

Multiple time level adjustment for data assimilation

By S. ZHANG*, J. L. ANDERSON, ANTHONY ROSATI, MATTHEW HARRISON, SHREE P. KHARE and ANDREW WITTENBERG, *GFDL/NOAA, Princeton University, P.O. Box 308, Princeton, NJ 08542, USA*

(Manuscript received 11 February 2003; in final form 25 September 2003)

ABSTRACT

Time-stepping schemes in ocean–atmosphere models can involve multiple time levels. Traditional data assimilation implementation considers only the adjustment of the current state using observations available, i.e. the one time level adjustment. However, one time level adjustment introduces an inconsistency between the adjusted and unadjusted states into the model time integration, which can produce extra assimilation errors. For time-dependent assimilation approaches such as ensemble-based filtering algorithms, the persistent introduction of this inconsistency can give rise to computational instability and requires extra time filtering to maintain the assimilation.

A multiple time level adjustment assimilation scheme is thus proposed, in which the states at times t and $t - 1$, $t - 2, \dots$, if applicable, are adjusted using observations at time t . Given a leap frog time-stepping scheme, a low-order (Lorenz-63) model and a simple atmospheric (global barotropic) model are used to demonstrate the impact of the two time level adjustment on assimilation results in a perfect model framework with observing/assimilation simulation experiments. The assimilation algorithms include an ensemble-based filter (the ensemble adjustment Kalman filter, EAKF) and a strong constraint four-dimensional variational (4D-Var) assimilation method. Results show that the two time level adjustment always reduces the assimilation errors for both filtering and variational algorithms due to the consistency of the adjusted states at times t and $t - 1$ that are used to produce the future state in the leap frog time-stepping. The magnitude of the error reduction made by the two time level adjustment varies according to the availability of observations, the nonlinearity of the assimilation model and the strength of the time filter used in the model. Generally the sparser the observations in time, the larger the error reduction. In particular, for the EAKF when the model uses a weak time filter and for the 4D-Var method when the model is strongly nonlinear, two time level adjustment can significantly improve the performance of these assimilation algorithms.

1. Introduction

Data assimilation uses model dynamics and observations to reconstruct the structure of a geofluid in time and space, providing on one hand the best estimate of initial conditions for numerical prediction, and on the other hand, reconstructed time series to further understanding of the dynamical and physical mechanisms of the geofluid evolution. Two typical representatives of modern data assimilation methods are variational approaches, e.g. four-dimensional variational data assimilation (4D-Var) (Le Dimet and Talagrand, 1986) and filtering approaches, e.g. the ensemble Kalman filter (Evensen, 1994) or the ensemble adjustment Kalman filter (EAKF) (Anderson, 2001). The former solves for an optimal initial input state by minimizing a distance measurement (cost function) between the model and observational trajectories over a time verification window, while the latter updates a set of ensemble members that represent the product of the modeled (prior) and observational probability distributions.

The traditional implementation of both variational and filtering approaches considers only the adjustment of the state at the most recent time using observations available, i.e. the one time level adjustment. However, for finite-differencing accuracy one often chooses a multiple-level time-stepping scheme to advance the model, using model states at previous time levels to derive the next model state. The two time level leap frog scheme is often used; for a simple linear equation such as $\partial\phi/\partial t = \kappa\phi$, where ϕ is a scalar and κ is a constant, this scheme advances ϕ for one step, Δt , as

$$\phi_{t+1} = \phi_{t-1} + \kappa\phi_t 2\Delta t. \quad (1)$$

Equation (1) shows that the state at the next time step, ϕ_{t+1} depends on the present state ϕ_t and the previous state ϕ_{t-1} . A traditional one time level adjustment uses the observation at time t to adjust only ϕ_t , introducing inconsistency between the unadjusted ϕ_{t-1} and the adjusted ϕ_t into the time integration.

In a realistic assimilation system, the resulting analysis error from the one time level adjustment is much more complicated than the simple case above. First, a realistic numerical model

*Corresponding author
e-mail: snz@gfdl.noaa.gov

contains nonlinear terms, which may amplify the error introduced by the one time level adjustment. Secondly, it is not clear how the observational interval affects the analysis error, since while a large observational interval reduces the frequency of introducing the inconsistency, it also reduces the constraint of observations on the analysis. Thirdly, the time filter in the model used to damp computational modes produced by the multiple time level differencing scheme may impact the analysis error since it can help damp the inconsistency introduced by the one time level adjustment.

In one time level adjustment scheme one may choose to restart the model for each analysis step using an initial forward (Euler) step. Yet the one time level forward restart time scheme also introduces extra errors into assimilation results, since the initial forward step may change the model trajectory in each observational interval (for filtering algorithms) or each iteration in assimilation window (for 4D-Var) due to spin-up.

Also important is the time dependence of the assimilation scheme itself in which the one/two time level adjustment has a different impact on the assimilation. Generally a filtering algorithm has a strong time dependence since each step analysis is based on the prior distribution derived from the previous analysis through the model integration, while the time dependence of a variational algorithm is relatively weak since the model forecast from the previous analysis serves only as a first guess of the next step analysis.

We shall look at two different methods with quite different time dependences. We then introduce a multiple time level adjustment scheme that is able to reduce the inconsistency between time levels. Using a low-order (Lorenz-63) model and a simple atmospheric (barotropic) model, given a leap frog time-stepping scheme this study examines the impact of the two time level adjustment on assimilation errors. The assimilation algorithms examined include an ensemble-based filter, the ensemble adjustment Kalman filter (EAKF, Anderson, 2001) and a strong constraint four-dimensional variational (4D-Var) assimilation method Le Dimet and (Talagrand, 1986). Section 2 describes methodology, including the fundamentals of models, the EAKF and the 4D-Var, as well as the implementation of two time level adjustment in the EAKF and the 4D-Var. Sections 3 and 4 examine the impact of the two time level adjustment on the EAKF and 4D-Var assimilation results, respectively. Section 5 gives conclusions and discussions.

2. Methodology

2.1. Models, time stepping and time filtering

The notation below follows Ide et al. (1997). The assimilation models (the Lorenz-63 model, see Appendix A and the global barotropic spectral model, see Appendix B) in this study have

the matrix form

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{F}(\mathbf{x}), \quad (2)$$

where \mathbf{x} represents the model state vector and $\mathbf{F}(\mathbf{x})$ is typically a nonlinear function of \mathbf{x} in ocean-atmosphere models. The model state vector \mathbf{x} consists of either the three Lorenz-63 model variables or the 64×54 gridpoints of the barotropic streamfunction. We may discretize eq. (2) using, for example, a leap frog scheme:

$$\mathbf{x}_{t+1} = \mathbf{x}_{t-1} + \mathbf{F}(\mathbf{x}_t) \cdot 2\Delta t, \quad (3)$$

where \mathbf{x}_{t+1} always depends on the state vectors \mathbf{x}_t and \mathbf{x}_{t-1} .

A disadvantage of the leap frog time stepping is that the computational mode introduced due to the use of two time level model states in time integration may amplify by nonlinear coupling with physical modes (Durrant, 1999). A time filter (Robert, 1969; Asselin, 1972) may be used to damp the computational modes:

$$\bar{\mathbf{x}}_t = \frac{1}{2}\epsilon\bar{\mathbf{x}}_{t-1} + (1 - \epsilon)\mathbf{x}_t + \frac{1}{2}\epsilon\mathbf{x}_{t+1}, \quad (4)$$

where ϵ controls the strength of the time filter. Combining eqs. (3) and (4) shows that $\bar{\mathbf{x}}_{t+1}$ is associated with \mathbf{x}_t , \mathbf{x}_{t-1} and \mathbf{x}_{t-2} . Since \mathbf{x}_{t-2} is only used for the recalculation of \mathbf{x}_{t-1} after the time integration and does not explicitly appear in the time integration for obtaining $\bar{\mathbf{x}}_{t+1}$, the experiment designs in Sections 2.3 and 2.4 only consider the adjustment of \mathbf{x}_t and \mathbf{x}_{t-1} using the observations at t .

2.2. observing/assimilation simulation experiments

In order to address the issue of the impact of the multiple time level adjustment on assimilation, here observing/assimilation simulation experiments are conducted in a ‘‘perfect model’’ context. For both the Lorenz-63 model and the barotropic model, ‘‘observations’’ are produced by adding a Gaussian random noise to the ‘‘truth’’, a single long time integration, independently for each model variable (gridpoint). The initial conditions used in the assimilation experiments are the 10^6 -step integration results starting from $(x_1, x_2, x_3) = (0, 1, 0)$ for the Lorenz-63 model and the streamfunction at 12 UTC 1 January 1991 derived from ECMWF (European Centre for Medium-Range Weather Forecasts) re-analysis 500-hPa u and v data for the barotropic model. The standard deviation of the observational errors for the Lorenz-63 model variables and the barotropic streamfunction is set to be 2 and $10^6 \text{ m}^2 \text{ s}^{-1}$, respectively.

Ensemble initial conditions (the ensemble size is 20) are formed by imposing a Gaussian random noise with the same standard deviation as the observations on each model variable independently. Then characteristics of the models such as variability and error growth rate can be examined by allowing the ensemble to evolve freely. For the Lorenz-63 model ($\epsilon = 0.005$, $\Delta t = 0.0001$) after 15 000 steps the freely evolving ensemble members start to separate to the different lobes on the attractor,

and the barotropic model ($\epsilon = 0.02$, $\Delta t = 30$ min) streamfunction after 30 days attains a “climatological” standard deviation of around $20 \times 10^6 \text{ m}^2 \text{ s}^{-1}$.

