Nonlinearity in Data Assimilation Applications: A Practical Method for Analysis

M. VERLAAN AND A. W. HEEMINK

Delft University of Technology, Delft, Netherlands

(Manuscript received 8 November 1999, in final form 12 September 2000)

ABSTRACT

A new method to quantify the nonlinearity of data assimilation problems is proposed. The method includes the effects of system errors, measurement errors, observational network, and sampling interval. It is based on computation of the first neglected term in a ''Taylor'' series expansion of the errors introduced by an extended Kalman filter, and can be computed at very little cost when one is already applying a second-order (or higher order) Kalman filter or an ensemble Kalman filter. The nonlinearity measure proposed here can be used to classify the ''hardness'' of the problem and predict the failure of data assimilation algorithms. In this manner it facilitates the comparison of data assimilation algorithms and applications.

The method is applied to the well-known Lorenz model. A comparison is made between several data assimilation algorithms that are suitable for nonlinear problems. The results indicate significant differences in performance for more nonlinear problems. For low values of V, a measure of nonlinearity, the differences are negligible.

1. Introduction and motivation

Data assimilation involving strongly nonlinear models is often considered to be a difficult and mainly unsolved problem (e.g., Miller et al. 1994; Evensen 1997). Although, computation of truly optimal estimates is not feasible for all but extremely small problems, approximate methods often work surprisingly well (Ghil 1997; Fukumori and Melanotte-Rizzoli 1995; Budgell 1986).

However, if in addition to strongly nonlinear dynamics measurements are sparse or inaccurate, then data assimilation algorithms may fail to track transitions and diverge from the true state. Moreover, for larger applications with real data this divergence may be difficult to detect if there is no a priori estimate of the expected performance. Divergence may, for example, also be due to inappropriate system noise or systematic errors in the observations.

Divergence caused by failure of the data assimilation algorithm has been studied by various authors (see, e.g., Miller et al. 1994, 1999; Evensen 1997). It was observed that the divergence is influenced by the sampling frequency, and the accuracy of the observations. The propagation of the error covariance in traditional Kalman filters is based on the use of a tangent linear approximation of the model. This method is only approximate for nonlinear models. However if the model is only weakly nonlinear, the approximation may still be useful. For a strongly nonlinear model one may be able to control the evolution of the error covariance with a large number of observations or very accurate observations. Evensen (1992) showed that for a quasigeostrophic (QG) model the instabilities of the tangent linear model cause an unbounded growth of the computed error variances, since nonlinear saturation effects were absent. Using a large number of obervations, one can however control these instabilities. Several extensions for Kalman filtering and Kalman smoothing algorithms have been proposed to overcome the problems related to the linear time propagation of the error covariance in traditional Kalman filters (Evensen and van Leeuwen 1995, 2000; Evensen 1997; Kushner 1967a,b; Cohn 1993). The comparison of algorithms between authors is often difficult because different parameter values and different numerical schemes are used.

In this work a method is proposed to measure the nonlinearity of a data assimilation problem quantitatively. Since this nonlinearity number attempts to incorporate all of the important effects, the method can be used to compare algorithms and experiments and possibly even different applications. The aim of the project is to predict the accuracy of the data assimilation algorithm associated with the nonlinearity for a particular application without the use of twin experiments. The use of such a method should allow a careful comparison of the pros and cons of various algorithms for a specific application.

Corresponding author address: Dr. M. Verlaan, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, Netherlands.
E-mail: M.Verlaan@math.tudelft.nl

2. Nonlinear data assimilation

Nonlinear data assimilation can be defined as data assimilation where either the model is nonlinear or the relation between model variables and the measured quantity is nonlinear. Although some of the concepts used can be extended to nonlinearities in the measurement functional, the following will deal entirely with nonlinearity of the model equations.

Consider a nonlinear discrete time system given by

$$\mathbf{x}(k+1) = \mathcal{F}[\mathbf{x}(k), \mathbf{u}(k), \mathbf{w}(k)] \quad \text{and} \tag{1}$$

$$\mathbf{y}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{v}(k). \tag{2}$$

Here $\mathbf{x}(\cdot)$ denotes the state vector, and $\mathbf{u}(\cdot)$ the model input. To allow for model errors an additional input vector $\mathbf{w}(\cdot)$ is constructed. These control variables, known as system noise in Kalman filtering literature, are considered unknown a priori. The numerical model is denoted mathematically as a map $\mathcal{F}$ of one discrete time step to the next. The relation between model variables $\mathbf{x}(\cdot)$ and measured quatities $\mathbf{y}(\cdot)$ is given by Eq. (2), which was assumed to be linear for clarity of exposition. Nonlinear measurement operators may be treated in a similar manner as the nonlinear time propagation. Finally, $\mathbf{v}(\cdot)$ denotes measurement errors including representation errors as, for example, arrising from insufficient resolution near the measurement location. Many data assimilation algorithms make use of the tangent linear model (TLM):

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k) + \mathbf{F}(k)\mathbf{w}(k), \tag{3}$$

which is just a local linearization of Eq. (1).

a. A Bayesian approach

One popular framework for performing data assimilation is provided by optimization of some function of the a posteriori probability density function (van Leeuwen and Evensen 1996; Evensen and van Leeuwen 2000). Many existing methods, such as optimal interpolation, three- and four-dimensional variational assimilation, and Kalman filtering, can be cast within this framework. These methods however are all approximate, so that the algorithms may deviate from the optimal solution or even fail when the nonlinearities become too severe.

