the Kalman filter must then satisfy

$$\frac{1}{\hat{B}_f} = \frac{1}{\sigma_o^2} + \frac{1}{\theta^2 \hat{B}_f + \theta \sigma_m^2}, \tag{4}$$

where $\hat{B}_f$ is the analysis error variance for the Kalman filter (subscripts b and s will be used for quantities associated with the backward filter and Kalman smoother). The equation for the backward filter is obtained if θ is replaced by θ^{-1} in eq. (4). Solving eq. (4), the analogous equation for $\hat{B}_b$ and combining the results to calculate the Kalman smoother analysis error covariance gives

$$\hat{B}_f = \frac{\sigma_o^2}{2\theta}[\sqrt{\mu^2 - 4} - (\theta^{-1} - \theta + x)], \tag{5}$$

$$\hat{B}_b = \frac{\theta \sigma_o^2}{2}[\sqrt{\mu^2 - 4} - (\theta - \theta^{-1} + x)], \tag{6}$$

$$\hat{B}_s = \frac{\sigma_m^2}{\sqrt{\mu^2 - 4}}, \tag{7}$$

where

$$\mu = \theta + \theta^{-1} + x, \qquad \text{and} \qquad x = \frac{\sigma_m^2}{\sigma_o^2} \tag{8}$$

is the ratio of model error to observational error. The analysis error variances are shown in Fig. 1, as functions of $\log_2 \theta$, for $x = 0.04$. This value of x corresponds to a situation in which the model error, which accumulates between observations is small compared with the observational error. For $\theta = 1$ the forward and backward Kalman filters have identical error variances, nearly twice as large as the error variance of the Kalman smoother. The error variance for the Kalman filter grows rapidly for $\theta > 1$. In

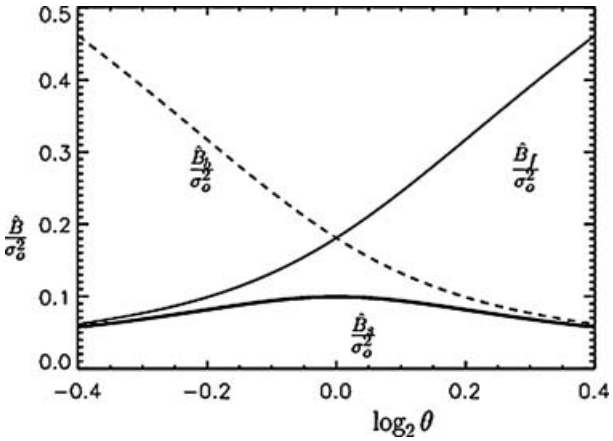

Fig. 1. Analysis error variance for a random walk plotted against $\log_2\theta$:
for the Kalman smoother (heavy solid line), Kalman filter (solid line) and backward Kalman filter (dashed). Equations (5)–(7) have been evaluated with $x = 0.04$.

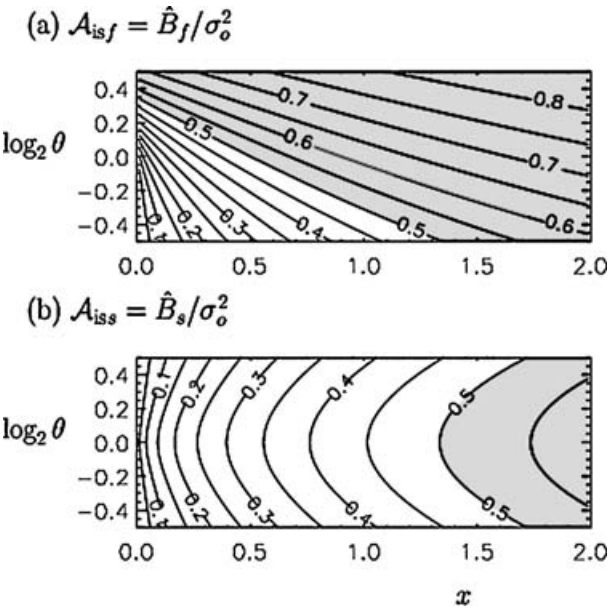

Fig. 2. Analysis information synchronicity for the random walk analysed by: (a) the Kalman filter and (b) the Kalman smoother. Contour interval 0.05, values greater than 0.5 shaded.

this situation the system is unstable, so errors in an analysis are amplified in the forecast used for the background of the next analysis. At large θ the error growth is such that the forecast has minimal weight and the analysis error approaches the observation error. The backward Kalman filter has error covariances which are identical to those of the forward Kalman filter mirrored about $\theta = 1$. The Kalman smoother combines the best of both worlds. It effectively takes the Kalman filter solution for damped modes and the backward KF solution for growing modes.