2.3. An ensemble adjustment Kalman filter

Filtering data assimilation approaches emphasize the probabilistic nature of the dynamical/observational system of the atmosphere and ocean (Jazwinski, 1970). An ensemble-based filter, in particular, computes the probability distribution of the model state using ensemble error statistics that account for the nonlinear evolution of model error covariance (Evensen, 1994; Miller et al. 1994, 1999; Houtekamer and Mitchell, 1998, 2001; Burgers et al. 1998; Van Leeuwen, 1999; Anderson and Anderson, 1999; Keppenne, 2000; Mitchell and Houtekamer, 2000; Bishop et al. 2001; Hamill et al. 2001; Anderson, 2001; Whitaker and Hamill, 2002). The core of these filtering algorithms solves for the product of the prior distribution of the system state, which is governed by the model dynamics, and the observational error distribution (a function of the observing system, typically assumed to be Gaussian), to compute a conditional probability distribution of the system state.

The ensemble adjustment Kalman filter (EAKF, Anderson, 2001) is chosen to examine the impact of two time level adjustment on a filtering assimilation algorithm. Ensemble-based filters like the EAKF can be applied sequentially to the individual scalar observations (if the observational errors are uncorrelated) or to a batch of correlated observations (Houtekamer and Mitchell, 2001). In addition, the impact of each scalar observation (uncorrelated) on each component of the state vector can be computed independently (Anderson, 2003). The EAKF defines a joint state/observation space $\mathbf{z} = \{x, h(\mathbf{x})\}$, where h is an operator that gives the expected value of the observation given a state vector. In this space, a statistically linear relation between state vectors and observation variables can be computed from the ensemble sample.

Like other filtering techniques, the EAKF can experience filter divergence (Jazwinski, 1970) in which the distribution produced by the filter drifts away from the truth. In order to avoid filter divergence, the EAKF algorithm is used here in conjunction with a covariance inflation parameter γ to increase the variance of the prior distribution, thereby enhancing the impact of the observations on the assimilated state (Anderson, 2001) as $\gamma(\mathbf{x}_i^p - \bar{\mathbf{x}}^p)$, where \mathbf{x}_i^p and $\bar{\mathbf{x}}^p$ represent the i th ensemble member and the ensemble mean respectively. Here, γ is chosen so that the ratio of the time-averaged root mean square (Rms) error of the ensemble mean (RmsEm) to the time-averaged mean Rms error of the individual ensemble members (MRms) is close to $\sqrt{(M+1)/2M}$. For an ensemble of size M , the expected value of the ratio of RmsEm to MRms is $\sqrt{(M+1)/2M}$ for an ensemble that has a variance that is consistent with the truth (Anderson, 1996). A ratio close to the expected value implies that the ensemble has

a variance that is approximately consistent with the truth. The use of relatively small ensemble samples can lead to spuriously large correlations between state variables and observation variables that are believed to be uncorrelated a priori due to sampling errors. To avoid the contamination of the assimilation by spurious correlations, the prior sample correlation between a state variable and observation variables is adjusted by a smoothly varying distance-dependent weight $\Omega(a, d)$ (see Appendix C for a definition) (Hamill et al. 2001) in the barotropic model experiments. Here, d is the Euclidean distance between the model grid point and the observation location and a is a measurement of the width of this observation impact window.

The EAKF adjustment consists of three steps. First, the numerical model (2) is advanced to the time of the next observation for each ensemble member to form a sample of the prior state distribution. The forward observation operator, h , is applied to each sample of the prior state to obtain a prior sample of the expected value of the observations. Next, the prior sample mean and covariance of the joint state/observation vector, Σ^p , are computed. The updated covariance matrix, Σ^u and mean $\bar{\mathbf{z}}^u$ are computed by taking the product of the Gaussian with the prior sample mean and covariance with the Gaussian distribution from the observations. Finally, a linear operator \mathbf{A} is computed (Appendix A in Anderson, 2001) using Σ^u and $\bar{\mathbf{z}}^u$ and each ensemble member is updated as

$$\mathbf{z}_i^u = \mathbf{A}^T (\mathbf{z}_i^p - \bar{\mathbf{z}}^p) + \bar{\mathbf{z}}^u, \quad i = 1, \dots, M \quad (5)$$

resulting in an updated sample with the mean and variance of the state variables exactly equal to the update values computed in the product.

Owing to the sequential nature of the EAKF, implementing two time level adjustment in the algorithm is straightforward, by extending the definition of the state being adjusted to include both \mathbf{x}_i and \mathbf{x}_{i-1} . When observations become available, their prior expected value can be computed from \mathbf{x}_i , as before, but the update impacts the state variables at both time levels. We compare two time level adjustment of leap frog (LF) time scheme (denoted by two^{LF} hereafter) to two one time level adjustment implementations: if the model is restarted after each analysis step from a forward (FR) time step starting from \mathbf{x}_i^a , the EAKF is referred to as an one^{FR} method; if the leap frog time stepping is continued using the unadjusted \mathbf{x}_{i-1} and the adjusted \mathbf{x}_i^a for the next analysis step, the EAKF is referred to as an one^{LF} method.

The essence of the multiple time level adjustment is the use of the temporal structure of the background error covariance of the model state variables. An example of the temporal correlation of the model state variables is given in Fig. 1 where the time or/and space mean of the temporal correlation of x_1 (panel a) in the Lorenz-63 model, the streamfunction (panel b) in the barotropic model was estimated by the EAKF over the assimilation period step $9 \times 10^5 \sim$ step 10×10^5 (for the Lorenz-63 model) and day 90 \sim day 100 (for the barotropic model) using observations

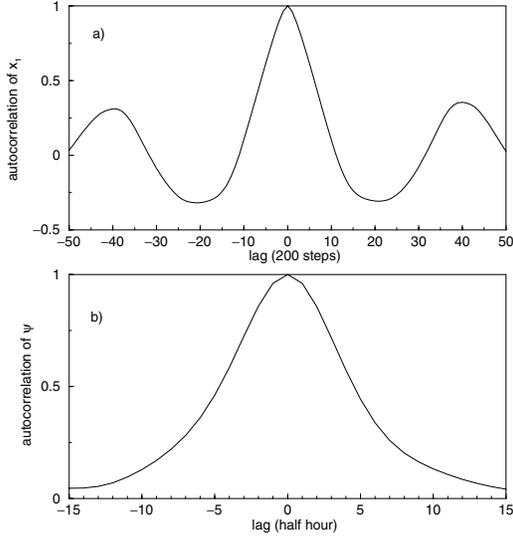


Fig 1. (a) Time-averaged auto-correlation from the EAKF of x_1 of the Lorenz-63 model over the last 10^5 steps of the 10^6 -step assimilation period. (b) Global-mean of the time-averaged auto-correlation of the streamfunction of the barotropic model over the last 10 days of the 100-day assimilation period.

available every 200 steps (for the Lorenz-63 model) and 6 h (for the barotropic model). An ensemble-based filter provides estimates of the covariance between the model state variables at different times (Zhang and Anderson, 2003), so as to be able to adjust the states at different time levels accordingly given observations available at a certain time level. This means that it is possible to create a limited time window ensemble smoother (Jazwinski, 1970; Fukumori, 2002), which can produce a consistent temporal evolution sequence of assimilation data. The development of this kind of fixed lag smoother is an on-going research topic, which is beyond the scope of this study.

2.4. Four-dimensional variational assimilation

For efficient implementation of the 4D-Var algorithm, the adjoint of the tangent linear version of the numerical model is required. In the tangent linear model (TLM), all nonlinear terms in eq. (2) are differentiated, giving a matrix form for the Lorenz-63 model (A4–A6) and the global barotropic spectral model (B4) of

$$\frac{\partial \delta \mathbf{x}}{\partial t} = \mathbf{F}'(\mathbf{x}) \delta \mathbf{x} \quad (6)$$

where a $\delta(\cdot)$ represents a perturbation and a (\cdot) represents the basic state. $\mathbf{F}'(\mathbf{x})$ is the first derivative of $\mathbf{F}(\mathbf{x})$ with respect to \mathbf{x} . Next, the adjoint of the finite difference of this equation (Sirkes and Tziperman, 1997) is formulated by transposing all DO loops and subroutines in the Fortran implementation of the TLM. If $\mathbf{M} = \mathbf{M}_n \cdots \mathbf{M}_2 \mathbf{M}_1$ represents the propagator of the TLM such as $\delta \mathbf{x}_t = \mathbf{M}_n \cdots \mathbf{M}_2 \mathbf{M}_1 \delta \mathbf{x}_0$, then a transposed version $\mathbf{M}^T = \mathbf{M}_1^T \mathbf{M}_2^T \cdots \mathbf{M}_n^T$ represents its adjoint. An inner product check

on the adjoint formulation, $\langle \mathbf{M} \delta \mathbf{x}, \mathbf{M} \delta \mathbf{x} \rangle = \langle \delta \mathbf{x}, \mathbf{M}^T \mathbf{M} \delta \mathbf{x} \rangle$ agreed to 15 decimal digits using 64-bit arithmetic when the Lorenz-63 model is run for 10^6 steps and the barotropic model is run at rhomboidal 21 truncation for 120 h with a 30-min time step.