In Bayesian methods, as well an other statistical methods, the statistics of the model errors and measurement errors are assumed to be known. For the system of Eqs. (1) and (2) it is assumed here that $\mathbf{w}(\cdot)$ and $\mathbf{v}(\cdot)$ are independent Gaussian white noise processes with covariances $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_o$, respectively. In addition some uncertainty for the initial condition $\mathbf{x}(0)$ is assumed, which is Gaussian with mean $E\mathbf{x}(0) = \mathbf{x}_0$ and covariance $\mathbf{P}_0$.

b. Kalman filtering for nonlinear models

The Kalman filter provides an algorithm for computation of a statistically optimal estimate for linear models (Kalman 1960). For weakly nonlinear applications the algorithm can be modified by a local Taylor series expansion around the current estimate. The popular extended Kalman filter is an example of a first-order approximation. Many algorithms based on this principle have been published (see, e.g., Jazwinski 1970). Since the computational requirements quickly grow with the order of the expansion, the number of applications reported in the literature rapidly decreases with the order of accuracy.

A disadvantage of the Kalman filter is that for high-dimensional systems the computational burden quickly grows beyond what is feasible. Therefore several suboptimal schemes (SOSs) have been proposed recently. Although the experiments in this work all have a small state dimension, only schemes that can be applied for large numerical models were used since our intended area of application is large-scale data assimilation.

c. A few approximate nonlinear methods

1) The ensemble Kalman filter

The first algorithm considered here is the ensemble Kalman filter (EnKF) (Evensen 1994; Burgers et al. 1998; Evensen and van Leeuwen 1996). This algorithm computes the forecast error covariance by integrating an ensemble of randomly perturbed initial analysis states in time with random perturbations added to the forcing. This Monte Carlo–type approach based on the full nonlinear model allows for consistent statistics in the case of nonlinear dynamics. The analysis of the pertubed states, known as ensemble members, is carried out with perturbed observations. For the system of Eqs. (1) and (2) the algorithm can be denoted as

$$\boldsymbol{\xi}_i(k+1|k) = \mathcal{F}[\boldsymbol{\xi}_i(k|k), \mathbf{u}(k), \mathbf{w}_i(k)] \tag{4}$$

$$\hat{\mathbf{x}}(k+1|k) = (1/q)\sum_{i=1}^{q} \boldsymbol{\xi}_i(k+1|k) \tag{5}$$

$$[\mathbf{L}_c(k+1|k)]_{1:n,i} = (1/\sqrt{q})[\boldsymbol{\xi}_i(k+1|k) - \hat{\mathbf{x}}(k+1|k)] \tag{6}$$

$$\begin{aligned}\mathbf{K}_c(k+1) = {} & \mathbf{L}_c(k+1|k)\mathbf{L}_c(k+1|k)^{\mathrm{T}}\mathbf{C}(k+1)^{\mathrm{T}} \\ & \times [\mathbf{C}(k+1)\mathbf{L}_c(k+1|k) \\ & \times \mathbf{L}_c(k+1|k)^{\mathrm{T}}\mathbf{C}(k+1)^{\mathrm{T}} \\ & + \mathbf{\Sigma}_o(k+1)]^{-1}\end{aligned} \tag{7}$$

$$\begin{aligned}\boldsymbol{\xi}_i(k+1|k+1) = {} & \boldsymbol{\xi}_i(k+1|k) \\ & + \mathbf{K}_c[\mathbf{y}(k+1) - \mathbf{C}(k)\boldsymbol{\xi}_i(k+1|k) \\ & - \mathbf{v}_i(k+1)],\end{aligned} \tag{8}$$

where $\boldsymbol{\xi}_i(k|l)$ is an ensemble of state vectors generated with the realizations $\mathbf{w}_i$ and $\mathbf{v}_i$ of the processes $\mathbf{w}$ and $\mathbf{v}$, respectively. These realizations are made using a

pseudorandom generator. Note that $\hat{\mathbf{x}}$ for finite samples depends on the actual realization used; that is, different random number generators or different initial seed values will give (slightly) different results.

2) The reduced rank square root filter (EKF version)

The reduced rank square root (RRSQRT) algorithm (Verlaan 1998; Verlaan and Heemink 1995) is an algorithm that is based on an eigendecomposition of the error covariance matrix. The leading eigenvectors of the analysis error covariance are used to form an ensemble of perturbed state vectors, which are integrated in time with the full nonlinear model. This ensemble has the property that for linear dynamics the covariance in the subspace spanned by the forecasts of the ensemble, relative to the central forecast, is exact. For nonlinear dynamics the method is approximate and somewhat comparable to the extended Kalman filter. For scalar measurements the algorithm can be denoted as

$$\hat{\mathbf{x}}(k+1|k) = \mathcal{F}[\hat{\mathbf{x}}(k|k), \mathbf{u}(k), 0] \tag{9}$$

$$\boldsymbol{\xi}_i(k|k) = \hat{\mathbf{x}}(k|k) + \delta[\mathbf{L}_c(k|k)]_{1:n,i} \tag{10}$$

$$\boldsymbol{\xi}_i(k+1|k) = \mathcal{F}[\boldsymbol{\xi}_i(k|k), \mathbf{u}(k), \mathbf{w}_i(k)] \tag{11}$$

$$[\mathbf{L}_c(k+1|k)]_{1:n,i} = (1/\delta)[\boldsymbol{\xi}_i(k+1|k) - \hat{\mathbf{x}}(k+1|k)] \tag{12}$$

$$\mathbf{L}_c(k+1|k)^{\mathrm{T}}\mathbf{L}_c(k+1|k) = \mathbf{U}_c(k+1)\mathbf{D}_c(k+1)\mathbf{U}_c(k+1)^{\mathrm{T}} \tag{13}$$

$$\mathbf{L}_c^*(k+1|k) = [\mathbf{L}_c(k+1|k)\mathbf{U}_c(k+1)]_{1:n,1:q} \tag{14}$$