The difference in accuracy of the real-time and retrospective analyses in the presence of strong instability can be illustrated by looking at the large θ limit of eqs. (5)–(7):

$$\hat{B}_f \approx \sigma_o^2\left(1 - \frac{1}{\theta^2}\right), \tag{9a}$$

$$\hat{B}_s \approx \hat{B}_b \approx \frac{\sigma_m^2}{\theta}. \tag{9b}$$

This shows that the optimal real-time analysis only gives a slight improvement over the observations in the highly unstable situation. For the retrospective analysis, on the other hand, the error actually decays towards zero as the degree of instability increases.

For neutral modes the ratio $\hat{B}_s/\hat{B}_f$ approaches unity for large x, when the analysis is, in both instances, dominated by a single observation, and approaches 0.5 for small x, when the analysis is dominated by information carried from asynchronous observations by the model. The Kalman smoother has double the number of asynchronous observations within a given time difference from the analysis, and hence half the analysis error covariance.

This dependence on x can be expressed in terms of an "analysis information synchronicity" (or "synopticity" for meteorologists):

$$\mathcal{A}_{\text{is}} = \left|\left|\hat{B}H^T R_k^{-1} H\right|\right|,$$

where R is the observation error covariance for observations at a particular time and H the observing operator and $||\cdot||$ is a suitable norm. $\mathcal{A}_{\text{is}}$ is the ratio of the information input to the analysis from the observations ($H^T R_k^{-1} H$) to the total information content of the analysis ($\hat{B}^{-1}$). In the present example the analysis information synchronicity is just the analysis error variance normalised by the observational error variance. This is shown in Fig. 2.

For the Kalman filter there is marked increase in the information synchronicity as θ increases. For $\theta \leq 0$ the information synchronicity falls to zero as the model skill increases ($x \to 0$). For unstable modes, on the other hand, the synchronous observation retains a finite weight even for a perfect model. Thus, for $\log_2(\theta) = 0.35$ ($\theta \approx 1.27$) a perfect model will still produce an analysis in which 50% of the information comes from the synchronous observation.

For the Kalman smoother the sign of α is immaterial: unstable modes are treated with the same accuracy as damped modes. This is a direct consequence of the time-symmetry of the analysis algorithm: time reversal, which swaps unstable with damped modes, has no impact on the accuracy of the result. Instability in the system actually improves the ability of the analysis process to bring in information from asynchronous observations. The greater transfer of information from observations after the analysis time more than compensates for the reduced transfer of information from observations before the analysis time.

For instance, when $\theta = 1.2$, $x = 0.04$, the error variance of the Kalman filter is more than five times that of the Kalman smoother.

4. Imperfect noise statistics and model errors

The results presented above have assumed that we have a perfect representation of the deterministic component of the model and perfect knowledge of the model error statistics. In real-life applications, we have to deal with imperfections in the deterministic model (e.g. incorrect specification of θ) and limited information about the noise statistics (here σ_m). In this case the calculations of the previous section, which predict the accuracy of the analysis on the assumption that these input parameters are well known, may not be meaningful. Here we repeat the analysis, taking into account imperfections in the input parameters. Suppose that Γ and σ_m are not accurately known, and instead we have to use an estimate $\Gamma^\dagger$ and $\sigma_m^\dagger$:

$$u^*(t+\Delta t) = \theta u^*(t) + \Gamma^\dagger,$$

where $u^*(t)$ is now a suboptimal estimate of $u(t)$. The analysis in this simple example depends on σ_m and σ_o only through the ratio σ_m/σ_o. That is, if both σ_m and σ_o are doubled, the analysis is completely unchanged. An overestimate of σ_m is equivalent to an underestimate of σ_o.

