Meteorologische Zeitschrift, Vol. **16**, No. 6, 795-818 (December 2007)


**Article**

# A review of issues in ensemble-based Kalman filtering

MARTIN EHRENDORFER*

Department of Meteorology, The University of Reading, United Kingdom

(Manuscript received February 28, 2007; in revised form September 17, 2007; accepted October 22, 2007)

**Abstract**

Ensemble-based data assimilation methods related to the fundamental theory of Kalman filtering have been explored in a variety of mostly non-operational data assimilation contexts over the past decade with increasing intensity. While promising properties have been reported, a number of issues that arise in the development and application of ensemble-based data assimilation techniques, such as in the basic form of the ensemble Kalman filter (EnKF), still deserve particular attention. The necessity of employing an ensemble of small size represents a fundamental issue which in turn leads to several related points that must be carefully considered. In particular, the need to correct for sampling noise in the covariance structure estimated from the finite ensemble must be mentioned. Covariance inflation, localization through a Schur/Hadamard product, preventing the occurrence of filter divergence and inbreeding, as well as the loss of dynamical balances, are all issues directly related to the use of small ensemble sizes. Attempts to reduce effectively the sampling error due to small ensembles and at the same time maintaining an ensemble spread that realistically describes error structures have given rise to the development of variants of the basic form of the EnKF. These include, for example, the Ensemble Adjustment Kalman Filter (EAKF), the Ensemble Transform Kalman Filter (ETKF), the Ensemble Square-Root Filter (EnSRF), and the Local Ensemble Kalman Filter (LEKF). Further important considerations within ensemble-based Kalman filtering concern issues such as the treatment of model error, stochastic versus deterministic updating algorithms, the ease of implementation and computational cost, serial processing of observations, avoiding the appearance of undesired dynamic imbalances, and the treatment of non-Gaussianity and nonlinearity. The discussion of the above issues within ensemble-based Kalman filtering forms the central topic of this article, that starts out with a brief overview of Bayesian updating and Kalman filtering theory. The article collects and discusses evidence related to these issues thus assessing also the status of knowledge regarding the performance of ensemble-based Kalman filtering methods.

**Zusammenfassung**

Datenassimilationsmethoden basierend auf Ensembles und der grundlegenden Theorie des Kalmanfilters sind im vergangenen Jahrzehnt in verschiedenen meist nicht-operationellen Problemstellungen mit zunehmender Intensität untersucht worden. Während mehrere Erfolg versprechende Eigenschaften berichtet wurden, verbleibt eine Reihe von Diskussionspunkten in der Entwicklung und Anwendung von Ensemble-basierten Datenassimilationsmethoden, wie sie auch in der grundlegenden Form des Ensemble Kalman Filters (EnKF) gegeben sind, die speziell beachtet werden müssen. Die Notwendigkeit, Ensembles von kleinem Umfang zu verwenden, stellt einen grundlegenden Diskussionspunkt dar, aus dem sich mehrere weitere Punkte ergeben, die sorgfältig beachtet werden müssen. Speziell ist hier die Notwendigkeit zu nennen, den durch das Betrachten der endlichen (kleinen) Stichprobe entstehenden "Lärm" in der geschätzten Kovarianzstruktur zu korrigieren. Die sogenannte Kovarianz-Inflation, die Lokalisierung durch ein Schur/Hadamard Produkt, das Vermeiden des Auftretens von Filterdivergenz, und des sogenannten Inbreeding, sowie des Verlusts dynamischer Balanzen, zählen alle zu den Gesichtspunkten, die direkt mit der geringen Anzahl von Ensemble-Mitgliedern im Zusammenhang stehen. Versuche, den "Lärm" in der geschätzten Kovarianzstruktur effektiv zu reduzieren und gleichzeitig den "Spread" des Ensembles so zu gestalten, dass die Fehlerstrukturen realistisch beschrieben werden, hat zur Entwicklung mehrerer Varianten des EnKF geführt. Hierzu zählen beispielsweise der Ensemble Adjustment Kalman Filter (EAKF), der Ensemble Transform Kalman Filter (ETKF), der Ensemble Square-Root Filter (EnSRF), und der Local Ensemble Kalman Filter (LEKF). Weitere wichtige Aspekte des Ensemble-basierten Kalmanfilters betreffen die Behandlung von Modellfehlern, stochastische und deterministische Aktualisierungsalgorithmen, die Einfachheit von Implementierung und entstehender Rechenaufwand, das serielle Prozessieren von Beobachtungen, das Vermeiden unerwünschter dynamischer Imbalanzen, sowie die Behandlung von Nichtlinearitäten und von nicht Gauß-verteilten Prozessen. Die Diskussion dieser Punkte innerhalb des Ensemble-basierten Kalmanfilters bildet das zentrale Thema dieses Artikels, an dessen Anfang ein kurzer Überblick über Bayes- und Kalmanfilter-Theorie steht. Die Arbeit sichtet und diskutiert Evidenz, die zu diesen Punkten vorliegt, wodurch gleichzeitig der Stand des Wissens zu den Eigenschaften dieser Ensemble-basierten Kalmanfiltermethoden dokumentiert wird.

*Author's address: Martin Ehrendorfer, Department of Meteorology, The University of Reading, P O Box 243, Reading, RG6 6BB, UK, e-mail: m.ehrendorfer@reading.ac.uk

0941-2948/2007/0256 $ 10.80

DOI: 10.1127/0941-2948/2007/0256

# 1 Introduction

Models describing atmospheric flows possess the property that two or more initially slightly different initial states – each evolving according to the same physical laws – do, in general, over time develop into states no more similar than two or more randomly observed states of the atmosphere (see, e.g., LORENZ, 1982, 1984, 1993). This inherent error growth is not an artifact of the models, but is a consequence of the nonlinearity and instability of atmospheric dynamics (LEITH, 1978), as also summarized by the statement "Unless we wish to maintain that the state-of-the-art model at the European Centre, and competitive models at the National Meteorological Center in Washington and other centers, do not really behave like the atmosphere, in spite of the rather good forecasts that they produce at short range, we are more or less forced to conclude that the atmosphere itself is chaotic" (LORENZ, 1993).

Such error growth coupled with the principal inevitability of initial-state errors implies an inherent limit on the predictability of atmospheric possesses. This limited atmospheric predictability manifests itself in an intrinsic uncertainty of weather forecasts made with atmospheric numerical weather prediction (NWP) models. This forecast uncertainty is further increased through errors within the model formulation itself. Increasing artificially the damping processes in the model suppresses the demonstration of chaos, but at the same time amounts to an increased model error.

In addition to the sensitive dependence on present condition (e.g., HOLLOWAY, 1986), the presence of a shallow slope of $-5/3$ in the observed spectrum of the kinetic energy of mesoscale atmospheric motions (see, e.g., HOUGHTON, 2005) puts a fundamental limitation on the predictability of flows at (e.g., ERRICO and BAUMHEFNER, 1987). To some extent such spectra are also reproduced in atmospheric models even when they are based on simplified dynamics and devoid of special forcing (see, e.g., TUNG and ORLANDO, 2003; SEIFERT et al., 2006; BURKHARDT and BECKER, 2006).

Two important implications arise from this behaviour of NWP models, and their treatment is strongly related to the application of statistical methods in the atmospheric sciences (WILKS, 2006). The first implication, forming the topic of this paper, is related to the optimal estimation of the initial state that has to be available in order to start the forward time integration of an NWP model in a prediction process (see, e.g., COURTIER, 1994). Such estimation needs to take into account both observational and a priori information. A related aspect concerns taking additional (targeted) observations in particularly sensitive regions if such can be identified before the forecast is initiated (see also, BERLINER et al., 1999; LANGLAND, 2005a, b).

The second implication is related to atmospheric predictability and concerns the a priori identification of forecast uncertainty. Such forecast-uncertainty prediction can, for example, be carried out through Monte Carlo methods with particular attention to the initial sampling procedure that provides the perturbed initial states to be evolved forward in time (see, e.g., EHRENDORFER and TRIBBIA, 1997; EHRENDORFER, 2006; EHRENDORFER and BECK, 2003; ZUPANSKI et al., 2006; EHRENDORFER, 1997; PALMER, 2000).

In this paper, data assimilation and statistical approaches for initial-state estimation (see, COHN, 1997) within NWP are reviewed, with primary emphasis on Kalman filtering (e.g., GHIL, 1989; DALEY, 1991; CATLIN, 1989; CHUI and CHEN, 1989; 1987; JAZWINSKI, 1970; TARANTOLA, 2005; EVENSEN, 2007; EUBANK, 2006; WUNSCH, 2006; LEWIS et al., 2006). Particular attention is given to methodological implementations of Kalman filtering in operational and experimental NWP contexts that are based on (necessarily) *finite-sized and small ensembles* used to estimate the state and its corresponding covariance properties; these methods are collectively referred to as *ensemble-based Kalman filtering* methods (EVENSEN, 2007; HAMILL, 2006; HOUTEKAMER and MITCHELL, 2005; LORENC, 2003b). [1]

In view of several relatively recent excellent overviews on this subject (e.g., HAMILL, 2006; HOUTEKAMER and MITCHELL, 2005; LORENC, 2003b), the present review is an attempt to revisit additional recent publications on the topic of ensemble-based Kalman filtering, and to discuss some issues (e.g., the work described by COMPO et al., 2006) from a different perspective, also in their relationship to atmospheric predictability. Further, this review attempts to provide a succinct summary of a few selected mathematical results of data assimilation theory when viewed from a Bayesian perspective. It also provides for a brief discussion on the likelihood of degrading the so-called background information when perfectly reasonably and accurate data are assimilated.

In very general terms, NWP models are based on mathematical statements of physical principles such as energy and mass conservation, and balance equations for momentum, as well as for components of, for example, water substance in the atmosphere (see, e.g. FRAEDRICH et al., 2005; KLEIDON et al., 2003; JACOBSON, 2005, for a description of NWP models). The fact that these principles govern atmospheric processes was realized early (e.g., BJERKNES, 1904), and the numerical implementation of these principles forms the basis for today's NWP models (for an historical account, see, EDWARDS, 2000; LYNCH, 2006).

[1] For a fascinating historical account of space-flight related Kalman filtering see, for example, MCGEE and SCHMIDT (1985).

The need for numerical implementation arises from the primary difficulty inherent in the formulation of the above-mentioned physical principles, as they are formulated in terms of *nonlinear* partial differential equations, making general analytical treatment impossible. In a generic form then, an NWP model represents a nonlinear mapping $\mathscr{M}$ between an initial model state $\mathbf{x}_0$ and a model state at a later time $t_1$ denoted as $\mathbf{x}_1$:

$$\mathbf{x}_1 = \mathscr{M}(\mathbf{x}_0)\,. \tag{1.1}$$

Beyond nonlinearity, the complexity of $\mathscr{M}$ and the high dimensionality $n$ of the state (on the order of $10^7$ for today's state-of-the-art models) pose additional challenges not only for NWP in general, but for the data assimilation process in particular. Somewhat in constrast to this high dimensionality, PATIL et al. (2001) report findings that atmospheric dynamics could be locally of low dimensionality. The importance of data assimilation in the process of NWP was already pointed out by BJERKNES (1904) who noted that for successful weather prediction it was necessary to firstly know as accurately as possible the laws according to which the atmosphere evolves, in a form such as Eq. (1.1), and to secondly know with high precision the present state of the atmosphere to be used as the initial condition for the model's forward integration.

## 2 Bayes theorem and Kalman filter

### 2.1 The Kalman filter: Analysis step

Consider the situation that observations $\mathbf{y}$ and an *a priori* estimate $\mathbf{x}^{\mathrm{b}}$ for the state $\mathbf{x}$ to be estimated are available, both with their respective uncertainties. Assuming normal distributions, the conditional probability density function (pdf) of the data $\mathbf{y}$ given the state is taken as:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}) \sim \mathscr{N}(\mathsf{H}\mathbf{x}, \mathsf{R})\,, \tag{2.1}$$

and the prior on the state $\mathbf{x}$ to be estimated is also assumed to be normal:

$$p_{\mathbf{x}}(\mathbf{x}) \sim \mathscr{N}(\mathbf{x}^{\mathrm{b}}, \mathsf{B})\,. \tag{2.2}$$

The mean $\mathbf{x}^{\mathrm{b}}$ of the prior pdf is referred to as background field, with $\mathsf{B}$ denoting the background-error covariance. The error covariance $\mathsf{R}$ of the observations $\mathbf{y}$ includes both observational error (measurement and representativeness), as well as error arising due to deficiencies in the forward operator $\mathsf{H}$ (see, also, e.g., IDE et al., 1997; JANJIĆ and COHN, 2006). Both $\mathsf{B}$ and $\mathsf{R}$ are assumed known.

The operator $\mathsf{H}$, assumed linear in the present discussion, takes the model state $\mathbf{x}$ into observation space for comparison with the data $\mathbf{y}$. The situation described above represents the basic data assimilation problem where an optimal estimate for $\mathbf{x}$ is to be determined on the basis of the information contained in $\mathbf{x}^{\mathrm{b}}$ and $\mathbf{y}$, together with their covariances and/or pdfs. Whereas a detailed discussion of the fundamental data-assimilation related issues is provided in a very comprehensive form by TARANTOLA (2005), COHN (1997), or DALEY (1991) (specifically his sections 2.2 and 4.2), a brief outline of these developments is given below as notational reference for the description of the ensemble-based Kalman Filter (KF).