$$\mathbf{H}(k+1) = \mathbf{L}_c^*(k+1|k)'\mathbf{C}(k+1)' \tag{15}$$

$$\boldsymbol{\beta}(k+1) = [\mathbf{H}(k+1)^{\mathrm{T}}\mathbf{H}(k+1) + \boldsymbol{\Sigma}_o(k+1)]^{-1} \tag{16}$$

$$\mathbf{K}_c(k+1) = \mathbf{L}_c^*(k+1|k)\mathbf{H}(k+1)\boldsymbol{\beta}(k+1) \tag{17}$$

$$\hat{\mathbf{x}}(k+1|k+1) = \hat{\mathbf{x}}(k+1|k) + \mathbf{K}_c(k+1) \times [\mathbf{y}(k+1) - \mathbf{C}(k+1)\hat{\mathbf{x}}(k+1|k)] \tag{18}$$

$$\mathbf{L}_c(k+1|k+1) = \mathbf{L}_c^*(k+1|k) - \mathbf{K}_c(k+1)\mathbf{H}(k+1)^{\mathrm{T}} \times \{\mathbf{I} + [\boldsymbol{\beta}(k+1)\boldsymbol{\Sigma}_o(k+1)]^{1/2}\}^{-1}, \tag{19}$$

where $\mathbf{L}_c(k|k)$ is the n by q estimate square root of the error covariance $\mathbf{P}(k|k)$ and $[\,\cdot\,,\cdot\,]$ denotes that the large matrix is built from the two block matrices [thus $\mathbf{L}_c(k+1|k)$ has $q+m$ columns]. The parameter δ controls the size of the perturbations of the state vectors that are propagated in time. It does not influence the results for linear models but for nonlinear models it does. In the limit for very small values the method converges to an approach with a tangent linear model. Experience has shown that $\delta = 1$ is often a reasonable value. The notation has been chosen in a way that stresses the similarities with the EnKF.

3) Two second-order extensions of the RRSQRT filter

For nonlinear models computation of the forecast error covariances with the TLM becomes approximate. In addition to this the traditional forecast equation [Eq. 9] becomes biased. This can be shown with a Taylor series expansion around the analysis:

$$\mathcal{F}(\mathbf{x}) = \mathcal{F}[\hat{\mathbf{x}}(k|k)] + \frac{\partial \mathcal{F}}{\partial \mathbf{x}}[\mathbf{x} - \hat{\mathbf{x}}(k|k)] + \frac{1}{2}\frac{\partial^2 \mathcal{F}}{\partial \mathbf{x}^2}[\mathbf{x} - \hat{\mathbf{x}}(k|k)] + \cdots. \tag{20}$$

Computing the mean value for this expresion one obtains (see, e.g., Jazwinski 1970)

$$\overline{\mathcal{F}(x)} = \mathcal{F}[\hat{\mathbf{x}}(k|k)] + \frac{1}{2}\frac{\partial^2 \mathcal{F}}{\partial \mathbf{x}^2}\mathbf{P}(k|k) + \cdots. \tag{21}$$

Thus when comparing a forecast based on one central analysis with the mean of an ensemble of forecasts based on perturbed initial conditions, several systematic differences occur. The leading term of these differences is known as the bias correction term. The well-known second-order truncated Kalman filter consists of an extended Kalman filter (EKF) with this term as the only difference.

For the SOSs used here the bias correction term can be computed when a symmetric (deterministic) ensemble is used instead of Eq. (10) (Verlaan 1998; Segers et al. 1999):

$$\boldsymbol{\xi}_i(k|k) = \hat{\mathbf{x}}(k|k) + \delta[\mathbf{L}_c(k|k)]_{1:n,i} \tag{22}$$

$$\boldsymbol{\xi}_{q+i}(k|k) = \hat{\mathbf{x}}(k|k) - \delta[\mathbf{L}_c(k|k)]_{1:n,i}. \tag{23}$$

This sample is twice the size and thus computationally more expensive. Much of the remaining algorithm is the same as the first-order algorithm. An important modification is the bias correction:

$$\hat{\mathbf{x}}^*(k+1|k) = \hat{\mathbf{x}}(k+1|k) + \sum_{i=1}^{2q} \frac{\boldsymbol{\xi}_i(k+1|k) - \hat{\mathbf{x}}(k+1|k)}{\delta^2}. \tag{24}$$

The corrected estimate $\hat{\mathbf{x}}^*(k+1|k)$ is used subsequently in Eq. (18). It should be noted that Eq. (24) is just a finite-difference approximation to Eq. (21).

A similar algorithm was proposed by Pham et al. (1998). In the singular evolutive interpolated Kalman (SEIK) filter a minimal sample of $q+1$ members is constructed that still exactly preserves the error covariance and mean. It can be shown that this algorithm is equivalent to a second-order truncated Kalman filter.

In order to construct a ''second order'' sample with $q + 1$ members,

$$\boldsymbol{\xi}_i(k|k) = \hat{\mathbf{x}}(k|k) + \delta\boldsymbol{\omega}_i^{\mathrm{T}}[\mathbf{L}_c(k|k)], \tag{25}$$

the weights $\omega_{i,j}$ have to satisfy some conditions; that is, the matrix

$$\boldsymbol{\Omega} \equiv \begin{bmatrix} 1 & \boldsymbol{\omega}_1^{\mathrm{T}} \\ 1 & \vdots \\ 1 & \boldsymbol{\omega}_{q+1}^{\mathrm{T}} \end{bmatrix} \tag{26}$$

must have orthogonal columns, which can be achieved with Householder reflections. The bias correction in this algorithm becomes

$$\hat{\mathbf{x}}^*(k+1|k) = \hat{\mathbf{x}}(k+1|k) + \sum_{i=1}^{q+1} \frac{\boldsymbol{\xi}_i(k+1|k) - \hat{\mathbf{x}}(k+1|k)}{\delta^2}. \tag{27}$$

Here the algorithm was generalized with respect to the original formulation. The original algorithm can be retrieved by selecting $\delta = \sqrt{q+1}$.