4.1. Inaccurate model noise variance

The Kalman filter and smoother equations in Appendix B give an estimate of the analysis error which depends only on the estimated system model and error statistics. Let the error covariance calculated by the analysis algorithm using the inaccurate inputs be $\hat{B}^\dagger$. This estimate will not, generally, equal the actual error variance of the associated analysis. Let the actual error variance be $\hat{B}^* \stackrel{\text{def}}{=} E[(u_a^* - u)^2]$, where u_a is the analysis. In the present case, we know the true model equations, so we can calculate how the true error evolves. The calculation is given in detail in Appendix C. The change in the actual error covariance, $\delta\hat{B}_f \stackrel{\text{def}}{=} \hat{B}_f^* - \hat{B}_f$, can be related to the error in the estimated optimal error covariance. Applying eq. (C6)

$$\delta\hat{B}_f = \frac{\left(\hat{B}_f^\dagger - \hat{B}_f\right)^2\left(\sigma_o^2 - \hat{B}_f\right)}{\sigma_o^4\left[\left(B_f\hat{B}_f^{-1}\right)^2 - \theta^2\right]}, \tag{10}$$

where B without the hat (ˆ) is the forecast error covariance (eq. B1b, Appendix B). Subscripts and superscripts are as for $\hat{B}$, so $B_f^\dagger$ is the Kalman filter forecast error estimate. The increase in the actual error covariance varies as the square of the error in the estimated optimal error covariance. For example, using an underestimate of σ_m will result in an underestimate of the analysis error variance ($\delta B_f^\dagger < 0$). Conversely, an overestimate of σ_m will give $\delta B_f^\dagger > 0$. In both cases the analysis will be suboptimal (as any incorrectly specified parameters reduce the accuracy of the analysis).

A sensitivity which is relevant for near-optimal condition can be calculated:

$$\frac{\partial^2\hat{B}_f}{\partial\left(\sigma_m^{\dagger\,2}\right)^2} = \frac{\partial^2\hat{B}_f}{\partial\left(\hat{B}_f^\dagger\right)^2}\left(\frac{\partial\hat{B}_f^\dagger}{\partial\left(\sigma_m^{\dagger\,2}\right)}\right)^2, \tag{11}$$

where, from eq. (10)

$$\frac{\partial^2\hat{B}_f}{\partial\left(\hat{B}_f^\dagger\right)^2} = \frac{2\left(\sigma_o^2 - \hat{B}_f\right)}{\sigma_o^4\left[\left(B_f\hat{B}_f^{-1}\right)^2 - \theta^2\right]},$$

and, from eq. (7a)

$$\frac{\partial\hat{B}_f^\dagger}{\partial\left(\sigma_m^{\dagger\,2}\right)} = \frac{1}{2\theta}\left(\frac{\mu}{\sqrt{\mu^2-4}} - 1\right).$$

Putting these together and rescaling to give a non-dimensional sensitivity

$$S_f \equiv \frac{\sigma_m^4}{\hat{B}_f}\frac{\partial^2\hat{B}_f}{\partial\left(\sigma_m^{\dagger\,2}\right)^2} = \frac{x^2(\mu - \sqrt{\mu^2-4})^4}{4\theta^4(\mu^2-4)^{3/2}(\sqrt{\mu^2-4} - \theta + \theta^{-1} - x)}. \tag{12}$$

Figure 3 plots the right-hand side of eq. (12) and the corresponding results for the reverse filter and the Kalman smoother. A typical value of around 0.2 implies that a 50% error in σ_m^2 will increase the analysis error by only 5%. The sensitivity of the Kalman filter error becomes very small for $\theta > 1$ because the analysis is dominated by the observations in that case.

4.2. Model bias

In real applications, we also expect there to be errors and pragmatic approximations in the deterministic components of our predictive model.

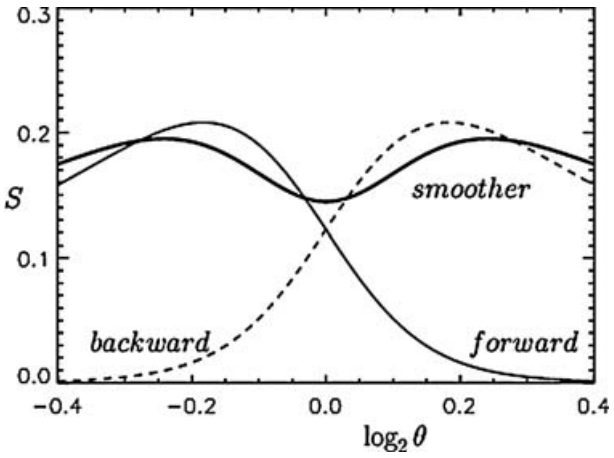

Fig. 3. As in Fig. 1, except showing sensitivity of error variance to inaccurate process model error statistics, as defined in eq. (12), evaluated at $x = 0.04$.