Using Bayes' Theorem (see, e.g., SIVIA and SKILLING, 2006; WILKS, 2006; WEST and HARRSION, 1989) the posterior pdf for the state given the data is obtained as:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}) \propto p_{\mathbf{y}|\mathbf{x}}(\mathbf{y})\; p_{\mathbf{x}}(\mathbf{x})\,. \tag{2.3}$$

Consequently, the posterior pdf $p_{\mathbf{x}|\mathbf{y}}(\mathbf{x})$ for the state given the data must be of the form:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}) \propto \exp\Big[-\frac{1}{2}\Big((\mathbf{y}-\mathsf{H}\mathbf{x})^{\mathrm{T}}\mathsf{R}^{-1}(\mathbf{y}-\mathsf{H}\mathbf{x}) + (\mathbf{x}-\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\mathrm{b}})\Big)\Big]\,. \tag{2.4}$$

Algebraic manipulation of the exponent in (2.4) reveals that – as expected – the posterior pdf $p_{\mathbf{x}|\mathbf{y}}(\mathbf{x})$ for the state given the data, as written in Eq. (2.4), is multivariate *normal* with mean $\mathbf{x}^{\mathrm{a}}$ and covariance $\mathsf{P}^{\mathrm{a}}$, as given below by Eqs. (2.9) and (2.6), respectively. For the derivation of this result, reference is made to Eq. (B.1) in appendix B, as well as to, for example, WEST and HARRISON (1989):

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}) \sim \mathscr{N}(\mathbf{x}^{\mathrm{a}}, \mathsf{P}^{\mathrm{a}})\,. \tag{2.5}$$

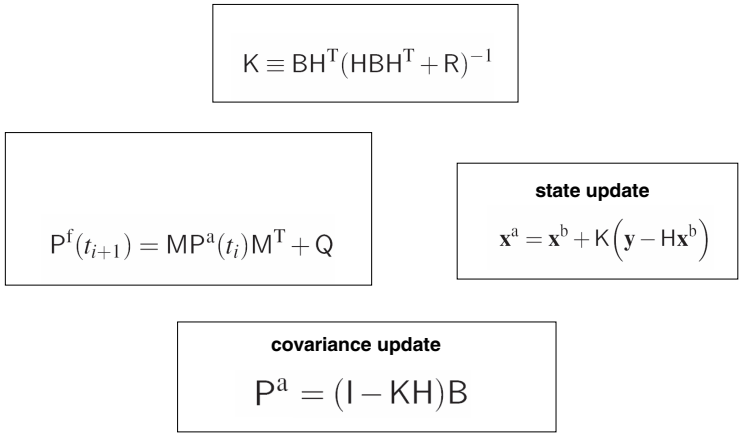
The mean $\mathbf{x}^{\mathrm{a}}$ is denoted as *analysis*, since it represents the Bayesian update of the prior given the (analysed) data, and the covariance $\mathsf{P}^{\mathrm{a}}$ is denoted as the *analysis-error covariance matrix*. Starting from the definition of $\mathsf{P}^{\mathrm{a}}$, as given during the manipulation of the exponent in Eq. (2.4), and spelled out in Eq. (B.1), it is possible to rewrite $\mathsf{P}^{\mathrm{a}}$, the covariance of the posterior pdf, also in the following three equivalent forms:

$$\begin{aligned}\mathsf{P}^{\mathrm{a}} &\equiv \left(\mathsf{B}^{-1}+\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}\right)^{-1} = \\ &= \left(\mathsf{I}-\mathsf{B}\mathsf{H}^{\mathrm{T}}(\mathsf{H}\mathsf{B}\mathsf{H}^{\mathrm{T}}+\mathsf{R})^{-1}\mathsf{H}\right)\mathsf{B} = \\ &= (\mathsf{I}-\mathsf{K}\mathsf{H})\mathsf{B} = \\ &= (\mathsf{I}-\mathsf{K}\mathsf{H})\mathsf{B}(\mathsf{I}-\mathsf{K}\mathsf{H})^{\mathrm{T}}+\mathsf{K}\mathsf{R}\mathsf{K}^{\mathrm{T}}\,,\end{aligned} \tag{2.6}$$

where result Eq. (A.1) has been used to obtain the first equality in Eq. (2.6), and the so-called Kalman gain $\mathsf{K}$ is defined as:

$$\mathsf{K} \equiv \mathsf{B}\mathsf{H}^{\mathrm{T}}(\mathsf{H}\mathsf{B}\mathsf{H}^{\mathrm{T}}+\mathsf{R})^{-1}\,. \tag{2.7}$$

The third equality in Eq. (2.6) represents an updating formulation that will preserve the positive definiteness


$\mathsf{K} \equiv \mathsf{B}\mathsf{H}^{\mathrm{T}}(\mathsf{H}\mathsf{B}\mathsf{H}^{\mathrm{T}} + \mathsf{R})^{-1}$
$\mathsf{P}^{\mathrm{f}}(t_{i+1}) = \mathsf{M}\mathsf{P}^{\mathrm{a}}(t_i)\mathsf{M}^{\mathrm{T}} + \mathsf{Q}$
state update
$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathsf{K}\left(\mathbf{y} - \mathsf{H}\mathbf{x}^{\mathrm{b}}\right)$
covariance update
$\mathsf{P}^{\mathrm{a}} = (\mathsf{I} - \mathsf{K}\mathsf{H})\mathsf{B}$

available) $\mathbf{x}^a$ (see also, EHRENDORFER and TRIBBIA, 1997) and $\mathsf{Q}$ is included here to account for the possibility of the so-called *model error* (see also, Eq. (3.8), as well as TRÉMOLET, 2007b). COHN (1993) points out that in stochastic-dynamic prediction theory, Eq. (2.11) should be considered as an equation for the expectation, just as (2.12) is, and, as such, Eq. (2.11) is missing a coupling term to the predictive covariance equation (2.12) (see, also EPSTEIN, 1969). For a further discussion of Eqs. (2.11) and (2.12), also with regard to the use of $\mathscr{M}$ versus $\mathsf{M}$, reference is again made to HAMILL (2006) and HOUTEKAMER and MITCHELL (2005) (their Eqs. (6.6) and (3), (4), respectively).

### 2.3 The Kalman filter: Implementation

The set of equations given by analysis equations (2.10) and (2.6), together with Eq. (2.7), and the prediction equations (2.11) and (2.12) – to be computed sequentially, as also indicated in Fig. 1 – are referred to as the KF equations (see KALMAN, 1960) in the atmospheric sciences data assimilation literature (e.g., GHIL et al., 1981; COURTIER, 1997a; COHN, 1997; TIPPETT et al., 2003; HOUTEKAMER and MITCHELL, 2005; HAMILL, 2006). A primary difficulty disallowing applying the KF equations directly in this form in realistic NWP and data assimilation contexts is the fact that the matrices involved are intractably large due to the large dimensionality $n$ of the state $\mathbf{x}$ (see, end of section **??**). Further, in operational contexts the number of observations contained in $\mathbf{y}$ is on the order of several million per analysis time. This difficulty concerns in particular the matrix equations (2.6) and (2.12), as well as the computation of the gain in Eq. (2.7).

Two basic approaches are feasible and presently used that allow for utilizing the Kalman filtering results described above within the operational and high-dimensional context of data assimilation in today's NWP. The first approach, described briefly below, is the variational implementation in which the solution to the Kalman filter equations is being sought through consideration of a cost function (COURTIER, 1994, 1997a, b; COURTIER et al., 1994; ANDERSSON et al., 1998; COURTIER et al., 1998). The second approach, discussed together with various issues in the present paper, employs an ensemble-based methodology (see sections 3, 4, and 5).

One possibility of avoiding the explicit handling of the very large matrices that appear here is based on the recognition that the estimate $\mathbf{x}^a$ given in Eq. (2.10) is, by referring back to Eq. (2.4), obtainable as the solution of the optimization problem that consists of finding the state $\mathbf{x}$ that *minimizes* the following cost function $\mathscr{J}(\mathbf{x})$ which is equivalent to *maximizing* the posterior pdf $p_{\mathbf{x}|\mathbf{y}}$:

$$\mathscr{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{y}-\mathsf{H}\mathbf{x})^{\mathrm{T}}\mathsf{R}^{-1}(\mathbf{y}-\mathsf{H}\mathbf{x}) + \\ + \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^{\mathrm{T}}\mathsf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)\,. \qquad (2.13)$$

It is easily checked that the choice $\mathbf{x}=\mathbf{x}^a$, as given by Eqs. (2.9) (or, Eq. (2.10)) together with (2.7), minimizes $\mathscr{J}$, since the gradient of $\mathscr{J}$, given by:

$$\nabla\mathscr{J}(\mathbf{x}) = \mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}(\mathsf{H}\mathbf{x}-\mathbf{y}) + \mathsf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)\,, \qquad (2.14)$$

vanishes precisely for that choice of $\mathbf{x}$ (for the proof of this statement, see, Eq. (D.1) in appendix D). On this basis, an operationally feasible implementation of the estimation problem is possible (see, COURTIER et al., 1994).

A discussion of least-squares estimation, expressed through the quadratic cost function (2.13) may be found in, for example, SORENSON (1970), or LEWIS et al. (2006). Such implementations, based on least-squares and aimed at minimizing cost functions such as $\mathscr{J}$ in Eq. (2.13) by varying $\mathbf{x}$, called *variational methods* (see, e.g., COURTIER, 1997b, 1994; LE DIMET and TALAGRAND, 1986; LEWIS and DERBER, 1985; TALAGRAND and COURTIER, 1987), are operationally feasible, even for very-high-dimensional state-of-the-art NWP systems, and are presently in use at, for example, the European Centre for Medium-Range Weather Forecasts (ECMWF) (see, e.g., COURTIER et al., 1994; TRÉMOLET, 2007a). The operational feasibility is related to the fact that $\mathbf{x}^a$, given in Eq. (2.9), namely, the mean of the posterior pdf (2.4), may be determined by finding the minimum of $\mathscr{J}$, defined in (2.13), a process much easier to implement than the direct computation of Eqs. (2.9) or, equivalently, (2.10).

It must be mentioned, however, that finding the minimum of $\mathscr{J}$ yields only $\mathbf{x}^a$, and not the associated analysis-error covariance $\mathsf{P}^a$ (see, however, FISHER and COURTIER, 1995; BECK and EHRENDORFER, 2005, as well as section 3). Further, still, all results developed here, and, in particular, the result that $\mathbf{x}^a$ may be determined through minimization of a quadratic cost function are only true in linear-dynamics situations, and normal choices for the likelihood and the prior pdfs (as assumed in Eqs. (2.1) and (2.2)); appropriate extensions, known as extended Kalman filters (EKFs), are required for nonlinear and non-Gaussian situations (see, e.g., EUBANK, 2006; LEWIS et al., 2006). An investigation of the initialization of ensemble-based KFs may be found in, for example, ZUPANSKI et al. (2006).

## 3 Ensemble-based Kalman filter

In its basic form, the ensemble-based Kalman filter, known as *ensemble Kalman filter* (EnKF), uses an ensemble of model forecasts and analyses for estimating

the first and second moments of the prior and the posterior pdfs given in Eqs. (2.2) and EVENSEN (1994), and since then the literature on such ensemble-based data assimilation methodology has grown enormously, with many authors having suggested a considerable number of variants of the EnKF (see, also, ANDERSON, 2003; LORENC, 2003b, as well as section 5), not in the least part since the EnKF possesses various attractive features, such as those discussed in section 3.1. Several recent reviews have summarized work related to such ensemble-based atmospheric data assimilation methods (e.g., HAMILL, 2006; Evensen, 2003, 2007; LORENC, 2003b; HOUTEKAMER and MITCHELL, 2005; NERGER et al., 2005).[2]

In this basic form, an ensemble of $N$ parallel data assimilation cycles is carried out with each of the $N$ members updated to a somewhat different background and set of observations:

$$\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{b}} + \hat{\mathsf{K}}\left(\mathbf{y}_i - \mathscr{H}(\mathbf{x}_i^{\mathrm{b}})\right) , \qquad i = 1,...,N , \tag{3.1}$$

with the perturbed observations:

$$\mathbf{y}_i = \mathbf{y} + \mathbf{y}_i' , \qquad i = 1,...,N , \tag{3.2}$$

and the $(n \times m)$ Kalman gain $\hat{\mathsf{K}}$ appropriate in this situation defined as ($n$ is the dimension of the state, and $m$ represents the dimension of the observation vector $\mathbf{y}$):

$$\hat{\mathsf{K}} = \hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}}(\mathsf{H}\hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}} + \mathsf{R})^{-1} . \tag{3.3}$$

Part of the attractiveness of the EnKF relates to the fact that the gain $\hat{\mathsf{K}}$, that is explicitly needed in Eq. (3.1), can be computed without ever explicitly building the $(n \times n)$ matrix $\hat{\mathsf{P}}^{\mathrm{b}}$ that is orders of magnitude too large to be stored even on the largest presently available computing machines. Defining the $(n \times 1)$ and $(m \times 1)$ vectors, respectively:

$$\overline{\mathbf{x}^{\mathrm{b}}} \equiv \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i^{\mathrm{b}} , \tag{3.4}$$

$$\overline{\mathscr{H}(\mathbf{x}^{\mathrm{b}})} \equiv \frac{1}{N}\sum_{i=1}^{N} \mathscr{H}(\mathbf{x}_i^{\mathrm{b}}) , \tag{3.5}$$

the $(n \times m)$ matrix product $\hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}}$ is formed as the average over outer vector products as (see, MITCHELL and HOUTEKAMER, 2006; HOUTEKAMER and MITCHELL, 2001):

$$\hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}}\right)\left(\mathscr{H}(\mathbf{x}_i^{\mathrm{b}}) - \overline{\mathscr{H}(\mathbf{x}^{\mathrm{b}})}\right)^{\mathrm{T}} . \tag{3.6}$$

[2]Many of the references related to ensemble-based Kalman filtering are available at http://enkf.nersc.no/Publications/.

Similarly, the $(m \times m)$ matrix product $\mathsf{H}\hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}}$ is formed in the EnKF as:

$$\mathsf{H}\hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}} = $$
$$= \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathscr{H}(\mathbf{x}_i^{\mathrm{b}}) - \overline{\mathscr{H}(\mathbf{x}^{\mathrm{b}})}\right)\left(\mathscr{H}(\mathbf{x}_i^{\mathrm{b}}) - \overline{\mathscr{H}(\mathbf{x}^{\mathrm{b}})}\right)^{\mathrm{T}} . \tag{3.7}$$

Here, $\mathscr{H}$ is the nonlinear form of $\mathsf{H}$; see, Eq. (2.1). In addition to (3.7), the analysis-error covariance $\hat{\mathsf{P}}^{\mathrm{a}}$ itself is also estimated from the ensemble (see also, section 5). A more complete summary of the basic EnKF equations may be found in HOUTEKAMER and MITCHELL (2005), or LORENC (2003a), and an algorithmic description of these basic forms is given by both EVENSEN (2003) and HAMILL (2006).

## 3.1 Attractive features of the EnKF

### 3.1.1 Statistical and physical nature

The EnKF represents a fundamentally statistically- and physically-based approach to the data assimilation problem. Use of ensembles to estimate statistics required, properly reflects the physical nature of the underlying system, the physics of which is captured and modeled as completely as possible through Eq. (1.1), or its model-error extended counterpart (3.8). Thus, in the EnKF observational (see, Eq. (3.2)) and model errors may be included in a straightforward fashion, somewhat in contrast to the purely variational data assimilation approaches (but, see, also, e.g., TRÉMOLET, 2007b).

### 3.1.2 Parallel nature

It is apparent from Eq. (3.4) in connection with Eq. (2.11) that the forward propagation of the $N$ ensemble members is "embarrassingly parallel" (see, p. 810 of HOUTEKAMER and MITCHELL, 1998). This claim about parallelism of the EnKF should make this step highly suitable for implementation on parallel computers (see, also LORENC, 2003b). Also, ingesting the observations, through the process described by Eq. (3.1), can be carried out in a completely parallel computation, assuming that the nonlinear operator $\mathscr{H}$, together with its linearized counterpart $\mathsf{H}$, are the same for each of the $N$ ensemble members; this condition is satisfied if observational network and observation-error characteristics are the same for each member.

However, it is apparent from Eqs. (3.3), (3.6), and (3.7) that the computation of the gain $\hat{\mathsf{K}}$ requires storage of vectors of length $n$, as well as of $(n \times m)$ and $(m \times m)$ matrices. As long as $m \ll n$, such matrices may be stored on presently available computing machines. If, however, $m \sim O(n)$, or $m \blacktriangleright\blacktriangleright n$, it may become a necessity to process batches of observations *serially* (see, e.g., ANDERSON, 2003; HOUTEKAMER and MITCHELL, 2001;

LORENC, 2003b, as well as section 4.7 and appendix E) for each one of the $N$ ensemble members. In that situation the parallel nature of the algorithm is disrupted because batch *after* batch must be processed, each one containing only such observations that are uncorrelated in terms of errors with all other observations; here, the resulting analysis obtained by processing one batch is the input for processing the next batch. However, in complete accordance with Bayes' theorem (see appendix Eand section 4.7), the order in which these *uncorrelated* batches are taken in is completely irrelevant (see, also ANDERSON, 2003). Thus, whenever such serial observation processing is being used, important aspects of the parallel nature of the ensemble-based assimilation algorithm are lost (see also, LORENC, 2003b).

#### 3.1.3 Nonlinearity

As indicated through the use of the symbols $\mathscr{M}$ and $\mathscr{H}$ (in contrast to their linear counterparts $\mathsf{M}$ and $\mathsf{H}$), the EnKF allows – in principle – for incorporating nonlinearities present in the assimilation process. However, it is unclear to what extent such nonlinear information will be reflected accurately enough in small ensembles to stand out against sampling noise. Also, essentially, by using the basic KF or EKF equations in the analysis and forecast step, the algorithm fundamentally remains a least-squares approach (SORENSON, 1970; LEWIS et al., 2006).