3. Measuring nonlinearity

In order to quantify the relative importance of nonlinear system dynamics in data assimilation applications a method is proposed here that is based on the combined application of a first-order extended and a truncated second-order Kalman filter. For near-linear applications the results of both methods will nearly coincide while for more nonlinear applications the differences will be larger, where the second-order results are likely to be more accurate. The magnitude of this divergence between the two algorithms can be compared to the error covariance, which indicates the magnitude of the errors if the algorithms were accurate. Let **b** denote the difference between the estimated state vectors for both algorithms, then for the scalar case, $\boldsymbol{b}/\sigma$ is likely to be a measure for the nonlinearity of the data assimilation problem. For state vectors of larger dimensions an extension of this expression is needed.

The only difference between the extended Kalman filter and the truncated second-order filter is a bias term in the time propagation of the truncated second-order filter [see Eq. (24)]. To obtain an algorithm that can recover after divergence it is more convenient to derive an equation for the bias directly. Although this equation has the same order of accuracy it is more likely to recover after divergence because the linearisation is now around the estimate of the more accurate second-order algorithm.

Let the error $\mathbf{e}(k|l)$ be defined as

$$\mathbf{e}(k|l) \equiv \mathbf{x}(k) - \hat{\mathbf{x}}(k|l), \tag{28}$$

where $\mathbf{x}(k)$ denotes the unknown true value and $\hat{\mathbf{x}}(k|l)$ the estimate computed with a data assimilation method. The time propagation and measurement propagation for the bias for the extended Kalman filter, that is, $\mathbf{b}(k|l) \equiv E[\mathbf{e}(k|l)]$, can easily be derived and are given by

$$\mathbf{b}(k+1|k) = \frac{\partial \mathcal{F}}{\partial \mathbf{x}}\mathbf{b}(k|k) + \frac{\partial^2 \mathcal{F}}{\partial \mathbf{x}^2}\mathbf{P}_c(k|k) \tag{29}$$

$$\mathbf{b}(k+1|k+1) = (\mathbf{I} - \mathbf{KC})\mathbf{b}(k+1|k), \tag{30}$$

where third- and higher-order terms in the time propagation have been neglected. The initial condition is $\mathbf{b}(0|0) = 0$. By evaluating the derivatives $\partial\mathcal{F}/\partial\mathbf{x}$ and $\partial^2\mathcal{F}/\partial\mathbf{x}^2$ around the second-order estimate the bias is likely to be more accurate. Also these additional computations for the bias can be performed at little additional cost if a truncated second-order filter is applied.

When the estimate $\hat{\mathbf{x}}(k|l)$ is biased, the expected magnitude of the errors is not determined by the covariance alone. A good estimate is given by Dee and da Silva (1998):

$$E(\mathbf{e}\mathbf{e}^{\mathrm{T}}) = \mathbf{P} + \mathbf{b}\mathbf{b}^{\mathrm{T}}. \tag{31}$$

As can be seen from this equation the bias contributes to the errors only in one direction (along the vector **b**). The relative importance in this direction can be defined in several ways.

One possibility for defining the relative importance of the bias compared to the random errors is by testing against a hypothesis of no bias. For unbiased normally distributed random variables with covariance **P** the log likelihood, disregarding some constants, becomes

$$L(\mathbf{e}) = \mathbf{e}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{e}. \tag{32}$$

It is well known that L has a χ_n^2 distibution if the hypothesis holds. However if **e** is biased, then the expected value of $L(\mathbf{e})$ becomes

$$E[L(\mathbf{e})] = E\{[\mathbf{e} - E(\mathbf{e})]^{\mathrm{T}}\mathbf{P}^{-1}[\mathbf{e} - E(\mathbf{e})]\} + E[E(\mathbf{e})^{\mathrm{T}}\mathbf{P}^{-1}E(\mathbf{e})] \tag{33}$$

$$= n + \mathbf{b}'\mathbf{P}^{-1}\mathbf{b}, \tag{34}$$

where it was assumed that $E\{[\mathbf{e} - E(\mathbf{e})][\mathbf{e} - E(\mathbf{e})]^{\mathrm{T}}\} = \mathbf{P}$. This shows that if $\mathbf{b}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{b} \ll n$, then the bias is very hard to detect, while if $\mathbf{b}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{b} \gg n$, the hypothesis of no bias is almost certainly rejected. Based on this motivation the following nondimensional number is defined:

$$V \equiv \sqrt{\mathbf{b}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{b}}, \tag{35}$$

where it is likely that the bias is significant if $V \gg \sqrt{n}$ and insignificant if $V \ll \sqrt{n}$.

It is easy to show that V is a nondimensional number, that is, it is not affected by scaling of the state variables or by any other linear state transformation. To show these properties, consider the linear state transformation $\mathbf{z} \equiv \mathbf{T}\mathbf{x}$. By definition the transformed covariance becomes

$$\tilde{\mathbf{P}} \equiv E(\tilde{\mathbf{e}}\tilde{\mathbf{e}}^{\mathrm{T}}) \tag{36}$$

$$= E(\mathbf{T}\mathbf{e}\mathbf{e}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}) \tag{37}$$

$$= \mathbf{T}\mathbf{P}\mathbf{T}^{\mathrm{T}}. \tag{38}$$

Similarly the bias transforms as

$$\tilde{\mathbf{b}} = E(\tilde{\mathbf{e}}) \tag{39}$$

$$= \mathbf{T}E(\mathbf{e}), \tag{40}$$

where the tilde denotes the tranformed value. Combining the expressions in Eqs. (38) and (40) yields

$$\tilde{V}^2 = \tilde{\mathbf{b}}^{\mathrm{T}}\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{b}} = \mathbf{b}^{\mathrm{T}}\mathbf{P}\mathbf{b} = V^2, \tag{41}$$

which shows that V is not affected by a linear state transformation if the bias and forecast error covariance are computed exactly. If the bias and forecast error covariance are approximated using a truncated second filter, is not so obvious that V is not affected by a linear state transformation.