The cost functions J_1 (for one time level adjustment) and J_2 (for two time level adjustment) are defined as

$$\begin{aligned} J_1(\mathbf{x}_0) &= \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_{t_0}^o)^T \mathbf{w}_0 (\mathbf{x}_0 - \mathbf{x}_{t_0}^o) \\ &\quad + \frac{1}{2} \sum_{t=t_1}^{t_R} (\mathbf{x}_t^f - \mathbf{x}_t^o)^T \mathbf{W} (\mathbf{x}_t^f - \mathbf{x}_t^o) \\ J_2(\mathbf{x}_{-1}, \mathbf{x}_0) &= \frac{1}{2} (\mathbf{x}_{-1} - \mathbf{x}_{t_0}^o)^T \mathbf{w}_{-1} (\mathbf{x}_{-1} - \mathbf{x}_{t_0}^o) \\ &\quad + \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_{t_0}^o)^T \mathbf{w}_0 (\mathbf{x}_0 - \mathbf{x}_{t_0}^o) \\ &\quad + \frac{1}{2} \sum_{t=t_1}^{t_R} (\mathbf{x}_t^f - \mathbf{x}_t^o)^T \mathbf{W} (\mathbf{x}_t^f - \mathbf{x}_t^o) \end{aligned} \quad (7)$$

where \mathbf{x}_{-1} and \mathbf{x}_0 represent the analysis of the model variables at t_{-1} and t_0 . \mathbf{x}_t^f and \mathbf{x}_t^o represent the modeled and observed vectors of the model variables at time t . All are column vectors that consist of either three (for the Lorenz-63 model) or 64×54 (for the barotropic model) elements. $[t_0, t_R]$ defines the time width of an assimilation window. \mathbf{W} is a weighting matrix, usually defined as the inverse of the covariance matrix of the model state variables (for the background term) or the inverse of the observational error covariance matrix (for the observational term). In this study \mathbf{W} is either the identity matrix (for the Lorenz-63 model) or a diagonal constant matrix with the inverse of the maximum variation of streamfunction in the assimilation window (for the barotropic model) and \mathbf{w}_{-1} and \mathbf{w}_0 are the corresponding weighting column vectors at times t_{-1} and t_0 (in this study, they are assumed to be the same). In practical data assimilation implementation, the background covariance matrix can be estimated by the correlation of time series at different locations (time-invariant) (Thiebaux, 1985; Hollingsworth and Lönnberg, 1986) or a Monte Carlo approach in ensemble filters (temporally varying) (Zhang and Anderson, 2003). Once the cost function is defined, the adjoint model is used to evaluate the gradient of the cost function with respect to \mathbf{x}_{-1} and \mathbf{x}_0 by a backward integration as

$$\begin{aligned} -\frac{\partial \hat{\delta \mathbf{x}}}{\partial t} - \left[\frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} \right]^T \hat{\delta \mathbf{x}} &= \mathbf{W}(\mathbf{x} - \mathbf{x}^o) \\ \nabla_{|\mathbf{x}_0} J_1 \text{ (or } \nabla_{|\mathbf{x}_0} J_2) &= \int_{t_R}^{t_1} \hat{\delta \mathbf{x}}_0 dt + \mathbf{W}_0 (\mathbf{x}_0 - \mathbf{x}_{t_0}^o) \\ \nabla_{|\mathbf{x}_{-1}} J_2 &= \int_{t_R}^{t_0} \hat{\delta \mathbf{x}}_{-1} dt + \mathbf{w}_{-1} (\mathbf{x}_{-1} - \mathbf{x}_{t_0}^o) \end{aligned} \quad (8)$$

where a $\hat{(\cdot)}$ represents the adjoint variable and $\hat{\delta \mathbf{x}}_0$, $\hat{\delta \mathbf{x}}_{-1}$ represent the adjoint variables that are related to \mathbf{x}_0 and \mathbf{x}_{-1} , respectively. The gradient test (Appendix D) shows that the integrations of these adjoint models correctly evaluate the gradient of the cost functions defined on different assimilation windows.

Once $\nabla|_{\mathbf{x}_0}J_1$ and $\nabla|_{\mathbf{x}_{-1}}J_2$, $\nabla|_{\mathbf{x}_0}J_2$ are available, a limited memory quasi-Newton optimization algorithm (Liu and Nocedal, 1989) is employed to minimize J_1 with respect to \mathbf{x}_0 or J_2 with respect to \mathbf{x}_0 and \mathbf{x}_{-1} so that the optimal \mathbf{x}_0 (for one time level adjustment) or the optimal \mathbf{x}_{-1} and \mathbf{x}_0 (for two time level adjustment, two^{LF}) are obtained. In the one time level adjustment minimization, if the model always is restarted for each iteration from a forward time stepping at the first integration step using \mathbf{x}_0^a produced by previous iteration, the minimization process is implementing an one^{FR} time level 4D-Var adjustment. If the model keeps the leap frog stepping using the unadjusted \mathbf{x}_{-1} and the adjusted \mathbf{x}_0^a the minimization process is implementing an one^{LF} time level 4D-Var adjustment.

3. Impact of two time level adjustment on the EAKF

3.1. Two^{LF} vs. one^{LF} (one^{FR}) in the Lorenz-63 model

Using the ensemble initial conditions and observations described in Section 2.2, the EAKF assimilations with the one^{LF}, one^{FR} and two^{LF} adjustments described in Section 2.3 are run with the Lorenz-63 model for 10^6 steps. Each experiment produces the time-averaged root mean square error of the ensemble mean and the time-averaged mean Rms error from the individual ensemble members and for each assimilation experiment, the ensemble filter is tuned by changing the covariance inflation factor (γ) to obtain a ratio of RmsEm to MRms that is within 1% of the expected value. Table 1 lists the RmsEm values for different observational intervals. The assimilation experiments are carried out for different Robert–Asselin time filters, $\epsilon = 0.005$ (upper rows in Table 1) and $\epsilon = 0.01$ (lower rows in Table 1). The statistics for Table 1 are computed using all analysis steps over the whole 10^6 -step period (e.g. 200 analysis steps for the 5000-step observational interval).

Generally, as the observational interval increases the assimilation produces a relatively larger error for both the one^{LF} and two^{LF} adjustments (columns 3 and 5) since fewer data are used in the assimilation to constrain the ensemble. However, columns 3, 5 in Table 1 also show that for each case the two^{LF} has significantly smaller assimilation errors than the one^{LF}. When the observational interval is small, the reduction of the assimilation error is relatively small while as the observational interval increases, the error reduction increases. For example, for the 200-step observational interval the assimilation errors are reduced by around 38% for the two^{LF} compared to the one^{LF} while for the 5000-step observational interval the assimilation errors are reduced by approximately 66%. Note that the assimilation results for the two^{LF} with the 5000-step observational interval are better than those of the one^{LF} using the 1000-step observational interval.

Table 1. Comparisons of assimilation results by the ensemble adjustment Kalman filter (EAKF) implemented by the one^{LF}, one^{FR} or two^{LF} adjustment over a 10^6 -step period for different observational intervals, using the Lorenz-63 model with different time filters

Time filtering coef (ϵ)	Obs interval (steps)	RmsEm		
		one ^{LF}	one ^{FR}	two ^{LF}
0.005	100	0.17	0.38	0.12
	200	0.27	0.34	0.17
	500	0.45	0.58	0.28
	1000	1.24	1.18	0.57
	2000	2.35	1.95	0.87
	5000	3.20	3.30	1.16
	free ens	–	–	6.73
0.01	100	0.15	0.21	0.12
	200	0.23	0.25	0.16
	500	0.52	0.49	0.29
	1000	1.07	1.03	0.44
	2000	1.65	1.98	0.83
	5000	3.98	3.98	1.35
	free ens	–	–	6.68

RmsEm – the time-averaged Rms error of the ensemble mean.

Free ens – all ensemble members are allowed freely evolving without any observation constraint.

Figure 2 presents time series of the first ten ensemble members (dotted), and the ensemble mean (long-dashed) of x_1 for the one^{LF} (panel a) and two^{LF} (panel b) assimilations using observations available every 5000 steps, between step 1.2×10^5 and step 1.6×10^5 , for the $\epsilon = 0.005$ case. Panel (a) shows that at each analysis step, the inconsistency between the unadjusted \mathbf{x}_{t-1} and the adjusted \mathbf{x}_t^a in the one^{LF} creates high-frequency oscillations (the black cones). During the period shown, although the high-frequency oscillations are damped by the time filter within the 1000-step model integration following each analysis, they prevent the assimilation from efficiently extracting the observational information and drive the ensemble away from the truth (thick-solid). This problem is especially severe when the model trajectory is experiencing a transition from one lobe of the attractor to the other, between step 1.3×10^5 and step 1.4×10^5 , for instance. If the observational interval is large enough that the ensemble members migrate to the lobe not occupied by the truth, the assimilation errors caused by the inconsistency between \mathbf{x}_{t-1} and \mathbf{x}_t^a are quite large. In the two^{LF} in which both \mathbf{x}_{t-1} and \mathbf{x}_t are adjusted using the observations \mathbf{x}_t^o accordingly, the consistently adjusted \mathbf{x}_{t-1}^a and \mathbf{x}_t^a does not create extra computational mode and the assimilated model trajectory remains close to the truth (Fig. 2b, step 1.3×10^5 to step 1.4×10^5 , for instance) after the introduction of a new observation. When observations are available more frequently, the amplitude of the high-frequency oscillations introduced by the one^{LF} is smaller due to smaller

