

Four-dimensional ensemble Kalman filtering

By B. R. HUNT¹, E. KALNAY¹, E. J. KOSTELICH², E. OTT¹, D. J. PATIL¹, T. SAUER^{3*},
I. SZUNYOGH¹, J. A. YORKE¹ and A. V. ZIMIN¹, ¹University of Maryland, College Park, MD 20742,
USA; ²Arizona State University, Tempe, AZ 85287, USA; ³George Mason University, Fairfax, VA 22030, USA

Manuscript received 18 September 2003; in final form 2 April 2004

ABSTRACT

Ensemble Kalman filtering was developed as a way to assimilate observed data to track the current state in a computational model. In this paper we show that the ensemble approach makes possible an additional benefit: the timing of observations, whether they occur at the assimilation time or at some earlier or later time, can be effectively accounted for at low computational expense. In the case of linear dynamics, the technique is equivalent to instantaneously assimilating data as they are measured. The results of numerical tests of the technique on a simple model problem are shown.

1. Introduction

Assimilation of observational data is a key component of numerical weather and ocean prediction (Daley, 1991; Kalnay, 2002). The extended Kalman filter provides a fundamental approach to assimilation in the non-linear dynamical setting. A full extended Kalman filter has always been considered computationally unfeasible for numerical weather forecasting due the large number of model variables. The method of ensemble Kalman filters (EnKF;¹ e.g. Evensen, 1994; Evensen and van Leeuwen, 1996; Houtekamer and Mitchell, 1998; Anderson and Anderson, 1999; Hamill and Snyder, 2000; Anderson, 2001; Bishop et al., 2001; Houtekamer and Mitchell, 2001; Ott et al., 2002; Whitaker and Hamill, 2002; Keppenne and Rienecker, 2003; Tippett et al., 2003) have been developed as a means of attacking this problem. In the ensemble-based schemes, an ensemble of initial conditions is evolved by the model and used to estimate the background covariance matrix, a key part of the extended Kalman filter computation.

In this paper we introduce a practical way of unifying the extended Kalman filter and the four-dimensional variational approach, in what we call a four-dimensional ensemble Kalman filter (4DEnKF). Instead of treating observations as if they occur only at assimilation times, we can take exact observation times into account in a natural way, even if they are different from the assimilation times. (Although the simplest approach would be to perform data assimilation at each observation time,

in an operational setting, frequent switching between assimilation and model evolution may be costly and detrimental to the accuracy of the numerical time integration.) The potential that such schemes can be constructed, based on the ensemble Kalman smoother of Evensen and van Leeuwen (2000), was pointed out by Evensen (2003). In our algorithm we use linear combinations of the ensemble trajectories to quantify how well a prospective model state at the assimilation time fits an observation from a different time. This extension of the EnKF to the 4DEnKF can be considered analogous to the extension of the three-dimensional variational technique (3D-Var) to the four-dimensional variational technique (4D-Var; e.g. Talagrand, 1981; Lorenc, 1986). The idea is to infer the linearized model dynamics from the ensemble instead of the tangent-linear map, as done in conventional 4D-Var schemes. Furthermore, in the case of linear dynamics, our technique is equivalent to instantaneous assimilation of measured data.

2. Background

We first describe a version of the ensemble square root Kalman filter approach in the case where the observations are synchronous with the analysis. Consider the continuous dynamical system defined by variables $x = (x_1, \dots, x_M)$ satisfying

$$\dot{x}_m = G_m(x_1, \dots, x_M) \quad \text{for } m = 1, \dots, M. \quad (1)$$

The goal is to track the evolution, under this dynamical system, of an M -dimensional Gaussian distribution centered at $\bar{x}(t)$ (called the most probable state) and with covariance matrix $P(t)$.