Consider the transformation of the equations for the covariance for forecast and analysis in a truncated second-order filter; that is,

$$\tilde{\mathbf{P}}(k+1|k) = \mathbf{T}\mathbf{P}(k+1|k)\mathbf{T}^{\mathrm{T}} \tag{42}$$

$$= \mathbf{T}\left[\frac{\partial \mathcal{F}}{\partial \mathbf{x}}\mathbf{P}(k|k)\frac{\partial \mathcal{F}^{\mathrm{T}}}{\partial \mathbf{x}} + \frac{\partial \mathcal{F}}{\partial \mathbf{w}}\boldsymbol{\Sigma}_s\frac{\partial \mathcal{F}}{\partial \mathbf{w}}\right]\mathbf{T}^{\mathrm{T}} \tag{43}$$

$$= \left(\mathbf{T}\frac{\partial \mathcal{F}}{\partial \mathbf{x}}\mathbf{T}^{-1}\right)\tilde{\mathbf{P}}(k|k)\left(\mathbf{T}\frac{\partial \mathcal{F}}{\partial \mathbf{x}}\mathbf{T}^{-1}\right)^{\mathrm{T}} + \left(\mathbf{T}\frac{\partial \mathcal{F}}{\partial \mathbf{w}}\right)\boldsymbol{\Sigma}_s\left(\mathbf{T}\frac{\partial \mathcal{F}}{\partial \mathbf{w}}\right)^{\mathrm{T}} \tag{44}$$

$$= \frac{\partial \tilde{\mathcal{F}}}{\partial \mathbf{z}}\tilde{\mathbf{P}}(k|k)\frac{\partial \tilde{\mathcal{F}}^{\mathrm{T}}}{\partial \mathbf{z}} + \frac{\partial \tilde{\mathcal{F}}}{\partial \mathbf{w}}\boldsymbol{\Sigma}_s\frac{\partial \tilde{\mathcal{F}}^{\mathrm{T}}}{\partial \mathbf{w}} \tag{45}$$

$$\tilde{\mathbf{P}}(k+1|k+1) = \mathbf{T}\mathbf{P}(k+1|k+1)\mathbf{T}^{\mathrm{T}} \tag{46}$$

$$= \mathbf{T}[\mathbf{P}(k+1|k) - \mathbf{K}(k+1)\mathbf{C}(k+1)^{\mathrm{T}}] \times \mathbf{P}(k+1|k)\mathbf{T}^{\mathrm{T}} \tag{47}$$

$$= [\tilde{\mathbf{P}}(k+1|k) - \tilde{\mathbf{K}}(k+1)\tilde{\mathbf{C}}(k+1)^{\mathrm{T}}] \times \tilde{\mathbf{P}}(k+1|k)\mathbf{T}^{\mathrm{T}} \tag{48}$$

$$\tilde{\mathbf{K}}(k+1) = \mathbf{T}\mathbf{K}(k+1) \tag{49}$$

$$= \mathbf{T}\{\mathbf{P}(k+1|k)\mathbf{C}(k+1)^{\mathrm{T}} \times [\mathbf{C}(k+1)\mathbf{P}(k+1|k)\mathbf{C}(k+1)^{\mathrm{T}} + \boldsymbol{\Sigma}_o(k+1)]^{-1}\} \tag{50}$$

$$= \{\tilde{\mathbf{P}}(k+1|k)\mathbf{C}(k+1)' \times [\tilde{\mathbf{C}}(k+1)\tilde{\mathbf{P}}(k+1|k)\tilde{\mathbf{C}}(k+1)^{\mathrm{T}} + \boldsymbol{\Sigma}_o(k+1)]^{-1}\}. \tag{51}$$

Thus, by using an induction argument, the approximate covariance computed with this algorithm transforms according to $\tilde{\mathbf{P}} = \mathbf{T}\mathbf{P}\mathbf{T}^{\mathrm{T}}$ as in the exact case. The same can be shown for the bias equations (29) and (30):

$$\tilde{\mathbf{b}}(k+1|k) = \mathbf{T}\mathbf{b}(k+1|k) \tag{52}$$

$$= \mathbf{T}\left[\frac{\partial \mathcal{F}}{\partial \mathbf{x}}\mathbf{b}(k|k) + \frac{\partial^2 \mathcal{F}}{\partial \mathbf{x}^2}\mathbf{P}_c(k|k)\right]\mathbf{T}^{\mathrm{T}} \tag{53}$$

$$= \frac{\partial \tilde{\mathcal{F}}}{\partial \mathbf{z}}\tilde{\mathbf{b}}(k|k) + \frac{\partial^2 \tilde{\mathcal{F}}}{\partial \mathbf{z}^2}\tilde{\mathbf{P}}_c(k|k) \tag{54}$$

$$\tilde{\mathbf{b}}(k+1|k+1) = \mathbf{T}\mathbf{b}(k+1|k+1) \tag{55}$$

$$= \mathbf{T}\{[\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)^{\mathrm{T}}] \times \mathbf{b}(k+1|k)\} \tag{56}$$

$$= [\mathbf{I} - \tilde{\mathbf{K}}(k+1)\tilde{\mathbf{C}}(k+1)] \times \tilde{\mathbf{b}}(k+1|k); \tag{57}$$

thus, $\tilde{\mathbf{b}} = \mathbf{T}\mathbf{b}$ as for the exact case. Combining these two expressions shows that $\tilde{V} = V$ also for the truncated second-order filter. Scaling with the standard deviation of each variable can be interpreted as a special linear transform. Since all variables will become nondimensional with this transform and since V is not affected, V must be nondimensional.