In strongly nonlinear situations, however (e.g., MILLER et al., 1999; Kim et al., 2003), this basic least-squares foundation of the KF may have to be abandoned. Then, alternative approaches must be considered, such as particle filters (e.g., PHAM, 2001; VAN LEEUWEN, 2003; XIONG et al., 2006; SMITH, 2007) that also allow for an efficient treatment of non-Gaussian pdfs that occur in conjunction with strongly nonlinear model dynamics.

#### 3.1.4 Model error

By augmenting Eq. (1.1), or, equivalently Eq. (2.11), with an appropriate model error term $\eta$ (e.g., TRÉMOLET, 2007b) to read:

$$\mathbf{x}^{\mathrm{b}}(t_{i+1}) = \mathscr{M}(\mathbf{x}^{\mathrm{a}}(t_i)) + \eta(t_{i+1}), \qquad (3.8)$$

the EnKF offers enhanced flexibility for dealing with model error, since the effects of such an augmentation will then automatically be included in the set of background forecasts $\mathbf{x}_i^{\mathrm{b}}$ at the next analysis time and as such be brought into the computation of the gain as well. Also, as pointed out by HOUTEKAMER and MITCHELL (2005), uncertainty caused through model error may also be incorporated through the consideration of sampling from different model versions, too. For selected examples of investigations of the impact of model error in state estimation and ensemble prediction reference is made to HANSEN (2002), ORRELL et al. (2001), and ORRELL (2003, 2005), as well as to the discussion in section 4.6.

#### 3.1.5 Dynamic background error covariances

In principle, the EnKF offers the possibility of maintaining and working with completely flow-dependent (i.e., dynamic) analysis-error and background-error covariances, as it does not in its basic form rely on restrictions or assumptions currently made when modeling the background error covariance matrix (e.g., DERBER and BOUTTIER, 1999; COURTIER et al., 1998; RABIER et al., 1998). However, as pointed out above in the context of subsection 3.1.3, and, in particular, in reference to problems with small ensemble sizes and localization (see, section 4), it will remain a difficult challenge to distinguish between structures reflecting sampling noise and truly physical, flow-dependent, and anisotropic features in the error statistics.

In addition, based on the experience obtained at the Meteorological Service of Canada, when assimilating real observations on the basis of realistic operational NWP models (see HOUTEKAMER et al., 2005), it has been found necessary to supplement the flow-dependent covariance structures, estimated from the ensemble, by an additional term. Addition of such a term becomes necessary as the ensemble mischaracterizes the background-error structure and as simple inflation (see, section 4.6) is insufficient to correct for this problem. In their study, HOUTEKAMER et al. (2005) resorted to using a fraction of the static background-error covariance statistics in a form as presently used in variational data assimilation (see, also MITCHELL and HOUTEKAMER, 2000), primarily though due to its easy accessibility, and not for theoretical reasons. The need for adding such a term also indicates that some of the flow-dependent error-covariance features as represented by the ensemble are unphysical structures, possibly due to sampling noise.

## 4 Specific Issues

In this section, various issues related to ensemble Kalman filtering are discussed. In this discussion, a distinction is made between issues that are related to the finite and small ensemble size $N$, as compared to issues that are not. In an ideal situation, where increasingly larger ensemble sizes were possible, the former issues would simply disappear. Since, however, such ideal conditions are nonexistent, it is very appropriate to refer to the difficulties related to finite and *small ensemble sizes* as a fundamental problem (e.g. ANDERSON, 2001), with the consequence that attempts and methods designed to partially resolve such difficulties deserve full attention.

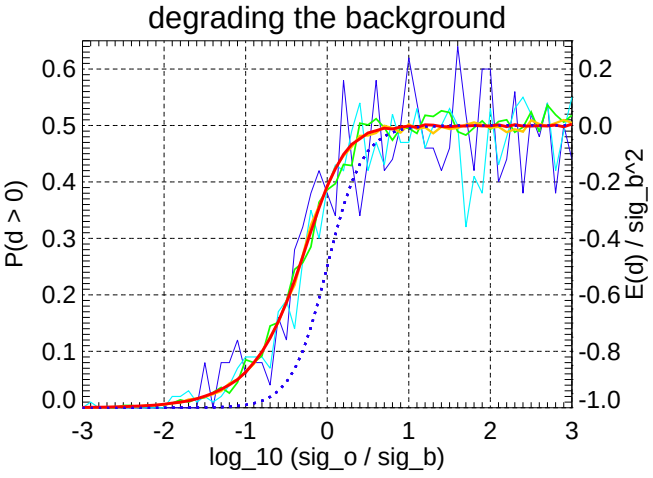



**Figure 2:** The probability for degrading the background $\mathbf{x}^b$ through analysing observations of varying accuracy, as parameterized through the ratio $\sigma_o/\sigma_b$ (shown in terms of its logarithm to the base ten on the abszissa), as obtained through Monte Carlo sampling of Eq. (F.8), with sample sizes of 50 (magenta), $10^2$ (blue), $10^3$ (green), $10^4$ (orange), and $10^5$ (red). The dotted blue curve shows the expectation of the quantity $d \equiv x_a^2 - x_b^2$ (see Eq. (F.8)), normalized by $\sigma_b^2$, with the ordinate to the right applying, again as a function of varying observation accuracy, as given in Eqs. (F.9) and (F.10).

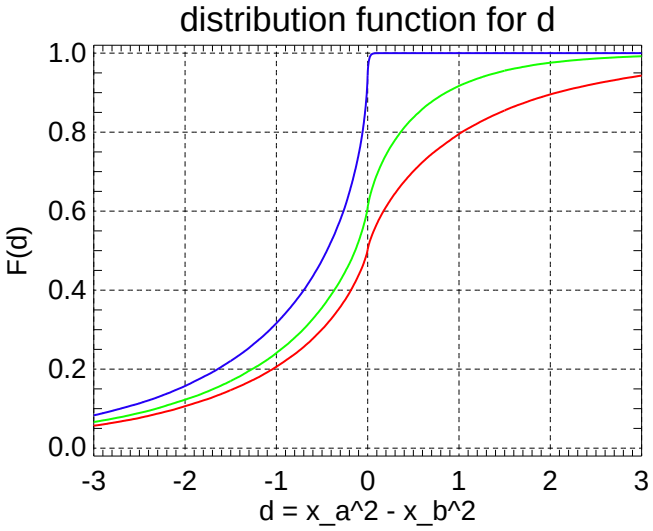



**Figure 3:** The distribution function $F(d)$ assessed empirically from the experiment with sample size equal to $10^5$ of Fig. 2 for $\sigma_o/\sigma_b = 10^{-1}$ (blue curve), $\sigma_o/\sigma_b = 1$ (green curve), and $\sigma_o/\sigma_b = 10^2$ (red curve). Note that the (theoretical) means of these distribution functions may be read off from the blue dotted curve in Fig. 2 as being approximately equal to $-1$, $-0.5$, and zero, respectively.

## 4.1 The accuracy of analyses

In this section the following question is considered: is it possible that the background field is degraded in terms of its quality through assimilating any given observation, even when it is assumed that all observational and background error statistics are correctly specified? In other words, is it possible that the analysis is further away from the truth than the presently available background?

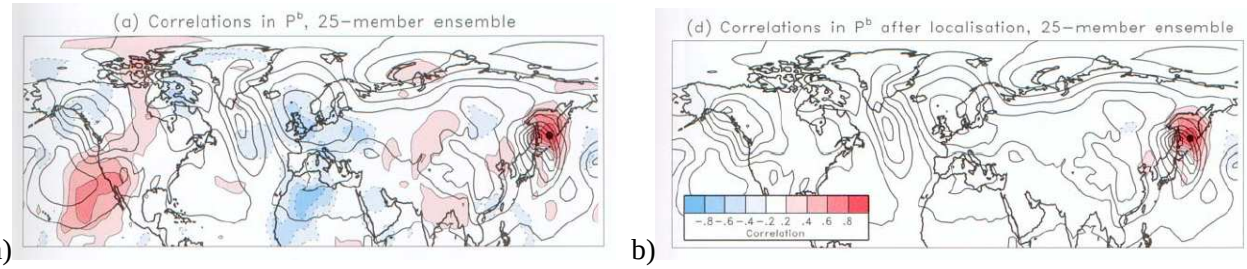
The answer to this question is affirmative as shown in the discussion below that is based on the one-dimensional formulation of the KF analysis equation. While it is true that through the assimilation of an observation the analysis-error variance is always smaller than the background-error variance (see, Eq. (2.6)), this fact does not imply that analysed field is closer to the true field on a case-to-case basis. There is in fact a certain non-zero probability (shown in Fig. 2, see below) that the analysed field is further away from the true field than the background is which translates into a degradation of the background field. The fundamental reason for seeing this behavior is that the analysed field exhibits variability, according to its analysis-error variance; or, stated differently, it is the analysis-error *variance* that is smaller than the background-error *variance*, and not the accuracy of any individual field on a case-to-case basis.

Apart from the desire to provide a clear and unambiguous answer to this often-raised question, at least in a one-dimensional context, the discussion of the above issue is included here to illustrate that sampling issues, namely the relationship between single realizations and the statistics of these realizations, play a role in its investigation, too. However, since the sampling involved here is now from the observation-error distribution, these latter sampling issues are not directly related to the finite small-sample-size issues in ensemble-based Kalman filtering forming the primary topic of this review.

For a quantitative discussion of the above-stated question, the quantity $d$ is defined as the difference between the analysis error and the background error (see, Eq. (F.8)). Then, whenever the situation $d > 0$ occurs, the background field has been degraded. The relationship of an individual realization of $d$ to a realization of an observation $y$ and a background field realization $x_b$, is discussed in detail in appendix F for the situation of the analysis step in a one-dimensional KF. In this discussion, the analysis step only is considered, without the subsequent implication on forecast errors. Further, in this purely static experiment, all statistics are assumed to be known correctly, and are not derived from cycling experiments.

As $d$ is a nonlinear function of other random variables, it is difficult to assess its properties analytically. Therefore, the chance for the event $d > 0$ is evaluated numerically through a Monte Carlo process, essentially by randomly generating realizations of $x_a$ and $x_b$ from their respective pdfs (see Eqs. (F.3) and (F.4)).

The result of these computations is shown in Figure 2 in terms of the (frequentist) probability for $d > 0$ as a function of $\sigma_o/\sigma_b$ (i.e., the ratio of observational-error and background-error standard deviations, respectively; see also appendix F) for various sampling sizes in this Monte Carlo process (where $\sigma_b = 1$ has been ar-




a) b)

**Figure 4:** (a) Correlations of sea-level pressure directly estimated from a 25-member data assimilation ensemble with the pressure at a point in the western Pacific (color shading; positive correlations red, negative correlations blue). The solid lines denote the ensemble mean background sea-level pressure with a contour interval of 8 hPa on 0000 UTC 14 December 2001. The domain shown is the entire northern hemisphere (including some of the southern hemisphere) with the zero meridian located near the middle of the picture. (b) Same as (a), but with correlation estimate from 25-member ensemble after application of covariance localization (taken from Fig. 6.4 of HAMILL, 2006).

bitrarily set). It is, for example, apparent that when $\sigma_o$ is only 10 % of $\sigma_b$, there is still a chance of about 6 % that the analysis is less accurate than the background (notice the variability apparent in assessing that fraction for the five different sample sizes used). If the two variances are equal, that probability goes up to about 39 % in the present one-dimensional example. Interestingly, if the observations are very inaccurate, they deteriorate the background only with a 50% chance, essentially since they then scatter with equal probability around both sides of $x_b$, with, according to Eq. (F.5), $\sigma_a \approx \sigma_b$. In this limiting situation of useless observations (as they are essentially obtainable without any effort on accuracy) 50 % of the cases will still see improved analysis accuracy.

Since the present reasoning carries over to multivariate situations, it is thus apparent that when a sample of background improvements is investigated (e.g., a large number of cases over several months or even years) observation content should be considered physically useful only, if an overall background improvement is found for more than 50% of the cases.

While it is difficult to work out analytically the nature of the pdf for $d$, it is possible to derive analytically (at least in the present one-dimensional situation) its expectation $E(d)$, as given in Eqs. (F.9) and (F.10). Fig. 2 also includes graphically this analytical formulation of $E(d)$ which is, after appropriate normalization, equal to the negative Kalman gain $k$, as shown in Eq. (F.10). Evidently, as necessary, essentially from Eq. (F.2), this curve is symmetric with respect to $\sigma_o/\sigma_b = 1$, and it shows that one would expect an analysis that is improved by an amount of $\sigma_b^2/2$ when $\sigma_o = \sigma_b$, or, for example, of $0.8 \times \sigma_b^2$ when $\sigma_o = 10^{-0.3}\sigma_b$. In the limiting situation of useless observations, one obtains $E(d) = 0$.

Further properties, beyond the expectation, of the distribution function for $d$ are difficult to derive analytically. Thus, to gain additional insight, Fig. 3 shows the empirically assessed – on the basis of the data of the Monte Carlo experiment with sample size equal to $10^5$ used to construct the red curve in Fig. 2 – distribution function $F(d)$ (e.g., DEGROOT, 1986) for three different values of $\sigma_o/\sigma_b$. From Fig. 3 the probabilities for the event $d < 0$ may be readily read off: for example, that probability is a little more than 60 % for $\sigma_o/\sigma_b = 1$, the complement of which matches the value of 39 % mentioned above. It is also apparent that $F(d)$ becomes highly asymmetric (e.g., blue curve in Fig. 3), the more accurate the observational data are. For example, the blue curve (for $\sigma_o/\sigma_b = 0.1$) in Fig. 3 shows that $d$ will be negative almost with certainty, indicating that the analysis is almost certainly improving upon the background. As the observational data become less accurate (red curve), the distribution function becomes almost symmetric around its mean of zero, indicating that it is just as likely to improve as it is to deteriorate the background, as already also indicated in Fig. 2. The mean value of $d$ when computed from these distribution functions takes on a (normalized) value of very close to $-k$ (see, also Fig. 2).

In concluding this discussion, it is reiterated that due to the random nature of the analysis it is possible to deteriorate the background field even on the basis of accurate data, and even in the situation when the relevant statistics are exactly known. In the example discussed above, a chance of about 6% for that event to happen was found for highly accurate observational data (left part of Fig. 2). Such deterioration may occur even though the equation for the analysis-error covariance (2.6) shows that the analysis is *when averaged over many realizations* – or, in other words, in a statistical sense – *always* more accurate than the either the data or the background. Clearly, this deterioration must be carefully considered, whenever the impact of observations on analysis fields is investigated.

It might be added that in real-world applications, whenever data to be assimilated are considered to be more accurate than the background, the degradation issue discussed above is then unlikely to be of large concern. Such accurate data may arise as the result of combining several pieces of information, prior to entering the assimilation algorithm, or from highly accurate in-

struments, such as radiosondes. Further, as mentioned above, it is important (but has not been considered here) for the subsequent forecast step how errors in the analysis project on growing and decaying modes. Obviously, if growing components are generated in the analysis step (see, e.g., JOHNSON et al., 2006), the forecast may end up degraded even though the analysis improves upon the background.

## 4.2 Sampling error in covariance estimates

One of the primary difficulties that appears within ensemble-based approaches to Kalman filtering is the appearance of noise in the forecast-error (co)variances $\hat{\mathsf{P}}^{\mathrm{b}}$ that are estimated from finitely-sized ensembles with small $N$ and that are subsequently used in the computation of the gain (see, e.g., Eq. (3.3)). Sampling theory indicates that this noise in the variances will decrease asymptotically in proportion to $1/N$ and possibly slower for covariances and/or correlations (see, e.g., LORENC, 2003b). Some further discussion of the convergence rates for estimates based on finitely-sized ensembles (also referred to as the "Monte Carlo $N$ barrier") may be found in VAN LEEUWEN (1999) and PRESS et al. (1992) (their section 7.6).