We can estimate the impact of such shortcomings by looking at the impact of a bias in the model: $\delta\Gamma \stackrel{\text{def}}{=} \Gamma^{\dagger} - \Gamma$. Using eq. (C8) from Appendix C, the bias in the analysis generated by an error in specification of Γ is given by

$$\hat{b}_f = \frac{1}{2\theta}\left(\sqrt{\frac{\mu+2}{\mu-2}} - 1\right)\delta\Gamma, \tag{13a}$$

$$\hat{b}_b = -\frac{\theta}{2}\left(\sqrt{\frac{\mu+2}{\mu-2}} - 1\right)\delta\Gamma, \tag{13b}$$

$$\hat{b}_s = \frac{1}{2(\mu-2)}\left(\sqrt{\frac{\mu-2}{\mu+2}} + 1\right)\left(\frac{1}{\theta} - \theta\right)\delta\Gamma. \tag{13c}$$

These functions are plotted for $x = 0.04$, $\delta\Gamma = 1$, in Fig. 4. The Kalman filter and backward Kalman filter show maximum bias amplitudes near $\theta = 0$, that is, for nearly neutral modes. The Kalman filter bias tends to zero for large θ, but this is merely a reflection of the fact that in this limit the model is given little weight and the analysis is dominated by the observation.

The bias in the backward filter has the opposite sign to that of the forward filter. The Kalman smoother bias is a weighted mean of forward and backward biases. Here it vanishes identically for neutral modes ($\theta = 0$).

The influence of bias on neutral, or near-neutral modes, becomes most severe when the random error of the model is small. For $\sigma_m \ll \sigma_o$, the bias of the Kalman filter at $\theta = 0$ approaches

$$\hat{b}_f(0) \approx \frac{\delta\Gamma\sigma_o}{\sigma_m}.$$

This shows that a decrease in model random error can lead to an increase in analysis bias. The maximum bias in the Kalman smoother in this limit occurs for slightly non-neutral modes, with $\theta \approx 1 \pm \sqrt{x}$. The corresponding bias is half that given by

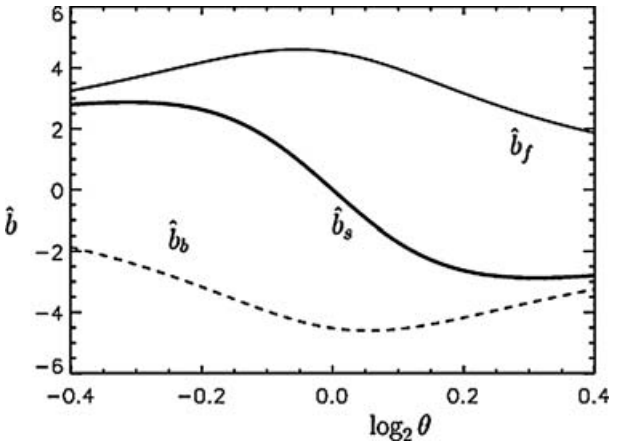

Fig. 4. As in Fig. 1, except showing bias induced by process model bias, normalised by the latter.

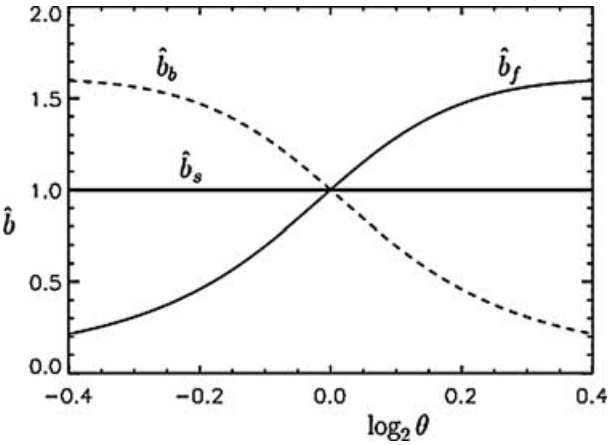

Fig. 5. As in Fig. 4, except showing bias induced by observational bias, normalised by the the latter.

the Kalman filter:

$$\max[\hat{b}_s] \approx \frac{\delta\Gamma\sigma_o}{2\sigma_m}. \tag{14}$$

The bias is more of a problem when the model is skillful (x small) because in this situation the analysis relies heavily on the model. In the small x limit the random error variance for near-neutral modes is $\sigma_m\sigma_o$. Comparison with the above expression for the bias shows that the bias dominates over the random error when