It should be noted that the definition of V used here is not the only way to quantify the relative importance of the nonlinear bias errors to the random error. An alternative is, for example, $V_2 \equiv \sqrt{(1/n)\ \Sigma_i\,(b_i/\sigma_i)^2}$, where the summation is over the elements of the state vector. Many other possibilities exist and much work still needs to be done to find the most valuable ones.

Since the nonlinearity number V is based on a first-order-accurate and a second-order-accurate algorithm it is expected to be valid for problems for which both methods perform reasonably well. If the nonlinearity of the problem becomes larger and these algorithms fail, it is likely that V is still large since the bias equation has a large source term. However, the magnitude of V will be less accurate. This property makes the present definition of V less useful for accurately computing the magnitude of nonlinearity for problems for which the extended Kalman filter fails completely. On the other hand, it can still be used to detect failure since V will be large for this type of problems.

4. Test case 1: The Burgers equation

a. Introduction

Consider the following stochastic extension of the Burgers equation:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2} \tag{58}$$

$$u|_{x=0} = u_c + w(t) \tag{59}$$

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{x=0} = 0 \tag{60}$$

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{x=L} = 0 \tag{61}$$

$$u|_{t=0} = u_c, \tag{62}$$

where the mean flow with velocity $u_c = 1$ m s^{-1} is perurbed with a randon forcing from the left boundary ω. This stochastic forcing has an exponential correlation in time with e-folding time τ. A similar example was analyzed in detail by Cohn (1993), where it was shown that the viscosity term is neglected then the correlation structure is completely determined by the initial and boundary conditions. Thus here the spatial correlation scale after the initial transient becomes τu_c. In this paper (Cohn 1993) equations for the time evolution of the mean and variance were derived. Contrary to the presentation in the previous section, where the model was assumed to be discretized in space and time, this derivation was based directly on the continuum formulation. To illustrate this alternative approach a short and illustrative derivation is shown here.

First substitute $u = \overline{u} + u'$ in Eq. (58), where the overbar denotes an ensemble average and primed values denote the deviation from this mean:

$$\frac{\partial \overline{u} + u'}{\partial t} + (\overline{u} + u')\frac{\partial \overline{u} + u'}{\partial x} = \nu\frac{\partial^2 \overline{u} + u'}{\partial x^2}. \tag{63}$$

In general this step is followed by a Taylor expansion around the mean value. Since Eq. (63) contains only linear terms and one quadratic term the expansion is exact with only these terms. An ensemble averaging now results in the equation for the mean:

$$\frac{\partial \overline{u}}{\partial t} + (\overline{u})\frac{\partial \overline{u}}{\partial x} + \overline{u'\frac{\partial u'}{\partial x}} = \nu\frac{\partial^2 \overline{u}}{\partial x^2} \tag{64}$$

$$\Leftrightarrow \frac{\partial \overline{u}}{\partial t} + (\overline{u})\frac{\partial \overline{u}}{\partial x} + \sigma\frac{\partial \sigma}{\partial x} = \nu\frac{\partial^2 \overline{u}}{\partial x^2}, \tag{65}$$

where σ denotes the standard deviation of the forcast errors. Thus the ensemble average contains one additional term $\sigma\partial\sigma/\partial x$ compared to the original model, also known as central forecast. This bias term is caused by the nonlinearity of the advection term in the Burgers equation. By defining the bias as $b \equiv \overline{u} - \hat{u}$, where $\hat{u}$ is the central forecast and subtracting Eq. (58) from Eq. (65), we obtain

$$\frac{\partial b}{\partial t} + \overline{u}\frac{\partial b}{\partial x} + b\frac{\partial \overline{u}}{\partial x} + \sigma\frac{\partial \sigma}{\partial x} = \nu\frac{\partial^2 b}{\partial x^2} \tag{66}$$

after neglecting the quadratic term in b. It should be noted that this derivation is very similar to what is well known as Reynolds averaging for turbulent velocity fluctuations (see, e.g., Tennekes and Lumley 1972; Pedlosky 1987, p. 181). The bias terms in that case are known as Reynolds stresses.

An expression for the forecast error variance can be obtained by subtracting Eq. (65) from Eq. (63), premultiplying with $2u'$, and computing the ensemble average of this expression. After a little algebraic manipulation this expression can be written as

$$\frac{\partial C}{\partial t} + \overline{u}\frac{\partial C}{\partial x} + 2C\frac{\partial \overline{u}}{\partial x} = \nu\frac{\partial^2 C}{\partial x^2} - 2\nu\overline{\left(\frac{\partial u'}{\partial x}\right)^2}, \tag{67}$$

where C denotes the error variance $\overline{u'^2}$. The term with $\overline{u'^2\partial u'/\partial x}$ is assumed to be zero here, which is a common assumption for Kalman filtering. However for turbulent fluctuations u' this term would be the production term for turbulent kinetic energy and very different closure assumptions are used. In this analogy Eq. (67) would be an equation for the turbulent kinetic energy ($C \approx k$) as, for example, is used in the well-known k–ε model. The term $2C\partial\overline{u}/\partial x$ may cause instabilities. Evensen (1992) observed unbounded growth of the variance when applying an extended Kalman filter to a QG model. The analysis showed similar terms to be the cause. Note also that Eq. (67) does not form a closed set of equations with Eq. (66) for viscous flow. Cohn (1993) showed that for hyperbolic models one can obtain a closed set of equations for mean and variance only, but for the parabolic term local correlations are needed.

In the model described by Eqs. (58)–(62). The initial estimate is assumed to contain no errors. After an initial transient the variance will become constant with a value equal to the variance of the stochastic term in the left boundary. At this time the source term in the bias equation (66) becomes zero. Thus after some time the nonlinearity measure V will converge to zero.