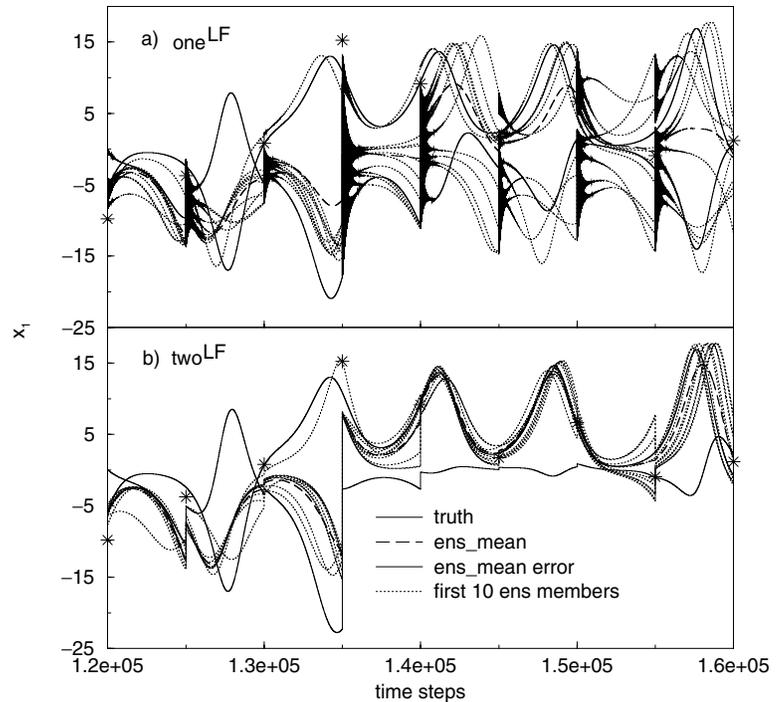


Fig 2. Time series of the first 10 ensemble members (dotted), the ensemble mean (dashed) and the truth (solid) of x_1 of the Lorenz-63 model ($\epsilon = 0.005$) in the EAKF assimilation using the observations available every 5000 steps (asterisk) by (a) the one^{LF} adjustment and (b) the two^{LF} adjustment, over the period between step 1.2×10^5 and step 1.6×10^5 . The black cones reflect the high-frequency oscillations introduced by one time level adjustment in the leap frog time scheme.

shock between \mathbf{x}_{t-1} and \mathbf{x}_t^a , and the ensemble members are less likely to diverge from the true trajectory.

When a relatively strong time filter ($\epsilon = 0.01$) is used in the Lorenz-63 model, for observational intervals greater than 500 steps, the magnitude of the assimilation error reduction by the two^{LF} from the one^{LF} stays around the same as $\epsilon = 0.005$. Although the high-frequency computational mode oscillations in the one^{LF} are more rapidly damped by a stronger time filter, these oscillations can still produce large errors by causing systematic errors soon after an assimilation step that lead to large errors after further integration. However, with a small observational interval (100 or 200 steps in Table 1), since observations constrain the model state more frequently the use of a stronger time filter that rapidly damps the introduced high-frequency oscillations by the one^{LF} reduces the difference between the one^{LF} and two^{LF} assimilations.

The comparison of the percentage of the RmsEm reduction by the two^{LF} from the one^{FR} and the one^{LF} (columns 3–5 in Table 1) shows that like the one^{LF}, the one^{FR} also introduces additional errors into the assimilation. When observations are available more frequently the assimilation errors of the one^{FR} are larger than those of the one^{LF}. Time series of the assimilation results (Fig. 3) indicate that the high-frequency oscillations induced by the one^{FR} are stronger than those from the one^{LF} (panels a and b). When these oscillations are introduced more frequently (for small observational intervals) the assimilation error of the one^{FR} is larger than that of the one^{LF}. Again, Fig. 3c shows the smallest assimilation error is produced by the two^{LF} due to the consistency of \mathbf{x}_{t-1}^a and \mathbf{x}_t^a .

In addition, if a fourth-order Runge Kutta time differencing is used to produce the control run (observations), similar relative assimilation quality from the two^{LF}, the one^{LF} and the one^{FR} is obtained.

3.2. Barotropic model

The ensemble initial conditions and observations (different control runs for different time filters used in the model) described in Section 2.2 are used to start the ensemble filtering assimilation experiments for the barotropic model. Here a in the distance dependent correlation envelope $\Omega(a, d)$ (Appendix C) is set to be 1000 km so that the observation impact radius is 2000 km, which means $\Omega(a, 10^6 m) = 0.5$ while $\Omega(a, 2 \times 10^6 m) = 0$. Assimilation experiments are carried out using 6-, 12- and 24-h observational intervals for the one^{LF}, one^{FR} and two^{LF} adjustments. Different time filters ($\epsilon = 0.02$ – 0.08) are tested for each observational interval.

3.2.1. Two^{LF} vs. one^{LF}. Results of the one^{LF} and two^{LF} adjustments are listed in Table 2. When a relatively weaker time filter is used in the barotropic model, the two^{LF} greatly reduces the assimilation error from the one^{LF}. For three different observational intervals (6, 12 and 24 h), using the weakest time filter ($\epsilon = 0.02$) by which the control run is stable, the assimilation of the one^{LF} blows up after only a few days, while the two^{LF} works well. Figure 4 shows time series of the streamfunction at (113°E, 38°N) for $\epsilon = 0.03$ (left-hand column, panels a, b) and $\epsilon = 0.02$ (right-hand column, panels c, d) using the one^{LF} (top row, panels a, c) and the two^{LF} (bottom row, panels b, d) with observations

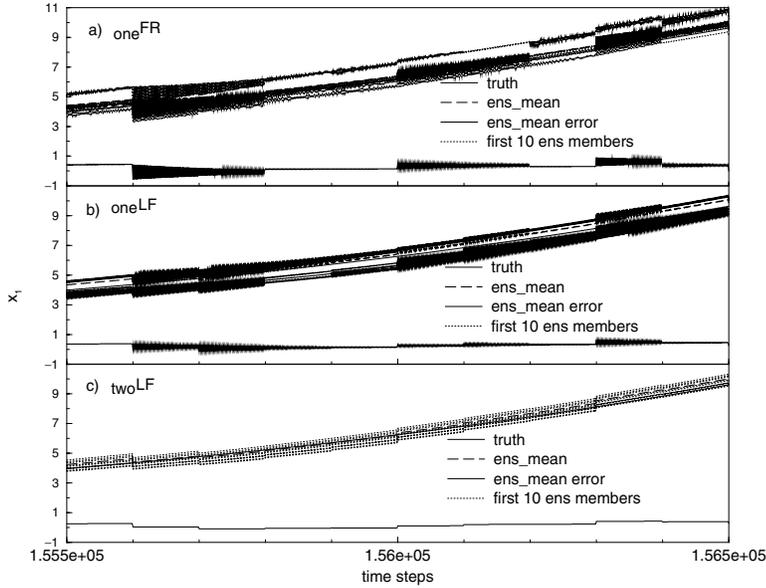


Fig 3. Time series of the first 10 ensemble members (dotted), the ensemble mean (dashed) and the truth (thick-solid) of x_1 of the Lorenz-63 model ($\epsilon = 0.005$ and $\Delta t = 0.0001$) in the EAKF assimilation using observations available every 100 steps by (a) the one^{FR} , (b) the one^{LF} and (c) the two^{LF} .

available every 6 h. The inconsistency between ψ_{t-1} and ψ_t^a in the one^{LF} always induces high-frequency oscillations (panels *a*, *c*) while the two^{LF} produces a much more consistent assimilation with the observations (panels *b*, *d*). The Robert–Asselin time filter in the model, which is used to damp the computational modes produced by the leap frog time stepping, is not sufficient to control the additional computational oscillations induced by the shock in the one^{LF} . Unlike the assimilation experiments in the Lorenz-63 model, here a relatively stronger time filter can always reduce the difference between the one^{LF} and the two^{LF} . Owing to the weak nonlinearity of the barotropic model, a stronger time filter adequately damps the high-frequency oscillations introduced in the one^{LF} and thereby reduces the assimilation errors. For a large observational interval in which the unbalance between ψ_{t-1} and ψ_t^a induced by previous observations has changed the model trajectory away from the truth, increasing the strength of time filtering cannot much improve the one^{LF} assimilation.

For a given filter strength, the improvement of the two^{LF} over the one^{LF} increases as the observational data become sparse in time. When the model prior trajectory is far from the observations, the one^{LF} has trouble constraining the assimilation back to the truth. In addition, Table 2 also shows that for a given observational interval, the strength of time filtering used in the model does not have a substantial impact on the two^{LF} adjustment which does not induce any extra computational oscillations.

3.2.2. One^{LF} vs. one^{FR} . Experiments using the one^{FR} EAKF assimilation show that the time filter used in the model has less impact on assimilation results than for the one^{LF} . For example, with $\epsilon = 0.02$ for a 6-h observational interval the one^{FR} EAKF assimilation gives small assimilation errors (close to those of the two^{LF}) while the one^{LF} does not work. The one^{LF} introduces inconsistency between the unadjusted \mathbf{x}_{t-1} and the adjusted \mathbf{x}_t^a into the initial integration step at each analysis step, while the

one^{FR} introduces a spin-up error for each analysis step by only using \mathbf{x}_t^a to restart the model. This is consistent with the fact that a forward restart scheme is an alternative way to deal with the slow growth of the computational model in the leap frog time scheme (Kalnay, 2002).

Unlike in the low-order model cases presented in Section 3.1, when observations are sparse in time the assimilation errors of the one^{FR} become much larger than those for the one^{LF} . An example ($\epsilon = 0.03$) is shown in Fig. 5 which presents the time series of the streamfunction for the one^{LF} and one^{FR} with $\epsilon = 0.03$ and a 24-h observational interval. Figure 5 shows that with a large observational interval, the model trajectories in the ensemble of the one^{FR} have large departures from the truth (panel *a*) while the one^{LF} does not.