Sampling noise in the estimated covariances is fundamentally inherent in ensemble Kalman filtering and has been addressed by a large number of authors (e.g., HAMILL, 2006; LORENC, 2003b; HOUTEKAMER and MITCHELL, 2001; HAMILL et al., 2001, MITCHELL and HOUTEKAMER, 2000) who suggested various routes to control this difficulty. One of the most obvious effects of sampling error on estimated forecast-error covariances is the appearance of spuriously large correlations between spatially very distant points throughout the physical assimilation domain (see, e.g., Figs. 4(a) and 5(a)). Such spurious correlations lead, when carrying out the analysis step using the gain, to spurious effects on analysed values at grid points very remote from the location of the observation. Through such spurious effects, the analysed ensemble may mischaracterize the analysis error distribution from which further undesirable consequences could arise. One of these consequences could be filter divergence (see section 4.5 below). It is thus evident that the effects of the above-mentioned sampling noise must be avoided through specific measures as much as necessary.

## 4.3 Covariance localization

Various measures have been suggested to reduce the noise in the estimated covariances as described in subsection 4.2. These measures are commonly referred to as *covariance localization* methods in view of the fact that the noisy covariances are much less local than would be expected on physical grounds. Covariance localization is thus a heuristic attempt to modify the model of background-error covariances so that a limited-size ensemble will not represent distant, distinct features as dynamically interrelated when in fact this long-distance dynamical coupling is simply due to noise that arises due to the small sample size (HAMILL, 2006).

A commonly used approach (e.g., LORENC, 2003b; HOUTEKAMER and Mitchell, 2001; HAMILL et al., 2001) consists of modifying the covariances with the aim to make them local by building the Schur product[3] of the sample estimate $\hat{\mathsf{P}}^{\mathrm{b}}$ with a compactly supported correlation function such as described, for example, in

a)

LORENC (2003b) provides an example where through the Schur product localization the wind increments are too small to balance the height gradient. MITCHELL et al. (2002) report from experimentation within a primitive-equation model context that severe localization causes imbalance that may, however, be reduced, by relaxing the localization. These results illustrate the difficulties inherent in ensemble Kalman filtering and that are ultimately related to the fact that ensemble sizes affordable in realistic assimilation contexts are relatively small. In fact, ANDERSON (2001) refers to the fact that sample sizes of practical ensembles are far too small for obtaining meaningful statistics within many of the variants of ensemble Kalman filtering as a *fundamental problem* inherent to these approaches.

### 4.5 Filter divergence and ensemble spread

The problem of filter divergence is discussed at length in HAMILL et al. (2001) (see, also, e.g., HAMILL, 2006; MAYBECK, 1979). As pointed out by HAMILL (2006), the basic problem of filter divergence occurs in the situation that analysis errors, represented through $\mathbf{P}^a$, are underestimated during the cycling process (see, Fig. 1). As shown by VAN LEEUWEN (1999), such underestimation of the total analysis error variance is a systematic effect in the EnKF due to the finite and small ensemble sizes used. This bias in the EnKF must be distinguished from the effects on $\hat{\mathbf{P}}^a$ occurring when observations are ignored which in turn implies that $\hat{\mathbf{P}}^a \approx \hat{\mathbf{P}}^b$ (see beginning of section 3). While the former effect (i.e., the systematic bias-effect described by VAN LEEUWEN, 1999) will lead to disregarding observations, the latter effect may arise for other reasons and may conceivably reflect correct behaviour of the filter.

This underestimation, in turn, implies that in the prediction step, forecast errors are underestimated, too, in the following-up cycle (see, e.g., Fig. 1), with the consequence that the background is considered to be – unrealistically – too accurate, in turn underweighting the new observations. Observations are then consequently more or less ignored, and if the process continues to feed back on itself, observational data are successively less and less used and the background is more and more used. Thus, during the subsequent assimilation cycle, the variance-deficient ensemble will provide another set of underestimated background-error statistics, discarding even more the influence of new observations (see, HAMILL et al., 2001; SORENSON, 1970). The problem may worsen progressively, eventually leading to filter divergence. The result of filter divergence is a useless ensemble in the sense that its mean is far distant from the true state; here far is measured by the ensemble covariances.

It is evident that it is critical to model the forecast error covariances $\mathbf{P}^b$ realistically, especially in terms of its magnitude, but also in terms of its spatial structure (e.g., BUEHNER, 2005). This requirement is equivalent to the statement made by LORENC (2003b) that "the major design problem is to keep the ensemble covariance matching that of the error", or the statement found in HAMILL (2006) that "one of the most crucial preventatives (to avoid filter divergence) is to model background-error covariances realistically". Further, meeting these requirements is also equivalent to maintaining an *ensemble spread* that realistically describes analysis and background error structures. Experience (pers. comm. J. KEPERT 2007) has shown that the ensemble mean can be rather sensitive to even moderate misspecifications of the spread. A recent discussion of quantitative spread formulations, as well as the use of ensemble-spread information in observation-impact studies may be found in TAN et al. (2007). A further discussion of verification of probabilistic forecasts is found in CANDILLE and TALAGRAND (2005).

It is further evident that filter divergence is also a sampling-size related issue (see also the comments above related to the study of VAN LEEUWEN, 1999), since sampling errors will tend to be large for small ensembles, thus increasing the chances that estimated background errors are estimated poorly (too small essentially due to the bias-effect described by VAN LEEUWEN, 1999). It is worth mentioning that filter divergence may also occur in situations when the predictive model is considered to be perfect, again due to the noise implied by finitely-sized ensembles.

### 4.6 Inflation, model error, and inbreeding

Various methods have been proposed that are aimed at attempting to avoid filter divergence. Clearly, the covariance localization methods discussed above (see section 4.3) with their purpose of modeling covariances realistically are among them. ANDERSON and ANDERSON (1999) have suggested another method consisting of broadening or inflating the prior distribution, as essentially described here through Eq. (2.2), by multiplying the covariance matrix by a factor slightly greater than one.

This general procedure is known as *covariance inflation*, and may be viewed as a measure complementing localization in order to keep covariances at a realistically large level. Symbolically it is written by HAMILL et al. (2001) as:

$$\mathbf{x}_i^b \leftarrow \gamma(\mathbf{x}_i^b - \overline{\mathbf{x}}^b) + \overline{\mathbf{x}}^b \,, \qquad (4.1)$$

meant to indicate that the deviation of an individual background forecast from the ensemble mean is (possibly just slightly) increased or inflated, before being used in the assimilation process. Eq. (4.1) also suggests that a more appropriate name for this procedure may be *ensemble inflation*, since the procedure amounts to more than just simply multiplying $\hat{\mathbf{P}}^b$ by a factor. Reports on

the increases used in ensemble inflation mention that these have – under some conditions – been taken to be as small as only on the order of a few percent (e.g., $\gamma = 1.01$), although considerably larger values may be necessary depending on the modeling context. Nevertheless, it is interesting to note that such a small amount of inflation appears sufficient to stabilize the filter (e.g., HAMILL et al., 2001; ANDERSON, 2001; WHITAKER and HAMILL, 2002). It might be added that the reasons for this behavior are not entirely clear at this point. Other approaches designed for ensemble inflation beyond Eq. (4.1) have been investigated, too (see, e.g., section 6.4.4 of HAMILL, 2006).

Inflation may also be interpreted as an effort to account for the model error term $\mathsf{Q}$ in Eq. (2.12). The model error (see also section 3.1.4) term reflects the fact that the ensemble, when time-evolved with $\mathscr{M}$, must in general be expected to carry additional uncertainty due to processes missing from the model, as well as processes being handled imperfectly within $\mathscr{M}$, with both of these error sources affecting the way in which observations are assimilated.

Since little is known about model error (see, e.g., TRÉMOLET, 2007b; HANSEN, 2002; ORRELL et al., 2001; ORRELL, 2003, 2005), formulations for $\mathsf{Q}$ are taken to be of a rather simple form. One of the simplest possible approaches to represent the effect of including $\mathsf{Q}$ within the EnKF consists of just using covariance or ensemble inflation (see, however, for example, HAMILL (2006), for a further discussion). Further, RODWELL and PALMER (2007) present a recent discussion on estimating model error through the assessment of model tendencies. In addition, the estimation of systematic errors, or biases, has recently been addressed by AUGIGNÉ et al. (2007) and DEE (2005).

HOUTEKAMER and MITCHELL (1998) refer to the problem of unrealistically small background-error covariances as represented by the ensemble as *inbreeding* which they suggest to avoid not through inflation, but by splitting the ensemble into two ensembles and using covariances from one ensemble to perform the analyses within the other ensemble (see, also HOUTEKAMER and MITCHELL, 2001). Their argument is thus analogous to problems appearing when applying a linear regression formula to the same data set that was used to develop the regression, resulting in an overly optimistic fit to the data. However, in a detailed theoretical analysis of this problem, VAN LEEUWEN (1999) points out that the difficulties encountered by HOUTEKAMER and MITCHELL (1998) are not only due to using a gain obtained from the same ensemble, but also due to small ensemble-size effects on the gain itself. It is thus consequently suggested by VAN LEEUWEN (1999) that the remedy proposed by HOUTEKAMER and MITCHELL (1998) will suffer from the same problem (to a lesser degree though), and that the difficulty they encountered can be effectively ameliorated only through the use of larger ensemble sizes. It is mentioned that such parallel ensemble data assimilation cycles in which the background-error covariances estimated by one ensemble are used to compute the gain utilized to perform the analyses in the other ensemble have also been considered by WHITAKER and HAMILL (2002).

### 4.7 Serial processing of observations

Eqs. (3.6) and (3.7) summarize in a generic form how – on the basis of the $N$-member ensemble – the parts that constitute the $(n \times m)$ gain $\hat{\mathsf{K}}$ are computed without ever forming the $(n \times n)$ matrix $\hat{\mathsf{P}}^{\mathrm{b}}$ explicitly, given $m$ observations. It is apparent from these equations and the matrix dimensions that the $(n \times m)$ matrix $\hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}}$, as well as the $(m \times m)$ matrix $\mathsf{H}\hat{\mathsf{P}}^{\mathrm{b}}\mathsf{H}^{\mathrm{T}}$ will become intractable if $m$ approaches the dimension $n$ of the state $\mathbf{x}$. If, however, observational errors are uncorrelated (or, even independent), and also if there is independence between background and observational error then it is possible to process observations serially, or, essentially one by one (see, e.g., HAMILL, 2006; ANDERSON, 2003; HOUTEKAMER and MITCHELL, 2001; LORENC, 2003b; GELB, 1974). The basis for such *serial processing of observations* is in fact the theory of the KF analysis step, as outlined in section 2.1 (note that there the analogous assumption was made that $\mathbf{x}^{\mathrm{b}}$ and $\mathbf{y}$ errors were independent or at least uncorrelated); for a more quantitative statement of this fact, reference is made to appendix E. If just one observation is assimilated at a time, then $m = 1$ and the matrix inversion needed to obtain $\hat{\mathsf{K}}$ in Eq. (3.3) is trivial.

However, four issues need to be mentioned in the context of serial processing; first, serial processing assumes uncorrelated errors as described above. Second, as pointed out by EHRENDORFER (1992), there appears to be a tradeoff in computing time between processing an entire batch of observations through just one update as compared to processing through many state updates each observation individually (see, his Fig. 16); for a given computing environment an optimal batch size is likely to exist. Third, serial observation processing disrupts the parallel nature of ensemble-based data assimilation since the output from processing one batch is the input for processing the next batch (see also, subsection 3.1.2). Fourth, through the serial processing of data, quality control is impeded in the sense that each observation can only be cross-checked against those already analysed (since these are reflected by the ensemble); the observation cannot be cross-checked though against those observations to be considered in future batches. Thus, in fact, the consideration of quality control and of the schemes available for this purpose is an issue deserving particular attention in the further development of EnKF methods.

### 4.8 Assimilation of Surface Pressure

Recently, ANDERSON et al. (2005) and WHITAKER et al. (2004) have reported results from data assimilation experiments within a primitive-equation model context, where only surface pressure data were used. Employing in this observational situation the ensemble adjustment Kalman filter, abbreviated as EAKF (see, e.g., ANDERSON, 2001; TIPPETT et al., 2003 and section 5, the data assimilation system is able to reconstruct details of the model's state at all levels (ANDERSON et al., 2005).

In a companion paper, COMPO et al. (2006) provide evidence showing that, on the basis of only surface pressure data, the assimilation systems tested are capable of producing 6-hourly useful analyses of the geopotential height of pressure surfaces at every level in the troposphere; here "useful" is defined by the criterion that errors have to be smaller than the climatological standard deviation (see, Fig. 3 of COMPO et al., 2006). The authors conclude that on the basis of only surface pressure data the entire extratropical tropospheric circulation could be reanalyzed, using an ensemble-based KF, beginning in the early 20th century, up to the present.

In addition, WHITAKER et al. (2004) show that the accuracy of analyses based on surface pressure data when using the ensemble square-root filter, abbreviated as EnSRF (see, e.g., WHITAKER and HAMILL, 2002; TIPPETT et al., 2003 and section 5), is about as high as the accuracy of today's 60-hour forecasts. Also, the comparison of the EnSRF results against results obtained by other candidate assimilation schemes that are restricted to using static background error covariance specifications suggests that it is primarily the flow-dependence contained in the ensemble-based covariances that enables obtaining analyses of such comparably high quality (see also, subsection 3.1.5).

However, in terms of further assessing the results of these studies, it still remains to investigate the impact of misspecification of observation-error covariances (WHITAKER et al., 2004), as well as of the relatively slow internal error growth characteristic for the assimilating model used here which may distort the assimilation results (see, ANDERSON et al., 2005, as well as next subsection). In addition, it is of interest to study the modal (in the vertical) decomposition of corrections made, in view of a strong hydrostatically-related connection between surface pressure and the geopotential height of the 500 hPa pressure surface (see, also, e.g., CULLEN, 2003). Such a modal-decomposition investigation should also be of interest in view of its implications on forecast error in the presence of non-modal error growth (see, e.g., FARRELL and IOANNOU, 1996). Finally, the results discussed in these assimilation studies should be expected to be very sensitive to any biases in the model.

### 4.9 Simplified models and error growth

Many of the studies carried out to investigate properties and performance of ensemble-based Kalman filtering have been performed within contexts of simplified or intermediate-complexity geofluid-dynamical models (e.g., CLAUSSEN et al., 2002), or with low-dimensional models (such as the presently widely used low-dimensional model proposed by LORENZ, 1996). While such models offer the clear advantage of a much reduced dimensionality $n$ of the state $\mathbf{x}$, their use limits to some unknown extent the possibilities of generalizing results so obtained to applications within more realistic contexts. It is in particular unclear how to 'scale' results obtained in situations with, for example, $n = 3$ and $m = 9$, to contexts with $n$ and $m$ both being on the order of $10^6$ or higher.

A fundamental limitation in terms of generalizing results may arise whenever the model used within a particular investigation has error-growth properties that are different from those found in state-of-the-art models (see, e.g., TRIBBIA and BAUMHEFNER, 2004; SIMMONS and HOLLINGSWORTH, 2002; SIMMONS et al., 1995; ZHANG et al., 2003). The importance of carefully considering the error-growth properties of the assimilating model, when interpreting data assimilation results becomes obvious from the equations describing the KF prediction step, Eqs. (2.11) and (2.12), since the error-growth characteristics of the assimilating model will determine magnitude and structure of the new background-error covariance described through the time-evolved ensemble.