When observations of the velocity are assimilated into the model the behavior of the system becomes more complex. First, it is important to note that the assimilation of direct observations of state variables will reduce the nonlinearity of the problem. One can easily show this effect with the relations derived above. First rewrite the standard Kalman filter analysis as

$$\mathbf{P}(k+1|k+1) = \mathbf{\Phi}\mathbf{P}(k+1|k)\mathbf{\Phi}^{\mathrm{T}} + \mathbf{K}\mathbf{\Sigma}_o\mathbf{K}^{\mathrm{T}}, \tag{68}$$

where $\mathbf{\Phi} \equiv (\mathbf{I} - \mathbf{KC})$. Using this equation and the analysis update for the bias, that is, Eq. (30), one can derive the following expression for V:

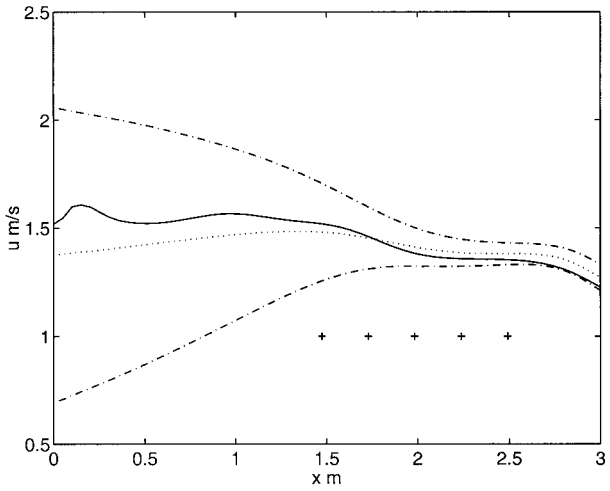

FIG. 1. True velocity (line) at $t = 5$ s, analysis (dotted) and 2σ bounds around the estimate (dashed).

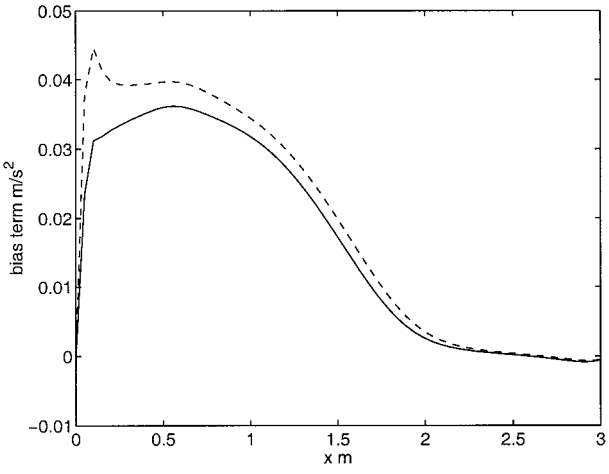

FIG. 2. Source term in the bias equation computed with the RRSQRT Kalman filter (line) and with $\sigma\partial\sigma/\partial x$ (dashed).

$$V(k+1|k+1)^2$$

$$= \mathbf{b}(k+1|k+1)^{\mathrm{T}}\mathbf{P}(k+1|k+1) \times \mathbf{b}(k+1|k+1) \quad (69)$$

$$= \mathbf{b}(k+1|k)^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}[\mathbf{\Phi P}(k+1|k)\mathbf{\Phi}^{\mathrm{T}} + \mathbf{K\Sigma}_o\mathbf{K}^{\mathrm{T}})]^{-1} \times \mathbf{\Phi b}(k+1|k) \quad (70)$$

$$< \mathbf{b}(k+1|k)^{\mathrm{T}}\mathbf{P}(k+1|k)^{-1}\mathbf{b}(k+1|k) \quad (71)$$

$$= V(k+1|k)^2. \quad (72)$$

The general tendency will therefore be a reduction of the nonlinearity by the analysis in well-observed areas. However, the gradient of σ at the edge of a data-rich and data-sparse region will lead to a large growth of bias there. In addition the bias will grow where $\partial\overline{u}/\partial x < 0$, but here also the variance of the errors will grow albeit not as fast. The largest nonlinearities are therefore expected at the end of the forecast.

b. A numerical example with data assimilation

An experiment was carried out with the model described in Eqs. (58)–(62). Observations of the velocity were assimilated at regular intervals in the right half of the domain. Figure 1 shows the true state, the estimated state, and the 2σ error bars, at $t = 5$ s.

The bias term can be computed both using the expression $-\sigma\partial\sigma/\partial x$ and with a second-order Kalman filter. Figure 2 shows the source term of the bias equation for both methods. The differences are numerical and due to the fact that the equality $2u\partial u/\partial x = \partial u^2/\partial x$ is not preserved exactly with the spatial discretization that is used here. As expected the largest values of the source term occur at the edge of the data-dense and data-sparse areas.

The nonlinearity measure at $t = 5$ s is $V \approx 0.1$. It is thus expected that the extended Kalman filter will perform well in this case. This is consistent with the fact that the true state is well within the 2σ bounds. For the computation of V in this application only the part of $\mathbf{b}$ lying in the space spanned by the largest 12 eigenvectors of $\mathbf{P}$ was used, because small eigenvalues of $\mathbf{P}$ may lead to a large value of V while at the same time it is not very likely that a data assimilation will fail on the basis of these differences. Further study is clearly necessary for this aspect. Instead of computing an overall measure, V, it is also possible to compute this value pointwise. For one state variable, $\mathbf{b}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{b}$ reduces to $V = b/\sigma$. In this manner the spatial distribution of the nonlinearities can be studied. Figure 3 shows the value of the bias and standard deviation and Fig. 4 their ratio. The negative value of the bias in the left half of the domain is caused by the fact that the Kalman gain is larger than one there because of the large ratio of the error variances at the left boundary compared to that at the observations locations. As expected the nonlinearities are smaller in