We have found that these conclusions are robust when changing the observation impact radius and/or the ensemble size, for example, doubling the value of a in Ω (*a*, *d*) and/or doubling the ensemble size.

4. Impact of two time level adjustment on 4D-Var

4.1. Two^{LF} vs. one^{LF} (one^{FR}) in the Lorenz-63 model

The same “observations” used in EAKF (Section 3.1) are used to carry out the 4D-Var one^{LF} , one^{FR} and two^{LF} assimilations as described in Section 2.4. The experiments are conducted for different size assimilation windows and different Robert–Asselin time filters. The minimizations for all experiments converge to the point at which the norm of the gradient of the cost function decreases by 10^{-8} from its initial value. This typically requires 10–20 iterations, but depending on the properties of the cost function and the first guess state, some extreme cases may require

Table 2. Comparisons of assimilation results by the ensemble adjustment Kalman filter (EAKF) implemented by the one^{LF} or two^{LF} adjustment over a 100-day assimilation period for different observational intervals, using the barotropic model with different time filters

Obs interval (h)	Time filtering coef (ϵ)	one ^{LF} /two ^{LF} adjustments	
		RmsEm ($10^5 \text{ m}^2 \text{ s}^{-1}$) (reduction %)	MRms ($10^5 \text{ m}^2 \text{ s}^{-1}$) (reduction %)
6	0.02	+/3.16	+/4.00
	0.03	3.95/2.94 (26)	5.17/3.76 (27)
	0.04	3.36/3.06 (9)	4.55/3.89 (15)
	0.05	3.15/2.97 (6)	4.33/3.79 (12)
	0.06	3.09/2.97 (4)	4.25/3.78 (11)
	0.08	2.62/2.67 (-2)	3.61/3.51 (3)
12	0.02	-/3.43	-/4.35
	0.03	4.06/3.20 (21)	5.45/4.12 (24)
	0.04	3.80/3.34 (12)	5.13/4.24 (17)
	0.05	3.58/3.23 (10)	4.90/4.14 (16)
	0.06	3.52/3.22 (9)	4.82/4.11 (14)
	0.08	3.16/2.93 (7)	4.35/3.84 (12)
24	0.02	*/3.92	*/4.96
	0.03	6.36/3.67 (42)	8.75/4.72 (46)
	0.04	5.81/3.77 (35)	7.94/4.80 (39)
	0.05	5.77/3.69 (36)	7.78/4.71 (39)
	0.06	5.54/3.67 (34)	7.50/4.68 (38)
	0.08	5.03/3.36 (33)	5.94/4.39 (26)

RmsEm – the time-averaged Rms error of ensemble mean.

MRms – the time averaged Rms error from individual ensemble members.

“+” – the assimilation blows up at step-468.

“-” – the assimilation blows up at step-624.

“*” – the assimilation blows up at step-480.

up to 100 iterations for the convergence. Once the minimization for a certain analysis step is done, the assimilation time window is shifted forward by one observational step until the assimilation procedure goes through the whole time period.

Since each iteration in minimization involves a line search that requires evaluating the cost function (via a nonlinear model run) and its gradient (via an adjoint model run), 4D-Var is much more expensive than the EAKF given the same frequency of observations available and the relatively small ensemble sizes used here. Obviously, the ratio of the computational costs strongly depends on the width of the 4D-Var assimilation time window and the EAKF ensemble size. For this Lorenz-63 model case, 4D-Var with a 10 000-step assimilation window is approximately 10 times more expensive than the EAKF with 20 ensemble members. All 4D-Var experiments span a 10^5 -step period for the error statistics to compare the assimilation results for the one^{LF}, one^{FR} and two^{LF} adjustments. Except for the first analysis step which uses the observed state as the first guess, the first guess is chosen as the optimized forecast state from the previous analysis step.

Table 3 lists the time mean of the Rms errors of the analyzed state of the one^{LF} and the two^{LF} using 10 000- and 20 000-step assimilation windows for different observational intervals and Robert–Asselin time filters. For all cases, the time mean Rms error is reduced by the two^{LF} compared to the one^{LF}. Generally, the improvement increases with the width of the assimilation window and decreases with the strength of the time filter. For example, when a weaker time filter ($\epsilon = 0.005$) is used, for a 10 000-step assimilation window, the analysis errors are only reduced by 5–13%. For a 20 000-step assimilation window the analysis errors are reduced by up to 70%. Moreover, when a stronger time filter ($\epsilon = 0.01$) is used in the model, for both 10 000- and 20 000-step assimilation windows, the Rms error reduction in all five cases is less than 10%.

These differences can be explained by examining the performance of the minimization at individual analysis steps. Figure 6 presents the modeled trajectories on the (x_1, x_2) phase-plane over the assimilation window in the one^{LF} (panel a) and the two^{LF} (panel b) at the 10th analysis step using the 1000-step observational interval for the $\epsilon = 0.005$ case using a 20 000-step assimilation window. These trajectories are displayed for selected iterations during the minimization including the initial (iteration 0) and the converged (iteration 23 for panel a, iteration 20 for panel b). Panel a shows that the inconsistency of the unadjusted \mathbf{x}_{t-1} and the adjusted \mathbf{x}_t^a produces high-frequency computational oscillations (the black cone in panel a and zoomed-in in panel c) and as a result, the optimized model trajectory (dotted-dash, iteration 23) only converges to one lobe of the attractor rather than the truth (thick-dashed) that covers both lobes. This is illustrated in Fig. 7 which plots the cost function as a function of x_1 for fixed values of x_2 and x_3 taken from the 10th analysis step for the 20 000-step (solid) 10 000-step (dashed) assimilation windows. The observational interval is 1000 steps and the time filter coefficient $\epsilon = 0.005$. Figure 7 shows that the lobes of the attractor define some local stationary points (minima) of the cost function, which implies that the model has a strong nonlinearity. The solid curve (20 000-step assimilation window case) in Fig. 7 shows that near the global minimum ($x_1 \approx -3.5$) the cost function has a local minimum at $x_1 \approx -5.5$. The extra error produced by the one^{LF} during the minimization makes it easier for the cost function to converge to the local minimum instead of the global one. On the other hand, in the two^{LF} since both \mathbf{x}_{t-1} and \mathbf{x}_t are consistently adjusted by \mathbf{x}_t^o , the observations better constrain the model trajectory toward the truth and the chance that the minimization converges to the local minima is reduced greatly. In contrast to panel a, panel b of Fig. 6 shows that after five iterations, the model trajectory (dotted-dash) in the two^{LF} starts to cover both lobes of the attractor and the optimized trajectory (iteration 20, solid) is able to approach to the truth (thick-dashed).

With $\epsilon = 0.005$, a 20 000-step assimilation window, and a 2000-step observational interval, the one^{LF} converges to both lobes of the attractor. However, for all other observational intervals, the optimized model trajectories remain in one lobe of

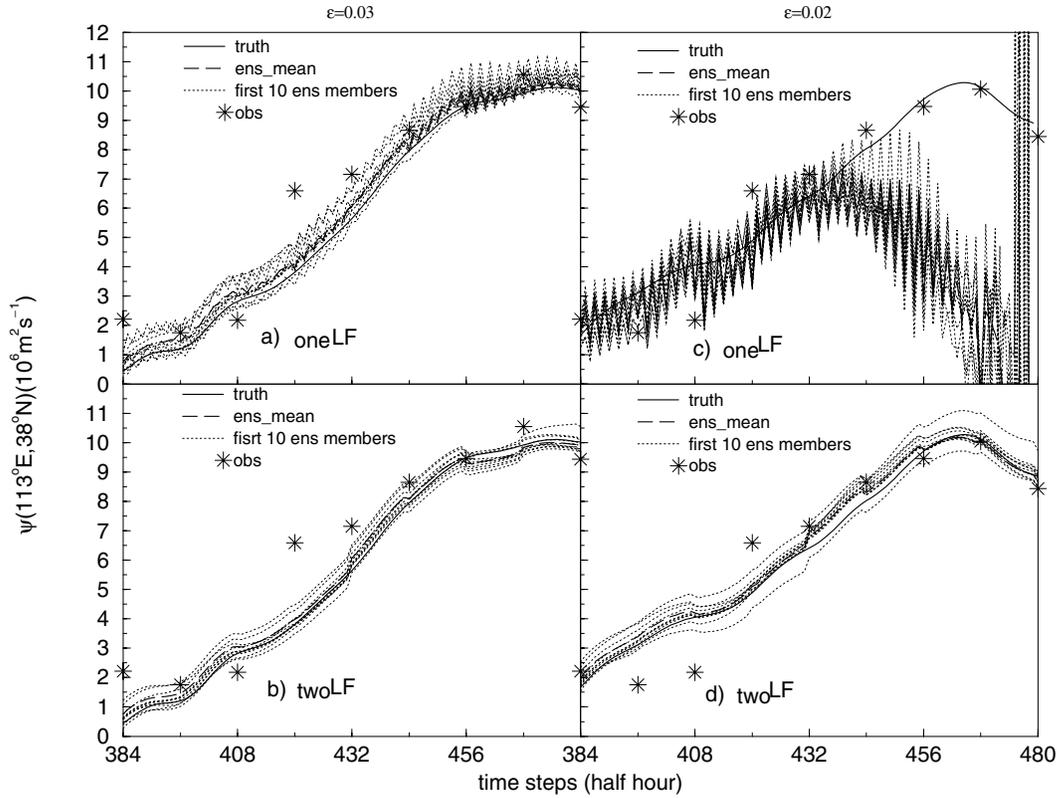


Fig 4. Time series of the first 10 ensemble members (dotted), the ensemble mean (dashed) and the truth (solid) of the streamfunction at $(113^\circ\text{E}, 38^\circ\text{N})$ of the barotropic model from the EAKF assimilation using observations available every 6 h by (a), (c) the one^{LF} (top) and (b), (d) the two^{LF} (bottom) for $\epsilon = 0.03$ (left) and $\epsilon = 0.02$ (right).