To use the EnKF approach $k + 1$ trajectories of eq. (1) are followed in parallel from initial conditions $x^{a(1)}, \dots, x^{a(k+1)}$ over a time interval $[t_a, t_b]$. Because the system is typically high-dimensional, we will assume that $k + 1 \leq M$. The $k + 1$ initial

*Corresponding author.
e-mail: tsauer@gmu.edu

¹Following the naming convention of Evensen (2003), we use EnKF to refer to all ensemble-based Kalman filter schemes, but we note that some other papers use EnKF to refer to the smaller group of stochastic schemes that involve perturbing the observations.

conditions are chosen so that their sample mean and sample covariance are $\bar{x}(t_a)$ and $P(t_a)$, respectively. After running the system over the time interval, we denote the trajectory points at the end of the interval by $x^{b(1)}, \dots, x^{b(k+1)}$, and compute a new sample mean \bar{x}^b and sample covariance P^b from these $k+1$ vectors. If we define the vectors

$$\delta x^{b(i)} = x^{b(i)} - \bar{x}^b$$

and the matrix

$$X^b = \frac{1}{\sqrt{k}} [\delta x^{b(1)} | \dots | \delta x^{b(k+1)}],$$

then

$$\begin{aligned} \bar{x}^b &= \frac{1}{k+1} \sum_{i=1}^{k+1} x^{b(i)} \\ P^b &= X^b (X^b)^T. \end{aligned} \quad (2)$$

Because the sum of the columns of X^b is zero, the maximum possible rank of P^b is k .

At this point, data assimilation is carried out using observations assumed to have been taken at time t_b . The goal is to use the observations to replace the dynamics-generated pair \bar{x}^b, P^b at time t_b with a revised pair \bar{x}^a, P^a that is used as $\bar{x}(t'_a)$ and $P(t'_a)$ on the next time interval $[t'_a, t'_b]$, where $t'_a \equiv t_b$.

Assume (as will hold in the typical case) that the rank of P^b is k . Then the column space S of P^b is k -dimensional, and equals the row space, because P^b is a symmetric matrix. The orthonormal eigenvectors $u^{(1)}, \dots, u^{(k)}$ of P^b , which correspond to non-zero eigenvalues, span this space. Because the variation of the ensemble members occurs in the directions spanning the vector space S , we look there for corrections to \bar{x}^b in the data analysis step. Set $Q = [u^{(1)} | \dots | u^{(k)}]$, the $M \times k$ matrix whose columns form a basis of S . To restrict P^b to the subspace S , define $\hat{P}^b = Q^T P^b Q$.

The details of the data analysis step vary among the different EnKF schemes. We illustrate our technique for the ensemble square root Kalman filter (Tippett et al., 2003, and references therein), although our technique can be applied to any EnKF scheme. The analysis step uses observations (y_1, \dots, y_ℓ) measured at assimilation time t_b , which we assume for simplicity are linearly related to the dynamical state x by $y = Hx$, where H is known as the observation operator. Denote by R the error covariance matrix of the observations. Define $\hat{H} = HQ$ to restrict the action of H to the subspace S . The formula for the solution to recursive weighted least-squares with current solution \bar{x}^b and error covariance matrix \hat{P}^b is

$$\begin{aligned} \hat{P}^a &= \hat{P}^b (I + \hat{H}^T R^{-1} \hat{H} \hat{P}^b)^{-1} \\ \Delta \hat{x} &= \hat{P}^a \hat{H}^T R^{-1} (y - H \bar{x}^b) \\ \bar{x}^a &= \bar{x}^b + Q \Delta \hat{x}. \end{aligned} \quad (3)$$

The corrected most likely solution is \bar{x}^a , with error covariance matrix $P^a = Q \hat{P}^a Q^T$.

To finish the step and prepare for a new step on the next time interval, we must produce a new ensemble of $k+1$ initial conditions $x^{a(1)}, \dots, x^{a(k+1)}$ that have the analysis mean \bar{x}^a and analysis covariance matrix P^a . This can be done in many ways, one of which is finding a matrix of the form

$$X^a = \frac{1}{\sqrt{k}} [\delta x^{a(1)} | \dots | \delta x^{a(k+1)}]$$

that satisfies

$$P^a = X^a (X^a)^T. \quad (4)$$

Defining the vectors

$$x^{a(i)} = \bar{x}^a + \delta x^{a(i)}$$

yields an ensemble with mean \bar{x}^a and covariance matrix P^a . It can be shown (e.g. Tippett et al., 2003) that a good (although not unique) choice is $X^a = X^b Y$ where

$$Y = \{I + (\hat{X}^b)^T (\hat{P}^b)^{-1} (\hat{P}^a - \hat{P}^b) (\hat{P}^b)^{-1} \hat{X}^b\}^{1/2}, \quad (5)$$

where $\hat{X}^b = Q^T X^b$, and the symmetric matrix square root is chosen to be positive.