$$\delta\Gamma^2 \geq \frac{\sigma_m^3}{\sigma_o}.$$

Thus, if $\delta\Gamma$ and σ_m are both much less than σ_o and $\delta\Gamma$ is of comparable or greater magnitude than σ_m, the analysis bias will be comparable to or greater than the observational error, no matter how small $\delta\Gamma$ and σ_m are. This statement applies to analyses carried out using an accurate estimate of the random model error variance: $\sigma_m^{\dagger} = \sigma_m$. The amplitude of the bias in the analysis can be reduced by using an artificially large value of $\sigma_m^{\dagger}$, but this will increase the random component of the error. In the small $\sigma_m^{\dagger}/\sigma_o$ limit the total error variance is given by

$$E_{\text{tot}} = \frac{\delta\Gamma^2\sigma_o^2}{4\sigma_m^{\dagger 2}} + \frac{\sigma_m^2\sigma_o}{2\sigma_m^{\dagger}} + \frac{\sigma_m^{\dagger}\sigma_o}{2}, \tag{15}$$

where the first term on the right-hand side is the squared bias and the second and third are derived by solving equation (C5) at $\theta = 1$.

If $\delta\Gamma$ and σ_m are of comparable size and both much smaller than σ_o, then E_{tot} of eq. (C5) is minimised by

$$\sigma_m^{\dagger} = \delta\Gamma^{2/3}\sigma_o^{1/3}$$

and the error variance is approximately

$$E_{\text{tot}} = \frac{3}{4}(\delta\Gamma)^{2/3}\sigma_o^{4/3} \ll \sigma_o^2.$$

This shows that a useful analysis can be obtained from a biased model provided the appropriate equivalent noise is used. For a skilful model this equivalent noise could be much larger than the square of the bias amplitude. This also means that tuning $\sigma_m^\dagger$ in order to minimise E_{tot} will not necessarily lead to a reasonable estimate of σ_m.

4.3. Observational bias

The impact of a bias δy in the observations is given by

$$\hat{b}_{fk} = \frac{1}{2}\left[1 + \theta^{-1} + (1-\theta^{-1})\sqrt{\frac{\mu+2}{\mu-2}}\right]\delta y, \tag{16a}$$

$$\hat{b}_{bk} = \frac{1}{2}\left[1 + \theta + (1-\theta)\sqrt{\frac{\mu+2}{\mu-2}}\right]\delta y, \tag{16b}$$

$$\hat{b}_{sk} \equiv 1. \tag{16c}$$

For neutral modes the bias in the observations is simply reflected in the analysis without any modification. For unstable modes, however, the Kalman filter amplifies the observational bias. False information fed into the analysis is amplified by the instability of the mode, so that the analysis is degraded relative to the observations. For damped modes the bias is also damped. In this case the relaxation towards zero reduces the impact of false information in the observation.

The Kalman smoother always carries the observational bias into the analysis without modification. The identity (16c), showing that the Kalman smoother reflects the observational bias unchanged in the present system, results from cancellation among a large number of complex algebraic terms in the formualtion used here. It can be derived more directly using an alternative, variational, formulation of the problem.

5. Discussion

A simple system has been used to illustrate the difference between real-time and retrospective analyses. The simple system is capable of representing damped, neutral and growing modes. The greatest difference between the real-time and retrospective analyses is, not surprisingly, found when the system represents unstable modes. In this case the retrospective analysis error variance was reduced to zero as the model error variance was reduced to zero (i.e. as the quality of the assimilation model improved). In the real-time analysis, on the other hand, there remains a finite analysis error even when a perfect model is used.

In the case of a stable system with neutral modes, sparse observations lead to little difference between the real-time and retrospective analyses, but as the data input increases, the analysis at any given time is dominated by information carried from other times by the model: for an analysis time in the middle of the available data, the retrospective analysis error variance tends towards half that of the real-time analysis.

An examination of the effects of model bias in this system shows that, as expected, it can lead to a severe degradation of both real-time and retrospective analyses. The degradation can be mitigated by boosting the assumed variance of the model random error. The amplitude which gives the best results depends not only on the model bias but also on the relationship between the model random and observational errors. Thus, depending on the latter, the amplitude correction may be either much larger or much smaller than the squared bias amplitude.

Observational bias in unstable modes is amplified by the Kalman filter.

In the context of current operational data assimilation systems these results may appear to be of purely theoretical interest because of the apparently insurmountable obstacle presented by the huge computational cost of applying the Kalman smoother to large systems (scaling as the cube of the number of control variables). The Kalman smoother algorithm for finding the optimal solution is not, however, the most efficient for large systems. The generalised inversion method (Bennett, 1992) derives the same optimal solution (apart from differences in finite difference approximations which might be associated with the different algorithms) at a cost proportional to the product of the total number of control variables and the total number of observations in a time period. This method has been applied to systems for which the Kalman smoother algorithm would be prohibitively expensive. Recent work (Juckes, 2005) deploys a new method which solves for the optimal solution defined by the Kalman smoother at a much reduced cost. In that case the cost grows only linearly with system size for large systems.