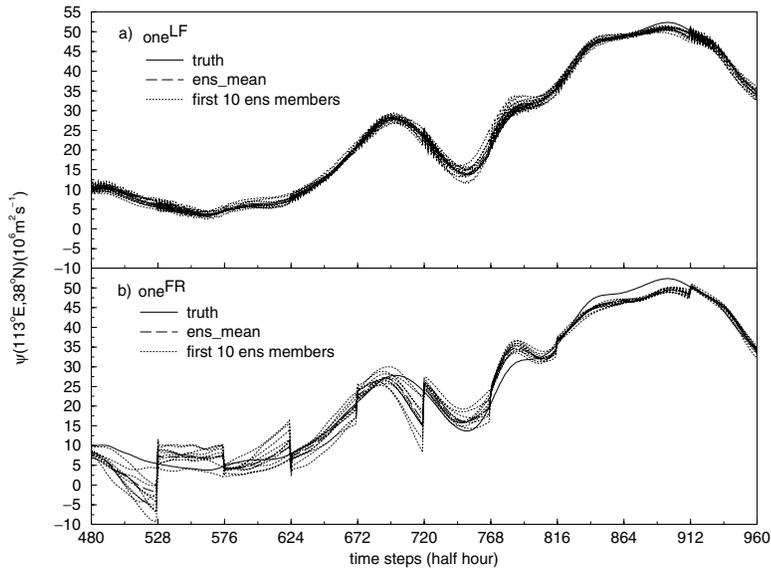


Fig 5. Time series of the first 10 ensemble members (dotted), the ensemble mean (dashed) and the truth (solid) of the streamfunction at $(113^\circ\text{E}, 38^\circ\text{N})$ of the barotropic model in the EAKF assimilation using observations available every 24 h by (a) the one^{LF} and (b) the one^{FR}.

the attractor instead of approaching the truth. The two^{LF} converge to the truth in all cases tested except for a single analysis step with a 5000-step observational interval. Not surprisingly, for $\epsilon = 0.005$ with a 20 000-step assimilation window, the Rms errors for two^{LF} are nearly always much smaller than those for one^{LF}.

For the cases with $\epsilon = 0.005$ and the 10 000-step assimilation window, since the model trajectory resides in one lobe of the attractor, the cost function has a unique minimum as shown by the dashed curve in Fig. 7. The optimized model trajectory in both one^{LF} and two^{LF} is able to converge toward the truth and the error reduction of the Rms error of the two^{LF} is small. For

Table 3. 4D-Var assimilation results in the Lorenz-63 model, implemented by the one^{LF} or two^{LF} adjustment over a 10⁵-step period, for different time filters, assimilation windows and observation intervals

Time filtering coef (ϵ)	Obs interval (steps)	Rms errors of the one ^{LF} /two ^{LF} adjustment	
		10k-step window (reduction %)	20k-step window (reduction %)
0.005	200	0.85/0.81 (5)	2.60/0.78 (70)
	500	1.20/1.10 (8)	3.11/1.05 (66)
	1000	1.28/1.18 (8)	3.30/1.12 (66)
	2000	1.44/1.34 (13)	1.37/1.26 (8)
	5000	1.42/1.25 (12)	2.87/2.02 (30)
0.01	200	0.89/0.84 (7)	0.84/0.80 (5)
	500	1.18/1.08 (8)	1.10/1.02 (7)
	1000	1.30/1.21 (6)	1.18/1.12 (5)
	2000	1.38/1.29 (7)	1.23/1.13 (8)
	5000	1.86/1.70 (9)	1.59/1.52 (5)

$\epsilon = 0.01$ cases, since the computational high-frequency oscillations are efficiently controlled by the stronger time filter, the optimized model trajectories in both one^{LF} and two^{LF} are able to converge toward the truth and therefore the error reduction is small too.

The Rms error of the one^{FR} is intermediate between the two^{LF} and the one^{LF}. When the time stepsize is small and the model trajectory in the assimilation window only covers one lobe of the attractor, the assimilation errors of the one^{FR} are very close to those of the two^{LF}. When the assimilation window is large and the model trajectory in the window covers both lobes of the attractor, as for the one^{LF}, the spin-up error introduced by the one^{FR} increases the chance that the model trajectory converges to one lobe of attractor; this increase the assimilation errors of the one^{FR}. This also explains that when the time stepsize is doubled, the difference of assimilation errors between the one^{FR} and the two^{LF} increases. However, the frequency of these events in the one^{FR} is much smaller than in the one^{LF}. Therefore the mean assimilation errors of the one^{FR} are smaller than the one^{LF}.

4.2. Two^{LF} vs. one^{FR} in the barotropic model

The same ‘‘observations’’ as used for the EAKF assimilation in Section 3.2 are used to carry out the one or two time level adjustment 4D-Var experiments described in Section 2.4. Experiments are conducted for different size assimilation windows and different observational intervals. Owing to the weak nonlinearity of the barotropic model within the assimilation windows (less than 24 h), different time filters in the model do not make a large difference in the assimilation results and the analysis and discussion in this section concentrates on the results of experiments with $\epsilon = 0.02$. For a chosen assimilation window, the maximum variation of the streamfunction in the window is used to rescale the control variable (the streamfunction) and its gradient in order to

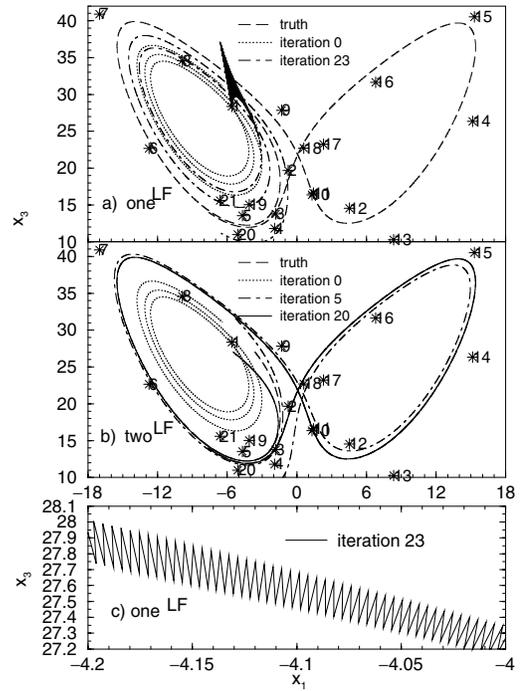


Fig 6. Projection on the (x_1, x_3) plane of the model trajectories at different iterations within the 20 000-step assimilation window in the 4D-Var minimization by (a) the one^{LF} and (b) the two^{LF}, for the 10th analysis step, using observations available every 1000 steps (asterisk). The number beside each asterisk is the observation index in the assimilation window. The black cone in (a) reflects the high-frequency oscillations introduced by one time level adjustment in the leap frog time scheme.

improve the minimization (Zupanski, 1993). After each analysis step is done, the time window is shifted forward to next time at which observations are available for the next analysis step until the whole 10-day analysis period is completed.

With this strong constraint 4D-Var using one^{LF}, the imbalance between the adjusted \mathbf{x}^a_{0-1} and the unadjusted \mathbf{x}_{-1} generated by the line search process becomes so severe that forward integrations of the model eventually blows up.

Table 4 shows that the assimilation errors of the two^{LF} are always smaller than for the one^{FR}. The error reduction made by the two^{LF} from the one^{FR} is not above 16% since in this weakly nonlinear (quasi-linear) situation the minimum (stationary point) of the cost function is unique and the extra spin-up assimilation errors never lead the model trajectory to approach a different regime. However, for the same observational interval, the extra assimilation errors made by the one^{FR} over a small assimilation window are larger than over a large one. These results are consistent with the results of the previous sections, i.e., when more observational data are used, the model trajectory is constrained by the observations promptly so that the introduced inconsistency cannot cause the model trajectory to depart too far from the truth. The error reduction of the two^{LF} from the one^{FR} therefore is relatively small. These results imply that in a realistic

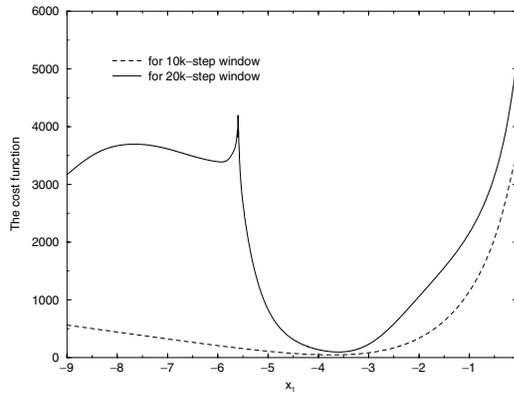


Fig 7. Variation of the cost function defined by the Lorenz-63 model on 10 000-step (dashed) and 20 000-step (solid) assimilation windows in the x_1 space, given x_2 and x_3 as the initial values of the 10th analysis step for the 20 000-step (solid) and 10 000-step (dashed) assimilation windows, with the 1000-step observational interval and an $\epsilon = 0.005$ time filter.