3. 4D-EnKF method

The description above assumes that the data to be assimilated were observed at the assimilation time t_b . The 4D-EnKF method adapts EnKF to handle observations that have occurred at non-assimilation times. The key idea is that because the analysis takes place in the space S spanned by the ensemble perturbations, we ask which linear combination of the ensemble trajectories best fits the observations at the appropriate times. In this way we avoid the need to linearize the original equations of motion, as is necessary in 4D-Var.

Notice that eqs. (3), (4) and (5) result in analysis vectors $x^{a(1)}, \dots, x^{a(k+1)}$ that lie in the space spanned by the background ensemble $x^{b(1)}, \dots, x^{b(k+1)}$. In order to carry out the analysis, we need only consider model states of the form

$$x_b = \sum_{i=1}^{k+1} w_i x^{b(i)}, \quad (6)$$

and the goal of the analysis is to find the appropriate set of weights $w_1^{(j)}, \dots, w_{k+1}^{(j)}$ for each analysis vector $x^{a(j)}$.

Now let $y = h(x)$ be a particular observation made at time $t_c \neq t_b$. We associate with the state (6) at time t_b a state

$$x_c = \sum_{i=1}^{k+1} w_i x^{c(i)}, \quad (7)$$

where $x^{c(i)}$ is the state of the i th ensemble solution at time t_c . We assign the observation $h(x_c)$ at time t_c to the state x_b given by eq. (6). Equation (7) has been used Bishop et al., 2001; Majumdar et al., 2002) to predict the forecast effects of changes in the analysis error. Here we use this approximation to propagate the

dynamical information within the time window of an analysis procedure designed for real-time numerical weather prediction.

Let

$$E_b = [x^{b(1)} \dots | x^{b(k+1)}]$$

and

$$E_c = [x^{c(1)} \dots | x^{c(k+1)}]$$

be the matrices whose columns are the ensemble members at the times t_b and t_c , respectively. Then eqs. (6) and (7) say that $E_b w = x_b$ and $E_c w = x_c$, respectively, where $w = [w_1, \dots, w_{k+1}]^T$. The orthogonal projection to the column span of E_b is given by the matrix $E_b(E_b^T E_b)^{-1} E_b^T$, meaning that the coefficients w in eq. (6) can be defined by $w = (E_b^T E_b)^{-1} E_b^T x_b$. The linear combination (7) is $x_c = E_c w = E_c (E_b^T E_b)^{-1} E_b^T x_b$. Therefore the observation $h(x_c)$, expressed as a function of the background state x_b at the time of assimilation, is

$$h(E_c w) = h[E_c (E_b^T E_b)^{-1} E_b^T x_b]. \tag{8}$$

The latter expression can be used as the observation operator in any ensemble Kalman filter. For example, a set of observations denoted by the matrix H and time-stamped at t_c can be represented at time t_b by the matrix $HE_c(E_b^T E_b)^{-1} E_b^T$. Therefore, the innovation $y - H\bar{x}_c$ learned from the observations is treated instead as $y - HE_c(E_b^T E_b)^{-1} E_b^T \bar{x}_b$ in the assimilation step. This technique is equivalent to the computation of the forcing of the observational increments at the correct time in 4D-Var; however, it propagates the increments forward or backward in time without the need for the tangent-linear model or adjoint.

Multiple observations are handled in the same manner. Assume the observation matrix is

$$H = \begin{pmatrix} h_1 \\ \vdots \\ h_\ell \end{pmatrix},$$

where the observation row vectors h_1, \dots, h_ℓ correspond to times $t_{c_1}, \dots, t_{c_\ell}$, respectively. Then the observation matrix H in eq. (3) is replaced with the matrix

$$\begin{pmatrix} h_1 E_{c_1} \\ \vdots \\ h_\ell E_{c_\ell} \end{pmatrix} (E_b^T E_b)^{-1} E_b^T. \tag{9}$$

In addition, it should be noted that t_{c_i} can be smaller or larger than t_b , allowing for observations to be used at their correct observational time even if it is after the nominal analysis time.