Deficiency or excess in growth will thus affect structure and magnitude of $\hat{\mathsf{P}}^{\mathrm{b}}$, as well as of the Kalman gain $\hat{\mathsf{K}}$ (see, Eq. (3.3)). Through the sequential nature of the filtering process, a further consequence may be filter divergence or convergence to inappropriate or unphysical stationary results. The nature of the stationary properties of the KF has been considered in detail in a modal context by DALEY and MÉNARD (1993). For an unstable mode, the stationary values (denoted by subscript 's') for background and analysis error variances (in the absence of model error) are given by:

$$\left(\frac{\sigma_{\mathrm{b}}}{\sigma_{\mathrm{o}}}\right)_s^2 = m^2 - 1\,, \qquad \left(\frac{\sigma_{\mathrm{a}}}{\sigma_{\mathrm{o}}}\right)_s^2 = \frac{m^2-1}{m^2}\,, \qquad (4.2)$$

where $m^2 \equiv \sigma_{\mathrm{b}}^2/\sigma_{\mathrm{a}}^2 > 1$ measures the mode's amplification (see, also, Eq. (2.12), as well as appendix F for notation; note that the relevant stationary value for the gain – see Eq. (F.2) – is obtained as $k_s = 1 - m^{-2}$). Evidently then, the results (4.2), stated by DALEY and MÉNARD (1993), indicate that the model's error-growth properties fundamentally determine the asymptotic behavior of the KF (note that the situation is somewhat different when model error is included in the analysis).

Obviously, these results will generalize to the situation of a system consisting of a large collection of (stable and unstable) modes, with additional aspects occurring of course when nonmodal growth plays an important role.

As an example, it is mentioned that the error-doubling time of the assimilating model used in the study by ANDERSON et al. (2005) to assimilate surface pressure observations (see, section 4.8) is quite long – suggesting an under-active model – in comparison to presently available estimates of doubling times (i.e., 5.5 days compared to 1.5 days as suggested by the study of SIMMONS et al., 1995). In view of the above discussion then it is conceivable that the assimilation results might turn out to be quite different when obtained on the basis of a more active assimilating model.

Another important factor that may strongly affect results and conclusions is linked to the properties of the assimulating model's climate that depends (among others) on terms added to the dynamical-core equations, such as forcing, damping, and relaxation terms, in order to have the model perform more realistically when compared to observed atmospheric behavior. As an example, HAMILL et al. (2003) report difficulties in obtaining good analyses in the tropical upper troposphere when experimenting with a primitive-equation based assimilating model using the forcing described by HELD and SUAREZ (1994) in order to have the model reflect a realistic climate.

Both of these above issues need to be taken into account, in particular, when trying to generalize the comparative behavior of one data assimilation method to another method to different and/or more complex assimilation problems (see, e.g., WANG and BISHOP, 2003; LORENC and RAWLINS, 2005; FISHER et al., 2005), or, when assessing the impact of new observation types on analysis quality (see, e.g., STOFFELEN et al., 2006; WEISSMANN and CARDINALI, 2007; TAN et al., 2007).

Nevertheless, the carefully considered usage of simplified models offers clear advantages and should be considered useful for the initial exploration of assimilation methods, in particular also within ensemble-based Kalman filtering. It is certainly possible to investigate within a simplified modeling context, and a controlled data assimilation environment (e.g., assuming truth is known), how a data assimilation method reacts to data-denial situations or how it compares to other data-assimilation methodology candidates.

For example, numerous studies have utilized the low-dimensional LORENZ (1996) model (see, e.g., LORENZ and EMANUEL, 1998; ANDERSON, 2001; WHITAKER and HAMILL, 2002; BOWLER, 2006; LAWSON and HANSEN, 2004; FISHER et al., 2005; ORRELL et al., 2001; ORRELL, 2003, 2005). Models based on quasigeostrophic (QG) atmospheric dynamics (such as, e.g., the widely used QG model of MARSHALL and MOLTENI, 1993) have been used, for example, in the studies by BISHOP et al. (2003), HAMILL and SYNDER (2002), HAMILL et al. (2002), and Beck and EHRENDORFER (2005).

Some questions, however, remain inaccessible for study given a certain model, such as issues related to the loss of dynamical balance (see, section 4.4) when a filtered (e.g., QG) model is used. Conclusions obtained in such situations may possibly thus best be split into conclusions of general versus limited validity.

### 4.10 Ease of implementation and costs

A detailed discussion of ease of implementation and costs of variational assimilation methods versus ensemble-based Kalman filtering may be found in LORENC (2003b). One of the major obstacles in implementing variational data assimilation is the need for an appropriate adjoint model (e.g., ERRICO, 1997), with no comparable "difficulty" present in the implementation of, for example, the EnKF or the EnSRF. This difficulty is, however, clearly ameliorated by the availability of so-called automatic adjoint generators. In terms of fully assessing both approaches in terms of costs required and analysis quality achievable, as well as in view of the important issue of keeping the ensemble spread matching the error, probably substantially more operational experience will be needed (e.g., LORENC, 2003a; HOUTEKAMER et al. 2005), also through enhanced collaboration within the data assimilation community (e.g., MCLAUGHLIN et al., 2005).

## 5 Stochastic and deterministic filters

In the basic form of the EnKF, Evenson (1994) suggested updating each member of the ensemble of short-term forecasts on the basis of the same set of observations. Subsequently though, as also pointed out by WHITAKER and HAMILL (2002), theoretical justification was provided for perturbing the observations in the EnKF by BURGERS et al. (1998), in order to avoid systematic underestimation of analysis error covariances. Treating the observations as random variables through the use of perturbed observations was also advocated by HOUTEKAMER and MITCHELL (1998) in an effort to maintain sufficient spread in the ensemble and thus avoid filter divergence (see, WHITAKER and HAMILL, 2002).

Thus, in this variant of the EnKF, each member of the ensemble is updated within the analysis step with the same gain, but with, however, different sets of perturbed observations, as indicated in Eq. (3.2). On the basis of such perturbed observations, this form of the EnKF provides a proper estimate of the uncertainty in the analysis (see, HOUTEKAMER and MITCHELL, 2005; HAMILL, 2006). A filter that uses such a stochastic updating algorithm in the analysis step is known as a *stochastic* filter

(see, e.g., HOUTEKAMER and MITCHELL, 2005; LAWSON and HANSEN, 2004).

The class of non-stochastic or *deterministic* filters (see, e.g., HAMILL, 2006; LAWSON and HANSEN, 2004) has been discussed in detail by TIPPETT et al. (2003). In these filters, observations are not perturbed before being assimilated, based on the argument that such perturbations can be a source of additional sampling error when samples are already small (HOUTEKAMER and MITCHELL, 2005). Instead, the update is performed such that the analysis-error covariance matrix obtained is of the form as given through the KF equations, assuming that the background-error covariance is modelled from the background ensemble $\mathbf{x}_i^b$ (see, also HAMILL, 2006).

The EnSRF, described by WHITAKER and HAMILL (2002), the EAKF, described by ANDERSON (2001), the Ensemble Transform Kalman Filter (ETKF), described by BISHOP et al. (2001), as well as the error-subspace statistical estimation (ESSE) method, described by LERMUSIAUX and ROBINSON (1999), are all deterministic filters in this sense. As such, these filters include an appropriate analysis or update step through a linear transformation of the background ensemble into the analysis ensemble (see, WHITAKER and HAMILL, 2002; LAWSON and HANSEN, 2004; KHARE and ANDERSON, 2006; ETHERTON, 2007), with the updated ensemble covariance constrained to satisfy Eq. (2.6). Regarding the cost of serial observation processing within the ETKF, also in view of the comments made in section 3.1.2, reference is made to BISHOP et al. (2001).

Comparisons of the performance of these deterministic filters with the performance of other candidate assimilation methods have been reported, for example, by WHITAKER and HAMILL (2002) and WANG and BISHOP (2003), with WHITAKER and HAMILL (2002) demonstrating that the elimination of sampling error associated with the perturbed observations improves the accuracy of estimated analysis error covariances over methods with perturbed observations, for the same ensemble size (see, also, LORENC, 2003b). On the other hand, WHITAKER and HAMILL (2002) point out the asymptotic correctness of using perturbed observations for large ensembles. Also, HAMILL (2006) provides an example comparison showing that in terms of describing the pdf, the EnKF works somewhat better than the EnSRF.

As discussed by TIPPETT et al. (2003) these non-stochastic ensemble filters, as well as the EnKF, all belong to the family of ensemble square-root filters which, by definition work with a square-root of the analysis error covariance matrix, such as the formulation (2.6), or the formulation (3.7). Working with a (nonunique) square-root has the advantage of implicitly avoiding loss of positive-definiteness of the covariance matrix. In addition, the non-uniqueness of the square-root of positive-definite matrices allows for a considerable variety of square-root filters (TIPPETT et al., 2003). Methods suggested in the literature differ in terms of details though, and, for example, as pointed out by both HAMILL (2006) and LORENC (2003b), it is claimed that it is not possible to apply covariance localization through inclusion of a Schur product formulation within the ETKF of BISHOP et al. (2001). Still, the algorithm may be localized by applying it repeatedly to different areas, with a different selection of observations for each.

## 6 Summary and discussion

In this paper an attempt has been made to provide an overview of ensemble-based data assimilation methods, related in particular to Kalman filtering. Originating with the fundamental work by KALMAN (1960), a very large number of studies have used this methodology that is fundamentally based on Bayesian ideas as pointed out in section 2. Difficulties that arise in typical earth sciences applications (e.g., atmospheric data assimilation) due to the large dimensionality of the state and the large number of observations to be assimilated have led to the consideration of ensemble-based implementations within this field (e.g., EVENSON, 1994, 2007) that allow for circumventing some of these problems.

Following this early work on ensemble-based Kalman filtering, a number of modifications, improvements, and variants of the basic Monte-Carlo approach were suggested, such as the EnKF, the EAKF, the ETKF, the EnSRF, and the LEKF. Consideration of the properties of these and other methods within various data assimilation contexts has led to an exceptionally high number of studies devoted to investigating, comparing, simplifying, improving, or applying ensemble-based Kalman filtering in various contexts (see the comprehensive reviews by HAMILL, 2006; HOUTEKAMER and MITCHELL, 2005; LORENC, 2003b). This high number of studies concerned with Kalman filtering in the atmospheric sciences is best illustrated by noting that over the past ten years, sixty-five journal articles have appeared in the journals of the American Meteorological Society carrying the words *Kalman filter* in their title. Similarly, twelve papers with *Kalman filter* in their title have appeared in *Tellus A* since 2004. These numbers also illustrate the increasing importance of data assimilation research (e.g., MCLAUGHLIN et al., 2005), which is further supported by the fact that a third of the papers that have been published in a recent issue (number 626) of *The Quarterly Journal of the Royal Meteorological Society* report directly on data assimilation research.

One fundamental issue (ANDERSON, 2001) in ensemble-based Kalman filtering relates to the fact that the number of ensemble members $N$ is usually quite small compared to the dimension of the state

*n*. Thus, considerable sampling variability must be expected when estimating certain quantities like variances and covariances from the ensembles. The discussion of this fundamental issue, and its consequences, such as the need for smoothing or localizing estimated covariances, as well as related issues, such as the serial processing of uncorrelated observations, and the ease of implementation and computational cost of these methods, formed the central topic of this article. Some of the variants of the basic ensemble-based Kalman filtering approach were discussed briefly by reviewing results obtained in various individual data assimilation settings. It remains to point out that a comparative assessment of the presently available methodologies needs careful consideration of dependencies on the specific data assimilation environment considered (see also end of section 4.9).

Beyond reviewing some of the literature within ensemble-based Kalman filtering, the primary intention of the paper has been to provide – in parallel to several other recent reviews (e.g., HAMILL, 2006; EVENSEN, 2003, 2007; LORENC, 2003b; HOUTEKAMER and MITCHELL, 2005) – a stepping stone for the improved understanding of the properties of these methods through pointing out issues that are involved and that will need to be addressed further. It would, for example, be highly insightful to perform a comparative study that applies one method throughout a hierarchy of increasingly more complex models, and to ascertain to what extent conclusions reached within the simplest assimilating model context do in fact generalize to the most complex model (see also, section 4.9). Or, a specific issue, such as balance considerations or serial observation processing together with data quality control, may be investigated in depth using various ensemble-based methods within one given data assimilation setup. Similarly, future efforts will probably increasingly concentrate on investigating the comparative performance of different methods within one specifically selected data assimilation problem, possibly even under operational constraints.

## Acknowledgements

Parts of the developments in section 4.1 are related to ideas discussed by M. FISHER, ECMWF, during the *Seventh International Workshop on Adjoint Applications in Dynamic Meteorology*, held at Universitätszentrum Obergurgl, Tyrol, Austria, in October 2006. I am indebted to Ron ERRICO, GMAO/NASA, for numerous very fruitful discussions and comments on the subject discussed in this article. Comments by four reviewers (including R. ERRICO and J. KEPERT) have been invaluable in improving this review.

## Appendix A: Inverse of a matrix sum

If the necessary inverses exist, it is possible, as described by, for example, Eq. (A.2.4f) of MARDIA et al. (1982), to write the inverse of the sum of two matrices as follows:

$$\begin{aligned}&\Big(\mathsf{A}+\mathsf{BCD}\Big)^{-1}=\\&=\mathsf{A}^{-1}-\mathsf{A}^{-1}\mathsf{B}\Big(\mathsf{C}^{-1}+\mathsf{DA}^{-1}\mathsf{B}\Big)^{-1}\mathsf{DA}^{-1}.\end{aligned}\qquad\text{(A.1)}$$

## Appendix B: Reformulation of Equation (2.4)

Manipulation of the exponent in (2.4) shows that the posterior pdf, given in Eq. (2.4), is multivariate normal with mean $\mathbf{x}^{\mathrm{a}}$ and covariance $\mathsf{P}^{\mathrm{a}}$, given by Eqs. (2.9) and (2.6), respectively, since, by changing constants (i.e., terms not containing $\mathbf{x}$), and essentially "completing the square", one obtains:

$$\begin{aligned}&(\mathbf{y}-\mathsf{H}\mathbf{x})^{\mathrm{T}}\mathsf{R}^{-1}(\mathbf{y}-\mathsf{H}\mathbf{x})+\\&+(\mathbf{x}-\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\mathrm{b}})=\\&=\underbrace{\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathbf{y}}_{const}-2\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}\mathbf{x}+\mathbf{x}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}\mathbf{x}+\\&+\mathbf{x}^{\mathrm{T}}\mathsf{B}^{-1}\mathbf{x}-2\mathbf{x}^{\mathrm{T}}\mathsf{B}^{-1}\mathbf{x}^{\mathrm{b}}+\underbrace{(\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}\mathbf{x}^{\mathrm{b}}}_{const}=\\&=\mathbf{x}^{\mathrm{T}}\underbrace{\Big(\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}+\mathsf{B}^{-1}\Big)}_{\equiv(\mathsf{P}^{\mathrm{a}})^{-1}}\mathbf{x}-\\&-2\Big(\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}+(\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}\Big)\mathbf{x}+const=\\&=\mathbf{x}^{\mathrm{T}}(\mathsf{P}^{\mathrm{a}})^{-1}\mathbf{x}-2\Big(\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}+(\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}\Big)\mathbf{x}+\\&+\Big(\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}+(\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}\Big)\\&\mathsf{P}^{\mathrm{a}}\Big(\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}+(\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}\Big)^{\mathrm{T}}+const'=\\&=\mathbf{x}^{\mathrm{T}}(\mathsf{P}^{\mathrm{a}})^{-1}\mathbf{x}-2\Big(\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}+(\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}\Big)\mathbf{x}+\\&+\Big(\mathbf{y}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}+(\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathsf{B}^{-1}\Big)\\&\mathsf{P}^{\mathrm{a}}\underbrace{\Big(\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathbf{y}+\mathsf{B}^{-1}\mathbf{x}^{\mathrm{b}}\Big)}_{\equiv(\mathsf{P}^{\mathrm{a}})^{-1}\mathbf{x}^{\mathrm{a}}}+const'\\&=\mathbf{x}^{\mathrm{T}}(\mathsf{P}^{\mathrm{a}})^{-1}\mathbf{x}-2(\mathbf{x}^{\mathrm{a}})^{\mathrm{T}}(\mathsf{P}^{\mathrm{a}})^{-1}\mathbf{x}+(\mathbf{x}^{\mathrm{a}})^{\mathrm{T}}(\mathsf{P}^{\mathrm{a}})^{-1}\mathbf{x}^{\mathrm{a}}=\\&=\Big(\mathbf{x}-\mathbf{x}^{\mathrm{a}}\Big)^{\mathrm{T}}(\mathsf{P}^{\mathrm{a}})^{-1}\Big(\mathbf{x}-\mathbf{x}^{\mathrm{a}}\Big)+const'.\end{aligned}\qquad\text{(B.1)}$$