Table 4. 4D-Var assimilation results from the one^{FR} or two^{LF} adjustment using the barotropic model, over a 10-day assimilation period (480 analysis steps) for different assimilation windows and observational intervals

Assim window size	Rms errors ($10^5 \text{ m}^2 \text{ s}^{-1}$) of one ^{FR} /two ^{LF} adjustment (reduction by the two ^{LF} adjustment %)		
	0.5-h obs interval	3-h obs interval	6-h obs interval
0.5 h	10.93/10.12 (7)	–	–
1 h	10.00/9.33 (7)	–	–
3 h	9.29/8.97 (4)	10.10/9.48 (6)	–
6 h	9.11/8.78 (4)	10.13/9.23 (9)	10.98/9.51 (13)
9 h	9.09/8.73 (4)	9.90/9.19 (7)	–
12 h	9.06/8.69 (4)	9.75/9.11 (7)	10.51/9.32 (11)
15 h	9.04/8.67 (4)	9.66/9.06 (6)	–
18 h	9.03/8.65 (4)	9.58/9.03 (6)	10.26/9.20 (10)

4D-Var assimilation with the temporally dense observations, if each iteration restarts from a forward stepping scheme, the difference of assimilation results between the two^{LF} and the one^{FR} may become less important.

In addition, for the same convergence criterion (the norm of the gradient of the cost function decreases by eight orders in magnitude, from 10^2 – 10^3 to 10^{-6} – 10^{-5}), when the width of the assimilation window increases the minimization convergence in both one^{FR} and two^{LF} becomes slow. This occurs because the nonlinearity increases as the width of the assimilation window increases and the hypersurface of the cost function becomes complex. However, for the same assimilation window and observational interval, the minimization of the one^{FR} always converges more slowly than the two^{LF}. This implies that the hypersurface of the cost function with respect to the control variables in the one^{FR} is more complex than in the two^{LF}. The character of the

hypersurface of the cost function may be examined using the eigenvalues of the Hessian matrix, which is beyond the scope of this study.

5. Summary and conclusions

For differencing accuracy, numerical ocean–atmosphere models usually choose a leap frog or an implicit time-differencing scheme to discretize the time tendency of state variables. In this type of time differencing scheme, the derivation of a future state requires multiple time levels, e.g. the present and the previous. Traditional data assimilation implementations only consider the adjustment of the present state using available observations. This one time level adjustment creates extra assimilation error by introducing inconsistency between the adjusted and unadjusted states into the time integration of the assimilation model. For time-dependent assimilation approaches such as ensemble-based filtering, the persistent introduction of this inconsistency may cause computational instability and require extra time filtering to maintain the assimilation.

To prevent this problem, a multiple time level adjustment assimilation scheme was proposed in which the states at times t and $t - 1$, $t - 2$, \dots , if applicable, are adjusted using observations at time t . Given a leap frog time differencing scheme, a low-order (Lorenz-63) model and a global barotropic atmospheric model were used to examine the impact of two time level adjustment on assimilation in a perfect model framework with observing/assimilation simulation experiments. Assimilation algorithms examined include an ensemble adjustment Kalman filter (EAKF, Anderson, 2001) and a four-dimensional variational (4D-Var) method (Le Dimet and Talagrand, 1986).

Results show that the two time level adjustment always reduces the assimilation errors for both filtering and variational algorithms by producing consistent adjusted states at times t and $t - 1$ that are used to produce the future state in the leap frog time-stepping scheme. The magnitude of the error reduction of the two time level adjustment relative to the one time level adjustment varies according to the sparseness of observations, the nonlinearity of the assimilation model and the strength of the model Robert–Asselin time filter. Generally, for sparse observations in time the error reduction by the two time level adjustment is relatively larger than for more frequent observations. Furthermore, when the EAKF is applied in a model with a weak time filter and when 4D-Var is applied with a strongly nonlinear model, the two time level adjustment can significantly improve the assimilation performance.

The 4D-Var implementation of the two time level adjustment described in Section 2.4 assumes a strong correlation between model state variables at adjacent time levels. This is true when the time step used in the model is reasonably small. In addition, Zhang et al. (2001) pointed out that as an implementation of the chain rule, the adjoint integration correctly computes the gradient when the model response is one-sided differentiable (caused by

discontinuities in model parametrizations). The examination of the gradients of cost functions using different width assimilation windows in this study further confirms that the same conclusion holds when the model response is strongly nonlinear (governed by the nonlinearity of the assimilation model, increasing the assimilation window width, for instance). Under this circumstance, the minimization may encounter difficulties (Pires et al. 1996).

In filtering algorithms, two time level adjustment is expected to nearly double the assimilation computational cost since the size of analysed state variables are doubled. In practical implementation, the cost of an assimilation system is usually dominated by the model integration. Therefore, a two time level adjustment may not significantly increase the cost of the entire assimilation procedure. In 4D-Var, the major cost comes from the iteration procedure for minimization and increasing the size of analysed state variables may not significantly increase the assimilation cost either.

It is worth mentioning that although the assimilation algorithms that were examined in this study only include an ensemble-based filter and a 4D-Var method, the conclusions may be used for other assimilation methods such as nudging and optimal interpolation etc. when a leap frog time scheme is used in the assimilation model.

In addition, throughout this study, the assimilation errors produced by the EAKF (Table 1, for instance) are smaller than the errors produced by 4D-Var (Table 3, for instance), which is consistent with a previous study (Anderson, 2001). Generally, a major advantage of the ensemble-based filtering algorithms is the use of temporally varying information about error statistics such as error covariance. Therefore, the relative superiority of the ensemble-based filtering algorithm or four-dimensional variational algorithm may depend on the internal variability of the system (model). Additional research on the relative capabilities of the two algorithm families, including both theoretical foundations and implementation technique, is needed to provide more insight on this issue.

6. Acknowledgments

The authors would like to thank Drs. Paul Kushner, A. Jacobson and C. Dong for their helpful comments on the earlier version of this manuscript. Thanks go to Dr. Eugenia Kalnay for her suggestions that were useful for improving the original manuscript. Thanks also go to an anonymous reviewer for thorough and helpful comments and suggestions.

7. Appendix A. Lorenz-63 model

The Lorenz-63 model (Lorenz, 1963) was derived from the convection equations of (Saltzman, 1962). Owing to the nonperiodic nature of its solutions, the model has become one of the mainstays for the study of the properties of nonlinear systems. The

three equations defining the model are

$$\dot{x}_1 = -\sigma x_1 + \sigma x_2 \quad (\text{A1})$$

$$\dot{x}_2 = -x_1 x_3 + \kappa x_1 - x_2 \quad (\text{A2})$$

$$\dot{x}_3 = x_1 x_2 - b x_3 \quad (\text{A3})$$

where x_1 , x_2 and x_3 are the model state variables and a dot represents the time derivative. The tangent linear version of the model is

$$\delta \dot{x}_1 = -\sigma \delta x_1 + \sigma \delta x_2 \quad (\text{A4})$$

$$\delta \dot{x}_2 = -\delta x_1 x_3 - x_1 \delta x_3 + \kappa \delta x_1 - \delta x_2 \quad (\text{A5})$$

$$\delta \dot{x}_3 = \delta x_1 x_2 + x_1 \delta x_2 - b \delta x_3. \quad (\text{A6})$$

In this study, the standard values 10, 8/3 and 28 for the parameters σ , b and κ are used. The time filter coefficient (ϵ) and the dimensionless time step (Δt) are tuned to $\epsilon = 5 \times 10^{-3}$ and $\Delta t = 10^{-4}$, for which the numerical model has the chaotic characteristics.

8. Appendix B. A global barotropic spectral model

The barotropic model used here is based on the equation of conservation of potential vorticity (Haltiner and Williams, 1980)

$$\frac{d}{dt} \left(\frac{f + \zeta}{H} \right) = 0 \quad (\text{B1})$$

where ζ is the relative vorticity, f is the planetary vorticity (Coriolis parameter) and H is the depth of the atmospheric layer.

Introducing the geostrophic streamfunction $\psi = \frac{g}{f_0} h$ and the Cressman parameter $\lambda^2 = \frac{f_0^2}{g H_0}$, the conservation of potential vorticity (B1) becomes

$$\frac{\partial}{\partial t} (\nabla^2 - \lambda^2) \psi + J(\psi, \nabla^2 \psi) + \beta \frac{\partial \psi}{\partial x} + J(\psi, h') = 0 \quad (\text{B3})$$

where $h' = \frac{f_0}{H_0} h_{\text{terrain}}$, representing the effect of topography.

To code the adjoint of this model, first all nonlinear terms in (B3) are differentiated to develop the tangent linear model (TLM). The tangent linear model governs the evolution of a perturbed streamfunction along the trajectory of the basic state, which can be written as

$$\begin{aligned} \frac{\partial}{\partial t} (\nabla^2 - \lambda^2) \delta \psi + J(\delta \psi, \nabla^2 \psi) + J(\psi, \nabla^2 \delta \psi) \\ + \beta \frac{\partial \delta \psi}{\partial x} + J(\delta \psi, h') = 0. \end{aligned} \quad (\text{B4})$$

In this study, the state variables for the time stepping are spectral coefficients in which a rhomboidal 21 truncation is applied for the transformation between spectral coefficients and grid values. For the assimilation, the state variables are redefined as the 54 (latitude) \times 64 (longitude) Gaussian gridpoints. Except for an initial forward (Euler) step ($\Delta t = 30$ min), a leap frog time step is used to advance the model and a Robert–Asselin time filter

(Robert, 1969; Asselin, 1972) is applied for damping spurious computational modes. The model is tuned to give a stable solution of the streamfunction with $\epsilon \geq 0.02$.