In the case of linear system dynamics, the 4DEnKF technique is equivalent to assimilating data at the time they is observed. As a simple example, assume an observation at time $t_c < t_b$ is assimilated at time t_c . Let the linear dynamics of the system on the time interval $[t_c, t_b]$ be denoted by the invertible $M \times M$ matrix A . The corrected covariance matrix P^d at time t_c is $P^d = P^c(I + H^T R^{-1} H P^c)^{-1}$ and the new mean is $\bar{x}^d = \bar{x}^c +$

$P^d H^T R^{-1} (y - H\bar{x}^c)$. Advancing both through linear dynamics on the time interval $[t_c, t_b]$ results in covariance

$$A P^d A^T = A P^c (I + H^T R^{-1} H P^c)^{-1} A^T$$

and

$$A \bar{x}^d = A \bar{x}^c + A P^d H^T R^{-1} (y - H\bar{x}^c). \tag{10}$$

On the other hand, using 4DEnKF, the covariance matrix advances without analysis to $A P^c A^T$ and \bar{x}^c to $A \bar{x}^c$ at analysis time t_b . For vectors x_b in the span of the columns of E_b , the matrix $E_b (E_b^T E_b)^{-1} E_b^T$ acts as the identity. Under the linear dynamics, $A E_c = E_b$, so that the revised observation matrix is $H E_c (E_b^T E_b)^{-1} E_b^T = H A^{-1}$. The assimilation formulae then give the new analysis covariance matrix

$$\begin{aligned} P^a &= A P^c A^T [I + (A^{-1})^T H^T R^{-1} H A^{-1} A P^c A^T]^{-1} \\ &= A P^c (I + H^T R^{-1} H P^c)^{-1} A^T = A P^d A^T \end{aligned}$$

and

$$\bar{x}^a = A \bar{x}^c + P^a (A^{-1})^T H^T R^{-1} (y - H\bar{x}^c),$$

which agrees with eq. (10).

4. Experiments with Lorenz model

To illustrate this implementation of four-dimensional assimilation of observations, we use the Lorenz model (Lorenz, 1996)

$$\dot{x}_m = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F \tag{11}$$

for $m = 1, \dots, M$ and with periodic boundary conditions $x_1 = x_{M+1}$. When the forcing parameter is set to $F = 8$ and $M = 40$, the attractor of this system has information dimension approximately 27.1 (Lorenz and Emanuel, 1998). A long integration of the model creates a long background trajectory x^* , to be considered as the true trajectory. Noisy observations are produced at each time interval Δt by adding uncorrelated Gaussian noise with variance 1 to the true state at each location. By comparison, the average root mean square deviation from the mean is approximately 3.61 for the true trajectory. Figure 1 shows a sample trajectory.

Figure 2 shows that, if we use 4DEnKF, assimilations can be skipped with little loss of accuracy in tracking the system state. The system is advanced in steps of size $\Delta t = 0.05$, but instead of assimilating the observations at each step, assimilation is performed only every s steps. For this simulation, 40 ensemble members were used, although the results are little changed when as few as 20 ensemble members were used. The resulting root mean square error (RMSE) is plotted as circles in Fig. 2 as a function of s . For $s \leq 6$, it appears that little accuracy is lost.