## Appendix C: Reformulation of Equation (2.7)

The Kalman gain, defined in Eq. (2.7), may be rewritten in the following equivalent formulations, on the basis of (A.1) and (2.6):

$$\begin{aligned}
\mathsf{K} &= \mathsf{B}\mathsf{H}^{\mathrm{T}}(\mathsf{H}\mathsf{B}\mathsf{H}^{\mathrm{T}}+\mathsf{R})^{-1} = \\
&= \mathsf{B}\mathsf{H}^{\mathrm{T}}\Big[\mathsf{R}^{-1}-\mathsf{R}^{-1}\mathsf{H}\Big(\mathsf{B}^{-1}+\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}\Big)^{-1}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\Big] = \\
&= \Big[\mathsf{B}-\mathsf{B}\underbrace{\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}}\Big(\mathsf{B}^{-1}+\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}\Big)^{-1}\Big]\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1} = \\
&= \Big[\mathsf{B}-\mathsf{B}\Big\{(\mathsf{P}^{\mathrm{a}})^{-1}-\mathsf{B}^{-1}\Big\}\mathsf{P}^{\mathrm{a}}\Big]\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1} = \\
&= \Big[\mathsf{B}-\mathsf{B}\Big\{\mathsf{I}-\mathsf{B}^{-1}\mathsf{P}^{\mathrm{a}}\Big\}\Big]\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1} = \mathsf{P}^{\mathrm{a}}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\,,
\end{aligned} \tag{C.1}$$

where it is noted that the underbraced term in the third line equals the term in curly brackets in the fourth line due to (2.6).

## Appendix D: Proof that $\nabla\mathscr{J}(\mathbf{x})=0$ for $\mathbf{x}=\mathbf{x}^{\mathrm{a}}$

Note first that $\mathscr{J}$, defined in Eq. (2.13), is the negative of the exponent in Eq. (2.4). It is shown here that $\mathscr{J}$ is minimized for $\mathbf{x}=\mathbf{x}^{\mathrm{a}}$ which is equivalent to showing that $\nabla\mathscr{J}(\mathbf{x})=0$ for $\mathbf{x}=\mathbf{x}^{\mathrm{a}}$. Through this equivalence a way is provided to determine $\mathbf{x}^{\mathrm{a}}$ given in Eq. (2.9) or, equivalently, (2.10), – as the mean of the posterior pdf (2.4) – through minimization of $\mathscr{J}$. The proof proceeds by inserting $\mathbf{x}^{\mathrm{a}}$, given in Eq. (2.9) or, equivalently, (2.10), into (2.14) and then showing that for this choice one obtains $\nabla\mathscr{J}(\mathbf{x}^{\mathrm{a}})=0$. Following this procedure the result is immediately obvious from:

$$\begin{aligned}
\nabla\mathscr{J}(\mathbf{x}^{\mathrm{a}}) &= \mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}(\mathsf{H}\mathbf{x}^{\mathrm{a}}-\mathbf{y})+\mathsf{B}^{-1}(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{b}}) = \\
&= \mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}(\mathsf{H}\mathbf{x}^{\mathrm{a}}-\mathbf{y})+(\mathsf{P}^{\mathrm{a}})^{-1}(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{b}})+ \\
&\quad +\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}(\mathbf{x}^{\mathrm{b}}-\mathbf{x}^{\mathrm{a}}) = \\
&= \mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}(\mathsf{H}\mathbf{x}^{\mathrm{a}}-\mathbf{y})+(\mathsf{P}^{\mathrm{a}})^{-1}\mathsf{K}\Big(\mathbf{y}-\mathsf{H}\mathbf{x}^{\mathrm{b}}\Big)+ \\
&\quad +\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}(\mathbf{x}^{\mathrm{b}}-\mathbf{x}^{\mathrm{a}}) = \\
&= \mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}(\mathsf{H}\mathbf{x}^{\mathrm{a}}-\mathbf{y}+\mathbf{y}-\mathsf{H}\mathbf{x}^{\mathrm{b}})+ \\
&\quad +\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}(\mathbf{x}^{\mathrm{b}}-\mathbf{x}^{\mathrm{a}}) = 0\,,
\end{aligned} \tag{D.1}$$

where Eqs. (2.6), (2.10), and (2.8) have been used in turn.

## Appendix E: Serial processing of observations

In order to prove that serial processing of batches of observations is equivalent to processing all observations at the same time (assuming uncorrelated observational error and Gaussian[4] pdfs, as well as linear operators), it is first noted that such a statement is already contained in the basic updating analysis equations of the KF, since the KF updating equations combine background and observational information in an optimal way. With slight change in notation, the basic KF result is therefore first rewritten as follows.

When given two pieces of information $\mathbf{z}_1$ and $\mathbf{z}_2$ with error characteristics $\mathsf{S}_1$ and $\mathsf{S}_2$ and operators $\mathsf{H}_1$ and $\mathsf{H}_2$ relating $\mathbf{z}_1$ and $\mathbf{z}_2$ to the state $\mathbf{w}$ to be estimated, the optimal state estimate is given by:

$$\mathbf{w}^{\mathrm{a}} \equiv \mathsf{W}^{\mathrm{a}}\Big(\mathsf{H}_1^{\mathrm{T}}\mathsf{S}_1^{-1}\mathbf{z}_1+\mathsf{H}_2^{\mathrm{T}}\mathsf{S}_2^{-1}\mathbf{z}_2\Big) \tag{E.1}$$

where:

$$\mathsf{W}^{\mathrm{a}} = \Big(\mathsf{H}_1^{\mathrm{T}}\mathsf{S}_1^{-1}\mathsf{H}_1+\mathsf{H}_2^{\mathrm{T}}\mathsf{S}_2^{-1}\mathsf{H}_2\Big)^{-1}\,. \tag{E.2}$$

Note that (E.1) and (E.2) are an obvious generalization of (2.9) and (2.6), respectively. This result is symbolically written as:

$$\mathbf{z}_1,\mathbf{z}_2,\mathsf{S}_1,\mathsf{S}_2,\mathsf{H}_1,\mathsf{H}_2 \quad\curvearrowright\quad \mathbf{w}^{\mathrm{a}},\mathsf{W}^{\mathrm{a}}\,. \tag{E.3}$$

On the basis of this result and utilizing this symbolic notation, it is now necessary to show that the one–step process (as actually contained in the KF equations):

$$\mathbf{x}^{\mathrm{b}},\mathbf{y},\mathsf{B},\mathsf{R},\mathsf{I},\mathsf{H} \quad\curvearrowright\quad \mathbf{x}^{\mathrm{a}},\mathsf{P}^{\mathrm{a}} \tag{E.4}$$

produces the same result as the two–step serial process written in this short–hand notation as:

$$\mathbf{x}^{\mathrm{b}},\mathbf{y}_1,\mathsf{B},\mathsf{R}_1,\mathsf{I},\mathsf{H}_1 \quad\curvearrowright\quad \mathbf{x}_1^{\mathrm{a}},\mathsf{P}_1^{\mathrm{a}} \tag{E.5}$$

$$\mathbf{x}_1^{\mathrm{a}},\mathbf{y}_2,\mathsf{P}_1^{\mathrm{a}},\mathsf{R}_2,\mathsf{I},\mathsf{H}_2 \quad\curvearrowright\quad \mathbf{x}_2^{\mathrm{a}},\mathsf{P}_2^{\mathrm{a}} \tag{E.6}$$

where the chosen notation is self–explaining and the full observation vector $\mathbf{y}$ has been partitioned, together with the error covariance and the observation operator into:

$$\mathbf{y} = \begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix} \tag{E.7}$$

$$\mathsf{R} = \begin{pmatrix}\mathsf{R}_1 & 0\\ 0 & \mathsf{R}_2\end{pmatrix} \qquad \mathsf{H} = \begin{pmatrix}\mathsf{H}_1\\ \mathsf{H}_2\end{pmatrix}\,. \tag{E.8}$$

In order to prove the correctness of the above statement, it is thus necessary to show that

$$\mathbf{x}_2^{\mathrm{a}} = \mathbf{x}^{\mathrm{a}} \tag{E.9}$$

$$\mathsf{P}_2^{\mathrm{a}} = \mathsf{P}^{\mathrm{a}} \tag{E.10}$$

with $\mathbf{x}^{\mathrm{a}}$ and $\mathsf{P}^{\mathrm{a}}$ given by (2.9) and (2.6), respectively, with, on the basis of (E.1) and (E.2), the relevant analyses $\mathbf{x}_1^{\mathrm{a}}$ and $\mathbf{x}_2^{\mathrm{a}}$ given by:

$$\mathbf{x}_1^{\mathrm{a}} = \underbrace{\Big(\mathsf{B}^{-1}+\mathsf{H}_1^{\mathrm{T}}\mathsf{R}_1^{-1}\mathsf{H}_1\Big)^{-1}}_{\equiv\mathsf{P}_1^{\mathrm{a}}}\Big(\mathsf{B}^{-1}\mathbf{x}^{\mathrm{b}}+\mathsf{H}_1^{\mathrm{T}}\mathsf{R}_1^{-1}\mathbf{y}_1\Big) \tag{E.11}$$

[4] Note that the assumption of Gaussianity facilitates the technicalities of considerations presented here (e.g., when precisely defining the meaning of 'optimal'); results may be generalized to non-Gaussian linear contexts.

and

$$\mathbf{x}_2^{\mathrm{a}} = \underbrace{\left((\mathsf{P}_1^{\mathrm{a}})^{-1} + \mathsf{H}_2^{\mathrm{T}}\mathsf{R}_2^{-1}\mathsf{H}_2\right)^{-1}}_{\equiv \mathsf{P}_2^{\mathrm{a}}} \left((\mathsf{P}_1^{\mathrm{a}})^{-1}\mathbf{x}_1^{\mathrm{a}} + \mathsf{H}_2^{\mathrm{T}}\mathsf{R}_2^{-1}\mathbf{y}_2\right) . \tag{E.12}$$

Considering first Eq. (E.12) with formulation (E.11) inserted one obtains:

$$\begin{aligned}(\mathsf{P}_2^{\mathrm{a}})^{-1}\mathbf{x}_2^{\mathrm{a}} &= \left(\mathsf{B}^{-1}\mathbf{x}^{\mathrm{b}} + \mathsf{H}_1^{\mathrm{T}}\mathsf{R}_1^{-1}\mathbf{y}_1 + \mathsf{H}_2^{\mathrm{T}}\mathsf{R}_2^{-1}\mathbf{y}_2\right) = \\ &= \mathsf{B}^{-1}\mathbf{x}^{\mathrm{b}} + \begin{pmatrix}\mathsf{H}_1^{\mathrm{T}} & \mathsf{H}_2^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\mathsf{R}_1 & 0 \\ 0 & \mathsf{R}_2\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{y}_1 \\ \mathbf{y}_2\end{pmatrix} = \\ &= \mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathbf{y} + \mathsf{B}^{-1}\mathbf{x}^{\mathrm{b}}\end{aligned} \tag{E.13}$$

which, on the basis of (E.7) and (E.8) is precisely equal to (2.9), as desired, **if**, however, (E.10) is correct. The correctness of (E.10), however, is immediately seen, by, in this second step, inserting the definition of $\mathsf{P}_1^{\mathrm{a}}$ from Eq. (E.11) into the definition of $\mathsf{P}_2^{\mathrm{a}}$ in Eq. (E.12), obtaining:

$$\begin{aligned}(\mathsf{P}_2^{\mathrm{a}})^{-1} &= \mathsf{B}^{-1} + \mathsf{H}_1^{\mathrm{T}}\mathsf{R}_1^{-1}\mathsf{H}_1 + \mathsf{H}_2^{\mathrm{T}}\mathsf{R}_2^{-1}\mathsf{H}_2 = \\ &= \mathsf{B}^{-1} + \begin{pmatrix}\mathsf{H}_1^{\mathrm{T}} & \mathsf{H}_2^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\mathsf{R}_1 & 0 \\ 0 & \mathsf{R}_2\end{pmatrix}^{-1}\begin{pmatrix}\mathsf{H}_1 \\ \mathsf{H}_2\end{pmatrix} = \\ &= \mathsf{B}^{-1} + \mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{H}\end{aligned} \tag{E.14}$$

which, on the basis of (E.8) is precisely equal to (2.6), as desired. This result concludes the elementary proof of the statement mentioned at the beginning of this appendix.

A few comments are in order. First, the simplicity of this proof, as contrasted, for example, with Appendix 3 of section 2 in WUNSCH (2006), or with section 3.3 of EHRENDORFER (1991), is based on formulating the KF updating equations in terms of the abstract notation (E.1) and (E.2), or (E.3), respectively (see, also, appendix B of BISHOP et al., 2001). Second, the result obtained here, shows the power contained in the KF equations that were derived in section 2.1 using Bayes' Theorem, as well as using the reformulations produced in appendices A, B, and C; from that point on, however, it is not be necessary to implicitly rederive these reformulations again, especially through using result (A.1), since they have already been used. Third, the notation used here also shows most clearly that in the linear – and Gaussian – situation present here the use of more than two pieces of information is achieved by just summing them up, weighted by their covariance structures; see, Eqs. (E.13) and (E.14). Fourth, the basic assumption for serial processing of observations (uncorrelated Gaussian observational error), or for combining observations $\mathbf{y}$ and background $\mathbf{x}^{\mathrm{b}}$ (uncorrelated observational and background errors), as in Eq. (2.9), becomes most clear in this notation. Finally, it is also very transparent from the above discussion that the result will not be established should covariance localization (see, section 4.3) be used when performing serial observation processing. An example of serial observation processing for a two–dimensional state and two observations may be found on p. 304 of GELB (1974).