9. Appendix C. Function $\Omega(a, d)$

$$\Omega(a, d) = \begin{cases} -\frac{1}{4} \left(\frac{d}{a}\right)^5 + \frac{1}{2} \left(\frac{d}{a}\right)^4 + \frac{5}{8} \left(\frac{d}{a}\right)^3 - \frac{5}{3} \left(\frac{d}{a}\right)^2, & 0 \leq d \leq a; \\ \frac{1}{12} \left(\frac{d}{a}\right)^5 - \frac{1}{2} \left(\frac{d}{a}\right)^4 + \frac{5}{8} \left(\frac{d}{a}\right)^3 + \frac{5}{3} \left(\frac{d}{a}\right)^2 & a < d \leq 2a; \\ 0, & d > 2a. \end{cases}$$

10. Appendix D. Gradient test

A gradient test was performed to ensure the correctness of the gradient calculated from the adjoint. From the first order approximation of the Taylor expansion of the cost function, one defines a ratio to measure the consistency between the linear increment along the gradient direction of the cost function and a perturbed cost function (Navon et al. 1992) as

$$\left. \begin{aligned} \Phi_1(\alpha) &= \frac{J_1(\mathbf{x}_0 + \alpha \mathbf{e}_0) - J_1(\mathbf{x}_0)}{\alpha \mathbf{e}_0^T \nabla|_{\mathbf{x}_0} J_1} \\ \Phi_2(\alpha) &= \frac{J_2(\mathbf{x}_{-1} + \alpha \mathbf{e}_{-1}, \mathbf{x}_0 + \alpha \mathbf{e}_0) - J_2(\mathbf{x}_{-1}, \mathbf{x}_0)}{\alpha \mathbf{e}_{-1}^T \nabla|_{\mathbf{x}_{-1}} J_2 + \alpha \mathbf{e}_0^T \nabla|_{\mathbf{x}_0} J_2} \end{aligned} \right\} = 1 + O(\alpha)$$

where α is a small scalar governing the magnitude of perturbations and \mathbf{e}_0 and \mathbf{e}_{-1} are unit vectors along the gradient. For the barotropic model case, the perturbation vector is set to be $\mathbf{e}_0 = \nabla|_{\mathbf{x}_0} J_1(J_2) \|\nabla|_{\mathbf{x}_0} J_1(J_2)\|^{-1} \times 10^6 \text{ m}^2 \text{ s}^{-1}$ and $\mathbf{e}_{-1} = \nabla|_{\mathbf{x}_{-1}} J_2 \|\nabla|_{\mathbf{x}_{-1}} J_2\|^{-1} \times 10^6 \text{ m}^2 \text{ s}^{-1}$. (C1) shows that for the correct gradient vectors $\nabla|_{\mathbf{x}_0} J_1(J_2)$ and $\nabla|_{\mathbf{x}_{-1}} J_2$, $\Phi_1(\alpha)$ and $\Phi_2(\alpha)$ go to 1 as α is small but not a machine zero. Figure D1 exhibits the curves of the logarithm of $\Phi_1(\alpha) - 1$ and $\Phi_2(\alpha) - 1$ with respect to $\log_{10}\alpha$ using the Lorenz-63 model (panel a) and the barotropic model (panel b) for the assimilation windows with different width, in which the observations described in Section 2.2 are used to compute the cost functions. From Fig. D1, it is found that the values of $\Phi_1(\alpha)$ and $\Phi_2(\alpha)$ are close to a unity for a small but non-machine-zero α . Particularly, when the values of α fall between 10^{-7} and 10^{-1} , the curves of the logarithm of $\Phi_1(\alpha) - 1$ and $\Phi_2(\alpha) - 1$ versus $\log_{10}\alpha$ are nearly straight lines for 10^4 -, 2×10^4 -step assimilation windows for the Lorenz-63 model (panel a) and 6- and 12-h assimilation windows for the barotropic model (panel b). The long-dashed line in Fig. D1a (the curve of $\log_{10}[\Phi_2(\alpha) - 1]$ versus $\log_{10}\alpha$ for the Lorenz-63 model using a 10^5 -step assimilation window) shows that when the width of the assimilation window increases, the $\log_{10}[\Phi_2(\alpha) - 1]$ curve still has a straight line section but the values of α falling in

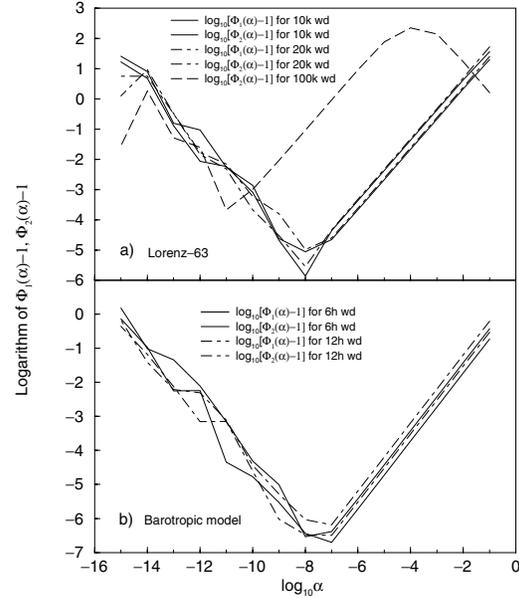


Fig D1. The variation of the logarithm of $\Phi_1(\alpha) - 1$ and $\Phi_2(\alpha) - 1$ with respect to the logarithm of α for the gradient test for (a) the Lorenz-63 model and (b) the barotropic model for different assimilation windows.

this section become smaller (10^{-11} to 10^{-5} in this case). Figure D1 shows that the integrations of these adjoint models correctly evaluates the gradient of the cost functions defined on different assimilation windows since in a certain range of small α values a linear increment along the direction of the gradient evaluated from the adjoint always exactly represents the perturbed cost function.

References

- Anderson, J. L. 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate* **9**, 1518–1530.
- Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.
- Anderson, J. L. 2003. A local least squares framework for ensemble filtering. *Mon. Wea. Rev.* **131**, 634–642.
- Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.
- Asselin, R. 1972. Frequency filter for time integrations. *Mon. Wea. Rev.* **100**, 487–490.
- Bishop, C. H., Etherton, B. J. and Majumdar, S. 2001. Adaptive sampling with the ensemble transform Kalman filter, part I. *Mon. Wea. Rev.* **129**, 420–436.
- Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**, 1719–1724.
- Durrant, D. R. 1999. *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*. Springer, New York, 465 pp.

- Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10 143–10 162.
- Fukumori, I. 2002. A partitioned Kalman filter and smoother. *Mon. Wea. Rev.* **130**, 1370–1383.
- Haltiner, G. J. and Williams, R. T. 1980. *Numerical Prediction and Dynamic Meteorology* 2nd edn. Wiley, New York, 477 pp.
- Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790.
- Hollingsworth, A. and Lönnberg, P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus* **38A**, 111–136.
- Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.
- Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137.
- Ide, K., Courtier, P., Ghil, M. and Lornec, A. 1997. Unified notation for data assimilation. Operational sequential and variational. *J. Meteor. Soc. Japan* **75**, 181–189.
- Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 376 pp.
- Kalnay, E. 2002. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, 341 pp.
- Keppenne, C. L. 2000. Data assimilation into a primitive equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.* **128**, 1971–1981.
- Le Dimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus* **38A**, 97–110.
- Liu, D. C. and Nocedal, J. 1989. On the limited memory BFGS method for large scale optimization. *Math. Program.* **45**, 503–528.
- Lorenz, E. N. 1963. Deterministic non-periodic flow. *J. Atmos. Sci.* **20**, 130–141.
- Miller, R. N., Ghil, M. and Gauthiez, P. 1994. Advanced data assimilation in strongly nonlinear dynamical system. *J. Atmos. Sci.* **51**, 1037–1056.
- Miller, R. N., Carter, E. F. and Blue, S. T. 1999. Data assimilation into nonlinear stochastic models. *Tellus* **51A**, 167–194.
- Mitchell, H. L. and Houtekamer, P. L. 2000. An adaptive ensemble Kalman filter. *Mon. Wea. Rev.* **128**, 416–433.
- Navon, I. M., Zou, X., Derber, J. and Sela, J. 1992. Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.* **120**, 1433–1446.
- Pires, C., Vautard, R. and Talagrand, O. 1996. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus* **48A**, 96–121.
- Robert, A. 1969. The integration of a spectral model of the atmosphere by the implicit method. *Proc. WMO/IUGG Symposium on NWP*. Japan Meteorological Society, Tokyo, Japan, 19–24.
- Saltzman, B. 1962. Finite amplitude free convection as an initial value problem – I. *J. Atmos. Sci.* **19**, 329–341.
- Sirkes, Z. and Tziperman, E. 1997. Finite difference of adjoint or adjoint of finite difference? *Mon. Wea. Rev.* **120**, 3373–3378.
- Thiebaux, H. J. 1985. On approximations to geopotential and wind-field correlation structures. *Tellus* **37A**, 126–131.
- Van Leeuwen, P. J. 1999. Comment on “Data assimilation using an ensemble Kalman filter technique”. *Mon. Wea. Rev.* **127**, 1374–1377.
- Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.
- Zhang, S. and Anderson, J. L. 2003. Impact of spatially and temporally varying estimates of error covariance on assimilation in a simple atmospheric model. *Tellus* **55A**, 126–147.
- Zhang, S., Zhou, X. and Ahlquist, J. E. 2001. Examination of numerical results from tangent linear and adjoint of discontinuous nonlinear models. *Mon. Wea. Rev.* **129**, 2791–2804.
- Zupanski, M. 1993. A preconditioning algorithm for large-scale minimization problems. *Tellus* **45A**, 478–492.