The RMSEs of two other methods are shown in Fig. 2 for comparison. The asterisks plotted in Fig. 2 are the RMSE found by using EnKF (eqs. 3–5), allowing s steps of length Δt to elapse between assimilations. Only those observations occurring at the

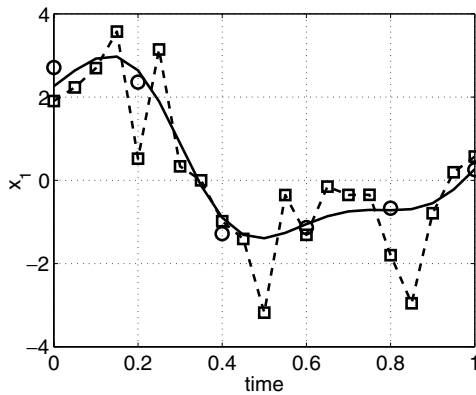


Fig 1. Typical solution of the Lorenz model (solid line) with noisy observations (squares) and analysis state from 4DEnKF (circles). Only the first component x_1 is shown. Observations are generated at each $\Delta t = 0.05$ time units. Assimilation is carried out every four Δt steps using eq. (3) with H replaced by eq. (9).

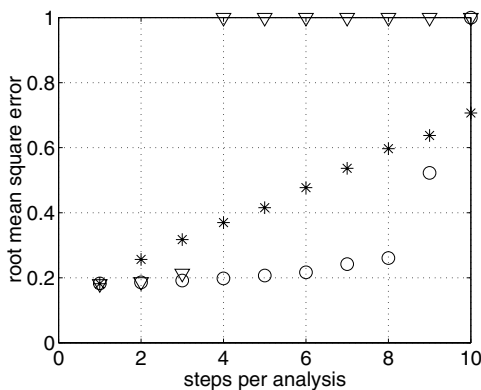


Fig 2. RMSE of proposed 4DEnKF method (circles) compared to standard EnKF (asterisks) and EnKF with time interpolation (triangles). Variance inflation is set at 0.5% per time-step. Symbols showing RMSE = 1 actually represent values ≥ 1 . RMSE is averaged over several runs of 40 000 steps.

assimilation time were used for assimilation. The triangles refer to time interpolation of the data since the last assimilation. In this alternative, linear interpolation of individual observations as a function of the ensemble background state evolved by the model is used to create an improved observation $y_{\Delta}(t_b)$ at the assimilation time, by adding $H(\bar{x}_b) - H(\bar{x}_c)$ to the observation. For the Lorenz example, where the observations are noisy states, this amounts to replacing the observation at time t_c with $y_{\Delta}(t_b) \equiv y(t_c) + \bar{x}_b - \bar{x}_c$ for assimilation at time t_b , which is carried out by EnKF. The idea behind this technique is widely used in operational 3D-Var systems to assimilate asynchronous observations (e.g. Huang et al., 2002; Benjamin et al., 2004). Our implementation provides somewhat optimistic results for this technique, because our background error covariance matrix is not static (independent of time) and homogeneous (independent of location) as it is assumed in a 3D-Var. As Fig. 2 shows,

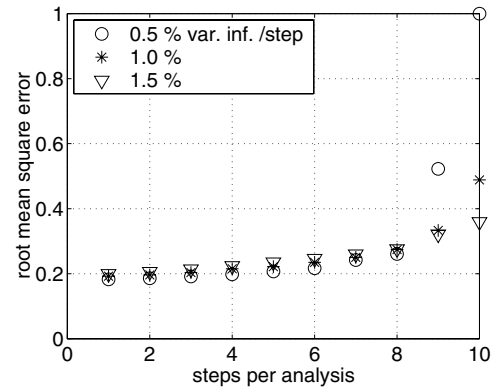


Fig 3. Effect of variance inflation on 4DEnKF. The RMSE of the method is shown for $\epsilon = 0.5\%$ (circles), 1% (asterisks) and 1.5% (triangles) per time-step.

for the last two methods, the accuracy of the assimilated system state becomes considerably worse compared to 4DEnKF as the steps per assimilation s increases.

Variance inflation was used in all experiments described above, meaning that the analysis covariance matrix was artificially inflated by adding $s\epsilon\lambda I$ to \hat{P}^a for small ϵ , where s is the number of time-steps per assimilation and λ is the average of the eigenvalues of \hat{P}^a . In Fig. 2, $\epsilon = 0.005$ per time-step was used for all methods. With $\Delta t = 0.05$, this amounts to additive variance inflation of 0.1 per unit time. Variance inflation helps to compensate for underestimation of the uncertainty in the background state due to non-linearity, limited ensemble size, and model error. (Conventionally, variance inflation is carried out by enlarging \hat{P}^b , rather than \hat{P}^a , in order to reflect errors that develop during model evolution that are not captured by the ensemble. We enlarge \hat{P}^a purely as a matter of convenience, because it does not require us to adjust the ensemble both before and after the data assimilation step. The two approaches yield very similar results.) Figure 3 shows the effect of variance inflation on the RMSE of the 4DEnKF method.