## Appendix F: Degrading the background

In order to gain additional insight into the nature of the KF equations presented in section 2.1, a one-dimensional version of the KF is investigated here, using the following state updating equation, in analogy to (2.9):

$$x_{\mathrm{a}} = ky + (1-k)x_{\mathrm{b}} , \tag{F.1}$$

with the Kalman gain given, from Eq. (2.7) as:

$$k = \frac{\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{b}}^2 + \sigma_{\mathrm{o}}^2} = \frac{1}{1 + \left(\frac{\sigma_{\mathrm{o}}}{\sigma_{\mathrm{b}}}\right)^2} . \tag{F.2}$$

The following question is posed: under the assumption that background and observation error variance, $\sigma_{\mathrm{b}}^2$ and $\sigma_{\mathrm{o}}^2$ are fixed and known[5], what is the fraction of cases in which the background is degraded so that the analysis is actually less accurate than the background was before the analysis increment was added. To answer this question consider the specification that the observation $y$ is a normal random variable with zero mean and variance $\sigma_{\mathrm{o}}^2$ and that the background is also normal with zero mean and variance $\sigma_{\mathrm{b}}^2$, as also assumed in Eqs. (2.1) and (2.2):

$$y \sim \mathcal{N}(0, \sigma_{\mathrm{o}}^2) \qquad\qquad x_{\mathrm{b}} \sim \mathcal{N}(0, \sigma_{\mathrm{b}}^2) . \tag{F.3}$$

Assuming, as usual, uncorrelated noise, $y$ and $x_{\mathrm{b}}$ are independent and therefore $x_{\mathrm{a}}$ is also normal with mean zero and with variance $\sigma_{\mathrm{a}}^2 = k^2\sigma_{\mathrm{o}}^2 + (1-k)^2\sigma_{\mathrm{b}}^2$, as stated in general terms in Theorem 2 of section 5.6 in DEGROOT (1986), or as also contained in the last equality of Eq. (2.6):

$$x_{\mathrm{a}} \sim \mathcal{N}(0, \sigma_{\mathrm{a}}^2) , \tag{F.4}$$

and:

$$\sigma_{\mathrm{a}}^2 = k^2\sigma_{\mathrm{o}}^2 + (1-k)^2\sigma_{\mathrm{b}}^2 = \left(\frac{1}{\sigma_{\mathrm{o}}^2} + \frac{1}{\sigma_{\mathrm{b}}^2}\right)^{-1} . \tag{F.5}$$

Note that the results in the present one–dimensional example correspond – not surprisingly – precisely to the results presented in section 2.1 specialized to the scalar situation. For example, result (F.5) corresponds to the fourth and the first expression for $\mathsf{P}^{\mathrm{a}}$ in Eq. (2.6).

Returning to the original question in the present situation, it is apparent that the analysis step may move the analysis closer to the true value of zero or further away from that value when compared to the position of

[5]Subscripts rather than superscripts are used here to allow for better readability when squared variables need to be considered.

the background, for any given realization of both observation and background. Thus, the analysis error will be identified as the expectation of the squared difference of $x_a$ from zero:

$$\varepsilon_a^2 = E(x_a^2) = \sigma_a^2 \ , \qquad \text{(F.6)}$$

and similarly for the background error:

$$\varepsilon_b^2 = E(x_b^2) = \sigma_b^2 \ . \qquad \text{(F.7)}$$

Obviously, since on the basis of (F.5), $\sigma_a^2 < \sigma_b^2$ is obtained, the analysis error is expected to be smaller than the background error which has to be the case since a new piece of independent information (namely the observation) with a finite variance has been introduced. However, if the observation is rather inaccurate, or due to the process of *sampling*, the analysis error may actually be larger than the background error on individual cases. To investigate this problem, the random variable $d$ is defined as the difference between $x_a^2$ and $x_b^2$:

$$d = x_a^2 - x_b^2 \ . \qquad \text{(F.8)}$$

Clearly, whenever $d > 0$, for a specific realization, the situation is faced in which the analysis error is greater than the background error on an individual case–by–case occurrence. The probability for seeing that situation is depicted in Fig. 2 in section 4, as a function of different values for $\sigma_o/\sigma_b$.

It is difficult to assess the distribution function for $d$ analytically, as it is the difference of two dependent squared random normal variables. The mean of $d$, however, is readily assessed as:

$$\begin{aligned} E(d) = E(x_a^2 - x_b^2) = \sigma_a^2 - \sigma_b^2 = \\ = k^2\sigma_o^2 + (1-k)^2\sigma_b^2 - \sigma_b^2 \ , \end{aligned} \qquad \text{(F.9)}$$

where Eqs. (F.5), (F.6), and (F.7) have been used. It is illustrative to rewrite Eq. (F.9) as:

$$\frac{E(d)}{\sigma_b^2} = k^2\left(\frac{\sigma_o}{\sigma_b}\right)^2 + (1-k)^2 - 1 = -k = \left(\frac{\sigma_a}{\sigma_b}\right)^2 - 1 \ , \qquad \text{(F.10)}$$

which is an interesting result, showing that the expectation of $d$ is always negative, which has to be the case, also in line of the comments following Eq. (F.7). Note that the last equality in the above result is obtained by rewriting $k$ using (F.2) and (F.5). A further discussion of (F.10) is given in section 4.1.

## References


ANDERSON, J.L., 2001: An Ensemble Adjustment Kalman Filter for Data Assimilation. – Mon. Wea. Rev. , **129**, 2884–2903.

—, 2003: A Local Least Squares Framework for Ensemble Filtering. –Mon. Wea. Rev. **131**, 634–642.

ANDERSON, J.L., S.L. ANDERSON, 1999: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. – Mon. Wea. Rev. **127**, 2741–2758.

ANDERSON, J.L., B. WYMAN, S. ZHANG, T. HOAR, 2005: Assimilation of Surface Pressure Observations Using an Ensemble Filter in an Idealized Global Atmospheric Prediction System. – J. Atmos. Sci. **62**, 2925–2938.

ANDERSSON, E., J. HASELER, P. UNDÉN, P. COURTIER, G. KELLY, D. VASILJEVIĆ, C. BRANKOVIĆ, C. CARDINALI, C. GAFFARD, A. HOLLINGSWORTH, C. JAKOB, P. JANSSEN, E. KLINKER, A. LANZINGER, M. MILLER, F. RABIER, A. SIMMONS, B. STRAUSS, J.-N. THÉPAUT, P. VITERBO, 1998: The ECMWF implementation of three–dimensional variational assimilation (3D–Var). III: Experimental results. – Quart. J. Roy. Meteor. Soc. **124**, 1831–1860.

AULIGNÉ, T., A.P. MCNALLY, D.P. DEE, 2007: Adaptive bias correction for satellite data in a numerical weather prediction system. – Quart. J. Roy. Meteor. Soc. **133**, 631–642.

BANNISTER, R.N., 2007: Can wavelets improve the representation of forecast error covariances in variational data assimilation? – Mon. Wea. Rev. **135**, 387–408.

BECK, A., M. EHRENDORFER, 2005: Singular-vector-based covariance propagation in a quasigeostrophic assimilation system. – Mon. Wea. Rev. **133**, 1295–1310.

BERLINER, L.M., Z.Q. LU, C. SNYDER, 1999: Statistical design for adaptive weather observations. –J. Atmos. Sci.**56**, 2536–2552.

BERLINER, L.M., R.A. LEVINE, D.J. SHEA, 2000: Bayesian climate change assessment. – J. Climate **13**, 3805–3820.

BISHOP, C.H., B.J. ETHERTON, S.J. MAJUMDAR, 2001: Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. – Mon. Wea. Rev. **129**, 420–436.

BISHOP, C.H., C.A. REYNOLDS, M.K. TIPPETT, 2003: Optimization of the Fixed Global Observing Network in a Simple Model. – J. Atmos. Sci. **60**, 1471–1489.

BJERKNES, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. – Meteorol. Z. **21**, 1–7.

BOWLER, N.E., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. – Tellus **58**A, 538–548.

BUEHNER, M., 2005: Ensemble-derived stationary and flow–dependent background–error covariances: Evaluation in a quasi–operational NWP setting. – Quart. J. Roy. Meteor. Soc. **131**, 1013–1043.

BURGERS, G., P.J. VAN LEEUWEN, G. EVENSEN, 1998: Analysis scheme in the Ensemble Kalman Filter. – Mon. Wea. Rev. **126**, 1719–1724.

BURKHARDT, U., E. BECKER, 2006: A Consistent Diffusion-Dissipation Parameterization in the ECHAM Climate Model. – Mon. Wea. Rev. **134**, 1194–1204.

CANDILLE, G., C. CÔTÉ, P.L. HOUTEKAMER, G. PELLERIN, 2007: Verification of an ensemble prediction system against observations. – Mon. Wea. Rev. **135**, 2688–2699.

CANDILLE, G., O. TALAGRAND, 2005: Evaluation of probabilistic prediction systems for a scalar variable. – Quart. J. Roy. Meteor. Soc. **131**, 2131–2150.

CATLIN, D.E., 1989: Estimation, Control, and the Discrete Kalman Filter. – Volume 71 of Applied Mathematical Science, Springer, 274 pp.

CHUI, C.K., G. CHEN, 1987: Kalman Filtering with Real-Time Applications. Springer, 191 pp.

—, —, 1989: Linear Systems and Optimal Control. Springer, 155 pp.

CLAUSSEN, M., L.A. MYSAK, A.J. WEAVER, M. CRUCIFIX, T. FICHEFET, M.F. LOUTRE, S.L. WEBER, J. ALCAMO, V.A. ALEXEEV, A. BERGER, R. CALOV, A. GANOPOLSKI, H. GOOSSE, G. LOHMANN, F. LUNKEIT, I.I. MOKHOV, V. PETOUKHOV, P. STONE, Z. WANG, 2002: Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. – Climate Dynam. **18**, 579–586.

COHN, S.E., 1993: Dynamics of short–term univariate forecast error covariances. – Mon. Wea. Rev. **121**, 3123–3149.

—, 1997: An introduction to estimation theory. – J. Meteor. Soc. Jap. **75** (1B), 257–288.

COHN, S.E., A. DA SILVA, J. GUO, M. SIENKIEWICZ, D. LAMICH, 1998: Assessing the Effects of Data Selection with the DAO Physical–Space Statistical Analysis System. – Mon. Wea. Rev. **126**, 2913–2926.

COMPO, G.P., J.S. WHITAKER, P.D. SARDESHMUKH, 2006: Feasibility of a 100–year reanalysis using only surface pressure data. – Bull. Amer. Meteor. Soc. **87**, 175–190.

COURTIER, P., 1994: Introduction to numerical weather prediction data assimilation methods. — In: Proceedings ECMWF Seminar on Developments in the Use of Satellite Data in Numerical Weather Prediction, 6–10 September 1993, 189–207.

—, 1997a: Dual formulation of four–dimensional variational assimilation. – Quart. J. Roy. Meteor. Soc. **123**, 2449–2461.

—, 1997b: Variational methods. – J. Meteor. Soc. Jap. **75**(1B), 211–218.

COURTIER, P., J. THÉPHAUT, A. HOLLINGSWORTH, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. – Quart. J. Roy. Meteor. Soc. **120**, 1367–1387.

COURTIER, P., E. ANDERSSON, W. HECKLEY, J. PAILLEUX, D. VASILJEVIĆ, M. HAMRUD, A. HOLLINGSWORTH, F. RABIER, M. FISHER, 1998: The ECMWF implementation of three–dimensional variational assimilation (3D–Var). I: Formulation. – Quart. J. Roy. Meteor. Soc. **124**, 1783–1807.

CULLEN, M.J.P., 2003: Four-dimensional variational data assimilation: A new formulation of the background–error covariance matrix based on a potential-vorticity representation. – Quart. J. Roy. Meteor. Soc. **129**, 2777–2796.

DALEY, R., 1991: Atmospheric Data Analysis. – Cambridge atmospheric and space science series, Cambridge University Press, 457 pp.

DALEY, R., R. MÉNARD, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. – Mon. Wea. Rev. **121**, 1554–1565.

DEE, D.P., 2005: Bias and data assimilation. – Quart. J. Roy. Meteor. Soc. **131**, 3323–3343.

DEGROOT, M.H., 1986: Probability and Statistics. – Addison-Wesley, 723 pp.

DERBER, J., F. BOUTTIER, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. – Tellus **51**A, 195–221.

EDWARDS, P.N., 2000: A brief history of atmospheric general circulation modeling. – In: General Circulation Model Development, D. A. RANDALL (Ed.): International Geophysics Series **70**, 67–90.

EHRENDORFER, M., 1991: Zur vergleichenden Untersuchung von Datenassimilationsmethoden an Hand der barotropen Vorticitygleichung (On the comparative investigation of data assimilation methods based on the barotropic vorticity equation). Doctoral Dissertation, Formal- und Naturwissenschaftliche Fakultät, University of Vienna, x+285 pp.

—, 1992: Four-dimensional data assimilation: comparison of variational and sequential algorithms. – Quart. J. Roy. Meteor. Soc. **118**, 673–713.

—, 1997: Predicting the uncertainty of numerical weather forecasts: a review. – Meteorol. Z., N.F. **6**, 147–183.

—, 2006: The Liouville equation and atmospheric predictability. – In: Predictability of Weather and Climate, T. PALMER, R. HAGEDORN (Eds.), Cambridge University Press, Cambridge, UK, 59–98.

EHRENDORFER, M., A. BECK, 2003: Singular vector-based multivariate normal sampling in ensemble prediction. – ECMWF Tech. Memo. No. **416**, 21 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, England].

EHRENDORFER, M., J.J. TRIBBIA, 1997: Optimal prediction of forecast error covariances through singular vectors. – J. Atmos. Sci.**54**, 286–313.

EPSTEIN, E.S., 1969: Stochastic dynamic prediction. – Tellus **21**, 739–759.

ERRICO, R.M., 1997: What is an adjoint model? – Bull. Amer. Meteor. Soc. **78**, 2577–2591.

ERRICO, R., D. BAUMHEFNER, 1987: Predictability experiments using a high-resolution limited-area model. – Mon. Wea. Rev. **115**, 488–504.

ETHERTON, B.J., 2007: Preemptive forecasts using an ensemble Kalman filter. – Mon. Wea. Rev. **135**, 3484–3495.

EUBANK, R.L., 2006: A Kalman Filter Primer. – CRC Press, 186 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** (C5), 10143–10162.

—, 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. – Ocean Dynamics **53**, 343–367.

—, 2007: Data Assimilation – The Ensemble Kalman Filter. – Springer, 279 pp.

FARRELL, B.F., P.J. IOANNOU, 1996: Generalized stability theory. Part I: autonomous operators. – J. Atmos. Sci.**53**, 2025–2040.

FERTIG, E.J., B.R. HUNT, E. OTT, I. SZUNYOGH, 2007: Assimilating non-local observations with a local ensemble Kalman filter. – Tellus **59**A, 719–730.

FISHER, M., 1998: Development of a simplified Kalman filter. – ECMWF Tech. Memo. No. 260, 16 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, England].

FISHER, M., P. COURTIER, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. – ECMWF Tech. Memo. No. **220**, 26 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, England].

FISHER, M., M. LEUTBECHER, G.A. KELLY, 2005: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. – Quart. J. Roy. Meteor. Soc. **131**, 3235–3246.

FRAEDRICH, K., E. KIRK, U. LUKSCH, F. LUNKEIT, 2005: The portable university model of the atmosphere (PUMA): Storm track dynamics and low–frequency variability. – Meteorol. Z. **14**, 735–745.

GASPARI, G., S.E. COHN, 1999: Construction of correlation functions in two and three dimensions. – Quart. J. Roy. Meteor. Soc. **125**, 723–757.

GASPARI, G., S.E. COHN, J. GUO, S. PAWSON, 2006: Construction and application of covariance functions with variable length–fields. – Quart. J. Roy. Meteor. Soc. **132**, 1815–1838.

GELARO, R., Y. ZHU, R.M. ERRICO, 2007: Examination of various-order adjoint-based approximations of observation impact. – Meteorol. Z. **16**, 685–692.

GELB, A., 1974: Applied Optimal Estimation. – The M.I.T. Press, 374 pp.

GHIL, M., 1989: Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. – Dynam. Atmos. Oceans **13**, 171–218.