Section 4.2 shows that empirical estimates of model error determined by minimising the analysis error need to be treated with caution: they do not necessarily have a simple relation to the actual model error if the bias is a significant factor.

While the results we present are only definitively applied to linear systems, they are indicative of what one might expect in non-linear systems, and so their implications for reanalysis work and analysis archives are important. If a good estimate of the flow is available and the difference between the analysis and the estimate is small, it may be that the assimilation problem is close to being linear even when the underlying behaviour of the system is non-linear. Tremolet (2004) suggests that this may be the case for the ECMWF analysis system. If linearity gives a reasonable approximation, so that a system of arbitrary complexity can be built up from a superposition of many modes, then results such as those presented here are of relevance. Even then, however, a multimode system has the additional complexity that observations will not generally be of the individual modes but of linear combinations of the modes, so that the errors associated with the modes interact even when the modes themselves are physically independent.

Given algorithms exist which provide (affordable) solutions consistent with those obtained with the Kalman smoother for some classes of analysis problems, then this paper demonstrates that using them could provide significantly better estimates of the state of those systems.

6. Appendix A: Analytic integration of the noise-driven linear differential equation

Analytic integration of the unforced equation is trivial: this appendix shows how to extend the calculation to include the random term. Equation (1) can then be integrated, without approximation, to give

$$u(t+\mathrm{d}t) = u(t)\exp(-\alpha \mathrm{d}t) - \frac{\gamma}{\alpha}[1-\exp(-\alpha \mathrm{d}t)] + \overline{W},$$

where the random component is

$$\overline{W} = \int_0^{dt} W(t+\tau)\exp\left[-\alpha(\mathrm{d}t-\tau)\right]\,\mathrm{d}\tau.$$

The variance of $\overline{W}$ now needs to be determined.

It follows from the definition of the Wiener processes W that

$$\begin{aligned} E[(\overline{W})^2] &= E\left[\int_0^{dt}\int_0^{dt} \varsigma^2 W(t+\tau)W(t+\tau') \right. \\ &\quad \left. \times \exp[-\alpha(2\,\mathrm{d}t-\tau-\tau')]\,\mathrm{d}\tau\,\mathrm{d}\tau'\right] \\ &= \int_0^{dt} \varsigma^2 \exp[-2\alpha(\mathrm{d}t-\tau)]\,\mathrm{d}\tau \\ &= \frac{\varsigma^2\,(1-\exp[-2\alpha \mathrm{d}t])}{2\alpha}. \end{aligned} \tag{A1}$$

For small $\alpha\,\mathrm{d}t$ this reduces to

$$E[(\overline{W})^2] \approx \varsigma^2\,\mathrm{d}t\,(1+\alpha \mathrm{d}t) \approx \varsigma^2\,\mathrm{d}t.$$

7. Appendix B: The Kalman filter and smoother

The Kalman filter equations (e. g. Rogers, 2000) are as follows:

$$x_{fk} = P\hat{x}_{fk-1} \tag{B1a}$$

$$B_{fk} = P\hat{B}_{fk-1}P^T + P^{1/2}Q_k P^{1/2\,T} \tag{B1b}$$

$$\hat{x}_{fk} = x_{fk} + K(y_k - Hx_{fk}) \tag{B1c}$$

$$\hat{B}_{fk} = (I-KH)\,B_{fk} \tag{B1d}$$

$$K = B_{fk}H^T\left(HB_{fk}H^T + R_k\right)^{-1} \tag{B1e}$$

Subscripts f, b and s will be used to indicate the Kalman filter, backward filter and smoother, respectively. A subscript k, $k\pm1$ will be used to label the time interval. Following standard notation, Q is the noise covariance matrix, H the observation operator and R the observation error covariance. B will be the analysis error covariance. The state estimate x and error covariance B at any time level are written with a $\hat{}$ if they include observations from the current time level, otherwise these quantities refer to estimates which do not use the current observations. A T superscript indicates a transpose.

The k subscript is omitted from H and K because for these operators the time is clear from the terms they are associated with. An algebraically more simple form of (B1c, d) is

$$\hat{B}_{*k}^{-1}\hat{x}_{*k} = B_{*k}^{-1}x_{*k} + H^T R_k^{-1} y_k, \tag{B2a}$$

$$\hat{B}_{*k}^{-1} = B_{*k}^{-1} + H^T R_k^{-1} H, \tag{B2b}$$

where the $*$ in the subscript can be f, b or s. (B1c, d) are preferred in numerical applications if the rank of R is less than the rank of B. The set (B1) avoids inversion of B, which is an enormous saving in many instances. Appendix C will, however, use the algebraically simpler form (B1b).