Although the examples shown in this paper describe assimilation using global states, we have achieved similar results by applying this modification to the local ensemble Kalman filter (LEnKF), as developed in Ott et al. (2002). In the local approach, based on the hypothesis that assimilation can be carried out on moderate-size spatial domains and reassembled, the same treatment of the asynchronous local observations can be exploited.

The computational savings possible with the 4DEnKF technique arise from the ability to improve the use of asynchronous observations without more frequent assimilations. The extra computational cost of 4DEnKF is dominated by inverting the $(k+1) \times (k+1)$ matrix $E_b^T E_b$ in eq. (8), which is comparatively small if the ensemble size $k+1$ is small compared to the number of state variables M . Moreover, applying this technique in conjunction with local domains as in LEnKF allows k to be greatly reduced in comparison with M . For very

high-dimensional systems, we expect the best results to be obtained when a combination of the two ideas is used, in what could be called 4DLEnKF.

5. Acknowledgments

The authors thank Jeff Whitaker and two anonymous reviewers for helpful comments. This work was supported by a James S. McDonnell 21st Century Research Award, by the Office of Naval Research (Physics), by the Army Research Office, by the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Integrated Program Office, and by the National Science Foundation (Grants DMS 0104087, DMS 0208092 and PHYS 0098632).

References

- Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.
- Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.
- Benjamin, S. G., Devenyi, D., Weygandt, S. S., Brundage, K. J., Brown, J. M. et al. 2004. An hourly assimilation/forecast cycle: the RUC. *Mon. Wea. Rev.* **132**, 495–518.
- Bishop, C. H., Etherton, B. J. and Majumdar, S. 2001. Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.
- Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, New York.
- Evensen, G. 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10 143–10 162.
- Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.
- Evensen, G. and van Leeuwen, P. J. 1996. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman Filter with a quasi-geostrophic model. *Mon. Wea. Rev.* **124**, 85–96.
- Evensen, G. and van Leeuwen, P. J. 2000. An ensemble Kalman smoother for non-linear dynamics. *Mon. Wea. Rev.* **128**, 1852–1867.
- Hamill, T. M. and Snyder, C. 2000. A hybrid ensemble Kalman Filter-3D variational analysis scheme. *Mon. Wea. Rev.* **128**, 2905–2919.
- 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.
- Huang, X.-Y., Morgensen, K. S. and Yang, X. 2002. First-guess at the appropriate time: the HIRLAM implementation and experiments. In: *Proceedings of the HIRLAM Workshop on Variational Data Assimilation and Remote Sensing*, Helsinki, Finland, 28–43.
- Kalnay, E. 2002. *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, Cambridge, 341 pp.
- Keppenne, C. L. and Rienecker, M. 2003. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. *J. Marine Sys.* **40-1**, 363–380.
- Lorenc, A. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**, 1177–1194.
- Lorenz, E. N. 1996. Predictability: a problem partly solved. In: *Proceedings of the Seminar on Predictability*, Vol. 1. ECMWF, Reading, Berkshire, UK.
- Lorenz, E. N. and Emanuel, K. A. 1998. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**, 399–414.
- Majumdar, S., Bishop, C. H., Etherton, J. and Toth, Z. 2002. Adaptive sampling with the ensemble transform Kalman filter. Part II: field program implementation. *Mon. Wea. Rev.* **130**, 1356–1369.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J. et al. 2002. A local ensemble Kalman filter for atmospheric data assimilation. Preprint (physics/0203058).
- Talagrand, O. 1981. A study of the dynamics of four-dimensional data assimilation. *Tellus* **33**, 43–60.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square-root filters. *Mon. Wea. Rev.* **131**, 1485–1490.
- Whitaker, J. S. and Hamill, T. H. 2002. Ensemble Data Assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.