GHIL, M., S. COHN, J. TAVANTZIS, K. BUBE, E. ISAACSON, 1981: Applications of Estimation Theory to Numerical Weather Prediction. – In: Dynamic Meteorology: Data Assimilation Methods, L. BENGTSSON, M. GHIL, E. KÄLLÉN (Eds.) Springer, 139–224.

HAMILL, T.M., 2006: Ensemble-based atmospheric data assimilation. – In: Predictability of Weather and Climate, T. Palmer, R. Hagedorn (Eds.), Cambridge University Press, Cambridge, , pp. 124–156.

HAMILL, T.M., C. SNYDER, 2002: Using Improved Background-Error Covariances from an Ensemble Kalman Filter for Adaptive Observations. – Mon. Wea. Rev. **130**, 1552–1572.

HAMILL, T.M., J.S. WHITAKER, C. SNYDER, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. – Mon. Wea. Rev. **129**, 2776–2790.

HAMILL, T.M., C. SNYDER, R.E. MORSS, 2002: Analysis–Error Statistics of a Quasigeostrophic Model Using Three-Dimensional Variational Assimilation. – Mon. Wea. Rev. **130**, 2777–2790.

HAMILL, T.M., C. SNYDER, J.S. WHITAKER, 2003: Ensemble Forecasts and the Properties of Flow-Dependent Analysis-Error Covariance Singular Vectors. – Mon. Wea. Rev. **131**, 1741–1758.

HANSEN, J.A., 2002: Accounting for Model Error in Ensemble-Based State Estimation Forecasting. – Mon. Wea. Rev. **130**, 2373–2391.

HARLIM, J., B.R. HUNT, 2007: Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model. – Tellus **59**A, 731–748.

HELD, I.M., M.J. SUAREZ, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. – Bull. Amer. Meteor. Soc. **75**, 1825–1830.

HOLLOWAY, G., 1986: Eddies, Waves, Circulation, and Mixing: Statistical Geofluid Mechanics. – Ann. Rev. Fluid Mech. **18**, 91–147.

HORN, R.A., C.R. JOHNSON, 1991: Topics in Matrix Analysis. – Cambridge Univ. Press, 607 pp.

HOUGHTON, J., 2005: The Physics of Atmospheres, (3rd ed.). – Cambridge University Press, 320 pp.

HOUTEKAMER, P.L., H.L. MITCHELL, 1998: Data assimilation using an ensemble Kalman filter technique. – Mon. Wea. Rev. **126**, 796–811.

—, —, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. – Mon. Wea. Rev. **129**, 123–137.

—, —, 2005: Ensemble Kalman filtering. – Quart. J. Roy. Meteor. Soc. **131**, 3269–3289.

HOUTEKAMER, P.L., H.L. MITCHELL, G. PELLERIN, M. BUEHNER, M. CHARRON, L. SPACEK, B. HANSEN, 2005: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. – Mon. Wea. Rev. **133**, 604–620.

HUNT, B.R., E.J. KOSTELICH, I. SZUNYOGH, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. – Physica D **230**, 112–126.

IDE, K., P. COURTIER, M. GHIL, A.C. LORENC, 1997: Unified notation for data assimilation: Operational, sequential and variational. – J. Meteor. Soc. Jap. **75** (1B), 181–189.

JACOBSON, M.Z., 2005: Fundamentals of Atmospheric Modeling. – Cambridge University Press, 813 pp.

JANJIĆ, T., S.E. COHN, 2006: Treatment of observation error due to unresolved scales in atmospheric data assimilation. – Mon. Wea. Rev. **134**, 2900–2915.

JAZWINSKI, A.H., 1970: Stochastic Processes and Filtering Theory. – Academic Press, 376 pp.

JOHNSON, C., B.J. HOSKINS, N.K. NICHOLS, S.P. BALLARD, 2006: A Singular Vector Perspective of 4DVAR: The Spatial Structure and Evolution of Baroclinic Weather Systems. – Mon. Wea. Rev. **134**, 3436–3455.

KALMAN, R.E., 1960: A New Approach to Linear Filtering and Prediction Problems. – Trans. ASME, Series D, Journal of Basic Engineering **82**, 35–45.

KHARE, S.P., J.L. ANDERSON, 2006: An examination of ensemble filter based adaptive observation methodologies. – Tellus **58**A, 179–195.

KIM, S., G.L. EYINK, J.M. RESTREPO, F.J. ALEXANDER, G. JOHNSON, 2003: Ensemble filtering for nonlinear dynamics. – Mon. Wea. Rev. **131**, 2586–2594.

KLEIDON, A., K. FRAEDRICH, T. KUNZ, F. LUNKEIT, 2003: The atmospheric circulation and states of maximum entropy production. – Geophysical Research Letters **30 (23)**, 2223, DOI:10.1029/2003GL018363.

LANGLAND, R.H., 2005a: Issues in targeted observing. – Quart. J. Roy. Meteor. Soc. **131**, 3409–3425.

—, 2005b: Observation impact during the North Atlantic TReC–2003. – Mon. Wea. Rev. **133**, 2297–2309.

LAWSON, G.W., J.A. HANSEN, 2004: Implications of Stochastic and Deterministic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth. – Mon. Wea. Rev. **132**, 1966–1981.

LE DIMET, F.X., O. TALAGRAND, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. – Tellus **38A**, 97–110.

LEITH, C.E., 1978: Objective methods for weather prediction. – Ann. Rev. Fluid Mech. **10**, 107–128.

LERMUSIAUX, P.F.J., A.R. ROBINSON, 1999: Data Assimilation via Error Subspace Statistical Estimation. Part I: Theory and Schemes. – Mon. Wea. Rev. **127**, 1385–1407.

LEWIS, J.M., J.C. DERBER, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. – Tellus **37**A, 309–322.

LEWIS, J.M., S. LAKSHMIVARAHAN, S. DHALL, 2006: DYNAMIC DATA ASSIMILATION – A LEAST SQUARES APPROACH. – Cambridge University Press, 654 pp.

LORENC, A.C., 2003a: Modelling of error covariances by 4D-Var data assimilation. – Quart. J. Roy. Meteor. Soc. **129**, 3167–3182.

—, 2003b: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. – Quart. J. Roy. Meteor. Soc. **129**, 3183–3203.

LORENC, A.C., F. RAWLINS, 2005: Why does 4D-Var beat 3D–Var? – Quart. J. Roy. Meteor. Soc. **131**, 3247–3257.

LORENZ, E.N., 1982: Atmospheric predictability experiments with a large numerical model. – Tellus **34**, 505–513.

—, 1984: Irregularity: a fundamental property of the atmosphere. – Tellus **36**A, 98–110.

—, 1993: *The Essence of Chaos*. University of Washington Press, 227 pp.

—, 1996: Predictability – A Problem Partly Solved. – In: Proceedings ECMWF Seminar on Predictability, Vol. I, 4–8 September 1995, 1–18.

LORENZ, E.N., K.A. EMANUEL, 1998: Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model. – J. Atmos. Sci. **55**, 399–414.

LYNCH, P., 2006: The Emergence of Numerical Weather Prediction – Richardson's Dream. Cambridge University Press, 279 pp.

MARDIA, K.V., J.T. KENT, J.M. BIBBY, 1982: Multivariate Analysis. – Academic Press, 521 pp.

MARSHALL, J., F. MOLTENI, 1993: Toward a dynamical understanding of planetary-scale flow regimes. – J. Atmos. Sci. **50**, 1792–1818.

MAYBECK, P.S., 1979: Stochastic models, Estimation, and Control, Volume I. Academic Press, 423 pp.

MCGEE, L.A., S.F. SCHMIDT, 1985: Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry. – National Aeronautics and Space Administration Technical Memorandum No. 86847, 21 pp. Available also online at http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19860003843_1986003843.pdf.

MCLAUGHLIN, D., A. O'NEILL, J. DERBER, M. KAMACHI, 2005: Opportunities for enhanced collaboration within the data assimilation community. – Quart. J. Roy. Meteor. Soc. **131**, 3683–3693.

MILLER, R.N., E.F. CARTER, JR., S.T. BLUE, 1999: Data assimilation into nonlinear stochastic models. – Tellus **51**A, 167–194.

MITCHELL, H.L., P.L. HOUTEKAMER, 2000: An Adaptive Ensemble Kalman Filter. – Mon. Wea. Rev. **128**, 416–433.

MITCHELL, H.L., P.L. HOUTEKAMER, G. PELLERIN, 2002: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter. – Mon. Wea. Rev. **130**, 2791–2808.

NERGER, L., W. HILLER, J. SCHRÖTER, 2005: A comparison of error subspace Kalman filters. – Tellus **57**A, 715–735.

ORRELL, D., 2003: Model error and predictability over different timescales in the Lorenz '96 systems. – J. Atmos. Sci.**60**, 2219–2228.

—, 2005: Ensemble Forecasting in a System with Model Error. – J. Atmos. Sci. **62**, 1652–1659.

ORRELL, D., L. SMITH, J. BARKMEIJER, T.N. PALMER, 2001: Model error in weather forecasting. – Nonlinear Processes in Geophysics **8**, 357–371.

OTT, E., B.R. HUNT, I. SZUNYOGH, A.V. ZIMIN, E.J. KOSTELICH, M. CORAZZA, E. KALNAY, D.J. PATIL, J.A. YORKE, 2004: A local ensemble Kalman filter for atmospheric data assimilation. – Tellus **56**A, 415–428.

PALMER, T.N., 2000: Predicting uncertainty in forecasts of weather and climate. – Rep. Prog. Phys. **63**, 71–116.

PATIL, D.J., B.R. HUNT, E. KALNAY, J.A. YORKE, E. OTT, 2001: Local low dimensionality of atmospheric dynamics. – Phys. Rev. Lett. **86**, 5878–5881.

PHAM, D.T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. – Mon. Wea. Rev. **129**, 1194–1207.

PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, B.P. FLANNERY, 1992: Numerical Recipes in FORTRAN – The Art of Scientific Computing. – Cambridge University Press, 963 pp.

RABIER, F., A. MCNALLY, E. ANDERSSON, P. COURTIER, P. UNDÉN, J. EYRE, A. HOLLINGSWORTH, F. BOUTTIER, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D–Var). II: Structure functions. – Quart. J. Roy. Meteor. Soc. **124**, 1809–1829.

RODWELL, M.J., T.N. PALMER, 2007: Using numerical weather prediction to assess climate models. – Quart. J. Roy. Meteor. Soc. **133**, 129–146.

SEIFFERT, R., R. BLENDER, K. FRAEDRICH, 2006: Subscale forcing in a global atmospheric circulation model and stochastic parametrization. – Quart. J. Roy. Meteor. Soc. **132**, 1627–1643.

SIMMONS, A.J., A. HOLLINGSWORTH, 2002: Some aspects of the improvement in skill in numerical weather prediction. – Quart. J. Roy. Meteor. Soc. **128**, 647–677.

SIMMONS, A.J., R. MUREAU, T. PETROLIAGIS, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. – Quart. J. Roy. Meteor. Soc. **121**, 1739–1771.

SIVIA, D.S., J. SKILLING, 2006: Data Analysis – A Bayesian Tutorial. – Oxford University Press, 246 pp.

SMITH, K.W., 2007: Cluster ensemble Kalman filter. – Tellus **59**A, 749–757.

SORENSON, H.W., 1970: Least-squares estimation: from Gauss to Kalman. – IEEE Spectrum **7**, 63–68.

STOFFELEN, A., G.J. MARSEILLE, F. BOUTTIER, D. VASILJEVIC, S. DE HAAN, C. CARDINALI, 2006: ADM–Aeolus Doppler wind lidar Observing System Simulation Experiment. – Quart. J. Roy. Meteor. Soc. **132**, 1927–1947.

SZUNYOGH, I., E.J. KOSTELICH, G. GYARMATI, D.J. PATIL, B.R. HUNT, E. KALNAY, E. OTT, J.A. YORKE, 2005: Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model. – Tellus **57**A, 528–545.

TALAGRAND, O., P. COURTIER, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. – Quart. J. Roy. Meteor. Soc. **113**, 1311–1328.

TAN, D.G.H., E. ANDERSSON, M. FISHER, L. ISAKSEN, 2007: Observing-system impact assessment using a data

assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission. – Quart. J. Roy. Meteor. Soc. **133**, 381–390.

TARANTOLA, A., 2005: Inverse Problem Theory and Methods for Model Parameter Estimation. – Society for Industrial and Applied Mathematics, 342 pp.

TIPPETT, M.K., S.E. COHN, R. TODLING, D. MARCHESIN, 2000: Low–dimensional representation of error covariance. – Tellus **52**A, 533–553.

TIPPETT, M.K., S.E. COHN, 2001: Adjoints and low–rank covariance representation. – Nonlinear Processes in Geophysics **8**, 331–340.

TIPPETT, M.K., J.L. ANDERSON, C.H. BISHOP, T.M. HAMILL, J.S. WHITAKER, 2003: Ensemble square root filters. – Mon. Wea. Rev. **131**, 1485–1490.

TRÉMOLET, Y., 2007a: Incremental 4D-Var convergence study. – Tellus **59**A, 706–718.

—, 2007b: Model-error estimation in 4D–Var. – Quart. J. Roy. Meteor. Soc. **133**, 1267–1280.

TRIBBIA, J.J., D.P. BAUMHEFNER, 2004: Scale interactions and atmospheric predictability: An updated perspective. – Mon. Wea. Rev. **132**, 703–713.

TUNG, K.K., W.W. ORLANDO, 2003: The $k^{-3}$ and $k^{-5/3}$ Energy Spectrum of Atmospheric Turbulence: Quasi-geostrophic Two-Level Model Simulation. – J. Atmos. Sci. **60**, 824–835.

VAN LEEUWEN, P.J., 1999: Comment on "Data assimilation using an ensemble Kalman filter technique". – Mon. Wea. Rev. **127**, 1374–1377.

—, 2003: A variance-minimizing filter for large-scale applications. – Mon. Wea. Rev. **131**, 2071–2084.

WANG, X., C.H. BISHOP, 2003: A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. – J. Atmos. Sci. **60**, 1140–1158.

WEISSMANN, M., C. CARDINALI, 2007: Impact of airborne Doppler lidar observations on ECMWF forecasts. – Quart. J. Roy. Meteor. Soc. **133**, 107–116.

WEST, M., J. HARRISON, 1989: Bayesian Forecasting and Dynamic Models. – Springer Series in Statistics, Springer, 704 pp.

WHITAKER, J.S., T.M. HAMILL, 2002: Ensemble Data Assimilation without Perturbed Observations. – Mon. Wea. Rev. **130**, 1913–1924.

WHITAKER, J.S., G.P. COMPO, X. WEI, T.M. HAMILL, 2004: Reanalysis without Radiosondes Using Ensemble Data Assimilation. – Mon. Wea. Rev. **132**, 1190–1200.

WILKS, D.S., 2006: Statistical Methods in the Atmospheric Science (2nd ed.), Volume 91, International Geophysics Series of International Geophysics Series. – Elsevier, Academic Press, 627 pp.

WUNSCH, C., 2006: Discrete Inverse and State Estimation Problems With Geophysical Fluid Applications. – Cambridge University Press, 371 pp.

XIONG, X., I.M. NAVON, B. UZUNOGLU, 2006: A note on the particle filter with posterior Gaussian resampling. – Tellus **58**A, 456–460.

ZHANG, F., C. SNYDER, R. ROTUNNO, 2003: Effects of Moist Convection on Mesoscale Predictability. – J. Atmos. Sci. **60**, 1173–1185.

ZUPANSKI, M., S.J. FLETCHER, I.M. NAVON, B. UZUNOGLU, R.P. HEIKES, D.A. RANDALL, T.D. RINGLER, D. DAESCU, 2006: Initiation of ensemble data assimilation. – Tellus **58**A, 159–170.