Similar equations apply for the backward filter, simply replacing P with P^{-1}, and appropriate changes in indexing:

$$x_{bk} = P^{-1}\hat{x}_{bk+1}, \tag{B3a}$$

$$B_{bk} = P^{-1}\hat{B}_{bk+1}P^{-1\,T} + P^{-1/2}Q_{k+1}P^{-1/2\,T}, \tag{B3b}$$

along with the general expressions in equation (B2).

The Kalman smoother is then

$$\hat{x}_{sk} = \hat{B}_{sk}\left(B_{fk}^{-1}x_{fk} + B_{bk}^{-1}x_{bk} + H^T R_k^{-1} y_k\right), \tag{B4a}$$

$$\hat{B}_{sk}^{-1} = B_{fk}^{-1} + B_{bk}^{-1} + H^T R_k^{-1} H. \tag{B4b}$$

8. Appendix C: Error evolution with imperfect models and error statistics

8.1. C.1 random error

If the deterministic model is not correct there will be both a random and a deterministic component of the analysis error. We first look at the random component.

Let $B_f^\dagger$ and $\hat{B}_f^\dagger$ be the estimated error covariances produced by equations (B1) using estimated, rather than perfectly known, input error statistics and model parameters. That is, they obey versions of (B1b) and (B2b) with the correct model and covariances replaced by estimates:

$$B_{fk}^\dagger = P^\dagger \hat{B}_{fk-1}^\dagger P^{\dagger T} + P^{\dagger\,1/2}Q_k^\dagger P^{\dagger\,1/2\,T}, \tag{C1}$$

$$\left(\hat{B}_{fk}^\dagger\right)^{-1} = \left(B_{fk}^\dagger\right)^{-1} + H^T R_k^{-1} H. \tag{C2}$$

Errors in H and R will not be considered here. The random component of the error, B^*, then evolves according to

$$B_{k+1}^* = P\hat{B}_k^* P^T + P^{1/2}QP^{1/2\,T}, \tag{C3}$$

$$\hat{B}^*_{k+1} = \hat{B}^\dagger_f B^{\dagger\,-1}_f B^*_{k+1}\big(\hat{B}^\dagger_f B^{\dagger\,-1}_f\big)^T + \hat{B}^\dagger_f H^T R^{-1} H \hat{B}^{\dagger\,T}_f, \quad \text{(C4)},$$

and hence

$$B^\dagger_f \hat{B}^{\dagger\,-1}_f \hat{B}^*_f \big(B^\dagger_f \hat{B}^{\dagger\,-1}_f\big)^T - P\hat{B}^*_f P^T = P^{1/2} Q P^{1/2\,T} + B^\dagger_f H^T R^{-1} H B^{\dagger\,T}_f. \quad \text{(C5)}$$

It can easily be verified from eqs. (B1) that the optimal error covariances B, $\hat{B}$ satisfy (C5) with $*$ and $\dagger$ superscripts removed. If the equation so obtained for the optimal solution is subtracted from (C5), then an equation for the difference between the actual and the optimal error covariance as a function of the difference between the estimated and the optimal errors is obtained:

$$\big(I + B_f H^T R^{-1} H\big)\big(\hat{B}^*_f - \hat{B}_f\big)\big(I + B_f H^T R^{-1} H\big)^T - P\big(\hat{B}^*_f - \hat{B}_f\big)P^T \approx \big(B^\dagger_f - B\big)H^T R^{-1} H\big(I - \hat{B}_f H^T R^{-1} H\big)\big(B^\dagger_f - B\big)^T, \quad \text{(C6)},$$

where terms in which $B^\dagger_f - B_f$ multiplies $\hat{B}^*_f - \hat{B}_f$ have been neglected. This approximation will be valid when $B^\dagger_f - B_f$ is small.

The difference terms on the right-hand side, $B^\dagger_f - B$ and its adjoint vary linearly with changes in $\sigma^\dagger_m$ when the perturbation from σ_m is small. This shows that the actual error variance varies quadratically with the perturbation:

$$\begin{aligned}\hat{B}^*_s - \hat{B}_s = {} & \hat{B}_s\big[\hat{B}^{-1}_f\big(B^*_f - B_f\big)\hat{B}^{-1}_f + \hat{B}^{-1}_b\big(B^*_b - B_b\big)\hat{B}^{-1}_b\big]\hat{B}_s \\ & + \big(\hat{B}^\dagger_s \hat{B}^{\dagger\,-1}_f - \hat{B}_s\hat{B}^{-1}_f\big)\hat{B}_f\big(\hat{B}^\dagger_s \hat{B}^{\dagger\,-1}_f - \hat{B}_s\hat{B}^{-1}_f\big)^T \\ & + \big(\hat{B}^\dagger_s \hat{B}^{\dagger\,-1}_b - \hat{B}_s\hat{B}^{-1}_b\big)\hat{B}_b\big(\hat{B}^\dagger_s \hat{B}^{\dagger\,-1}_b - \hat{B}_s\hat{B}^{-1}_b\big)^T \\ & + \big(\hat{B}^\dagger_s - \hat{B}_s\big)H^T R^{-1} H\big(\hat{B}^\dagger_s - \hat{B}_s\big)^T.\end{aligned}$$

This gives a combination of the errors in the forward and backward components and an additional error due to the inaccurate combination of the two.

8.2. C.2 bias

Since the equation system is linear, the bias evolves independently of the solution and the random error. The amplitude of the bias in the solution results from the competition between the generation of bias by the forward model and the removal of bias by the addition of unbiased observations:

$$b_{fk+1} = P\hat{b}_{fk} + \delta\Gamma, \quad \text{(C7a)}$$

$$\hat{b}_{fk+1} = \hat{B}_f B^{-1}_f b_{fk+1} + K\delta y. \quad \text{(C7b)}$$

If $\hat{b}_{fk+1} = \hat{b}_{fk} \equiv \hat{b_f}$ then

$$\hat{b}_f = \big(1 - \hat{B}_f B^{-1}_f P\big)^{-1}\big(\hat{B}_f B^{-1}_f \delta\Gamma + K\delta y\big), \quad \text{(C8a)}$$

$$\hat{b_b} = \big(1 - \hat{B}_b B^{-1}_b P^{-1}\big)^{-1}\big(-\hat{B}_b B^{-1}_b \delta\Gamma + K\delta y\big), \quad \text{(C8b)}$$

$$\hat{b_s} = \hat{B}_s\big(\hat{B^{-1}_f}\hat{b_f} + \hat{B^{-1}_b}\hat{b_b} - R^{-1}\delta y\big). \quad \text{(C8c)}$$

References

Bell, M. J., Martin, M. J. and Nichols, N. K. 2004. Assimilation of data into an ocean model with systematic errors near the equator. *Q. J. R. Meterol. Soc.* **130**, 873–893.

Bennett, A. F. 1992. Inverse methods in physical oceanography. Cambridge University Press, New York, NY, 347 pp.

Cohn, S. E., Sivakumaran, N. S. and Todling, R. 1994. A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Weather Rev.* **122**, 2838–2867.

Dee, D. P. and Da Silva, A. M. 1998. Data assimilation in the presence of forecast bias. *Q. J. R. Meterol. Soc.* **124**, 269–295.

Dee, D. P. and Todling, R. 2000. Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon. Weather Rev.* **128**, 3268–3282.

Griffith, A. K. and Nichols, N. K. 1996. Accounting for model error in data assimilation using adjoint methods. In: *Computational Differentiation: Techniques, Applications and Tools* (eds. M. Berz, C. Bischof, G. Corliss, and A. Greiwank). SIAM, Philadelphia, 195–204.

Juckes, M. N. 2005. The direct inversion method for data assimilation using isentropic tracer advection. *Atmos. Chem. Phys. Discuss.* **5**, 8879–8923.

Kalman, R. E. and Bucy, R. 1961. New results in linear filtering and prediction. *J. Basic Eng.* **83D**, 95–108.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. and co-authors. 1996. The NCEP/NCAR Reanalysis Project. *Bull. Am. Meteorol. Soc.* **77**, 437–471.

Kloeden, P. E. and Platen, E. 1992. *Numerical solution of stochastic differential equations*. Springer, Berlin, 292 pp.

Li, Z. and Navon, I. M. 2001. Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. R. Meterol. Soc.* **127**, 661–684.

Rogers, C. D. 2000. *Inverse methods for atmospheric sounding*. World Scientific Publishing, London, 238 pp.

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

Tremolet, Y. 2004. Diagnostics of linear and incremental approximations in 4D-Var. *Q. J. R. Meterol. Soc.* **130**, 2233–2251.