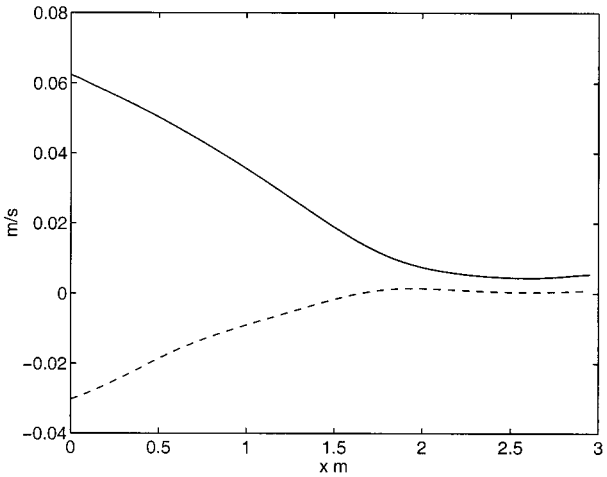

FIG. 3. Bias (dashed) and standard deviation (line).

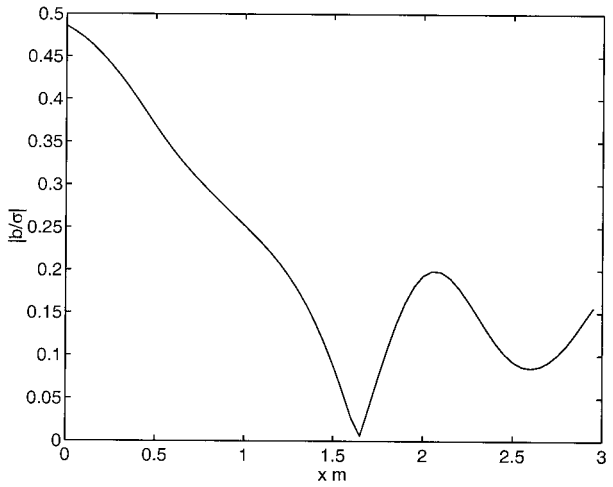

FIG. 4. Local nonlinearity $|b/\sigma|$.

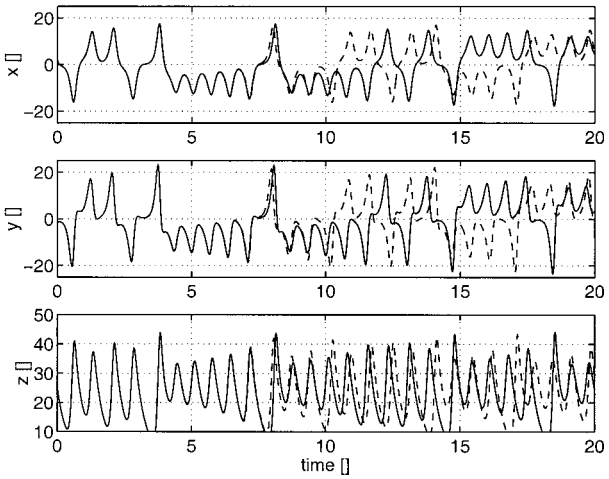

FIG. 5. Two realizations of the Lorenz model with slighty different initial conditions.

the data-dense region due to the linearizing effect of the analysis, as was shown in the previos paragraph.

5. Test case 2: The Lorenz model

a. Model description

A well-known example of a strongly nonlinear system is the Lorenz model. This model is a simplified description of the flow in a thermal convection experiment. For some values of the parameters this system shows chaotic behavior in the sense that very small pertubations in the initial condition will grow into completely different solutions very rapidly in time. These equations were originally derived by Lorenz (Lorenz 1963; Sparrow 1982) and are often used as a prototype example of nonlinear dynamic behavior (see, e.g., Miller et al. 1999). The system of Lorenz equations can be denoted as

$$\frac{dx}{dt} = \sigma(y - x) \tag{73}$$

$$\frac{dy}{dt} = \rho x - y - xz \tag{74}$$

$$\frac{dz}{dt} = xy - \beta z. \tag{75}$$

The following parameter values will be used: $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ together with the initial conditions $(x, y, z) = (1.508\,870, -1.531\,271, 25.460\,91)$. Figure 5 shows that a perturbation of 0.001 in x can alter the trajectory completely after some time.

There are several authors who studied these equations in connection with Kalman filtering or other types of data assimilation. Miller et al. (1994, 1999) showed that the extended Kalman filter fails if the observations are too far apart or if the observations are too inaccurate. A fourth-order extension of the Kalman filter (Kushner 1967a,b) performed much better, but failed also if these parameters were increased further. Of all the practical methods applied, the EnKF performed best. It was also shown that direct minimization of a criterion that is a function of the forecast errors with the model as a strong constraint is extremely difficult due to the many local minima (see also Miller et al. 1994). See also Evensen (1997) for an intercomparison of weak constraint methods and the EnKF for the Lorenz equations, and Evensen and van Leeuwen (2000) for an introduction of the ensemble Kalman smoother, which provides more accurate estimates if observations past the analysis time are available.

b. A twin experiment

The main aims of the experiments in this section are first to show that the measure V indeed has predictive power for the performance of data assimilation methods to nonlinear systems and, second, to compare several extensions of the Kalman filter for a wide range of parameter values.

When comparing various data assimilation methods, it is always very convenient if the ''true'' state of the system is available for determining the performance. For this reason the observations are often generated from a second model with a realization of the measurement noise and system noise with the statistics as used in the assimilation algorithm. This kind of experiment is generally referred to as a (identical) twin experiment.

For use in the twin experiment a stochastic extension of the Lorenz model was constructed. First Eqs. (73)–(75) were discretized using a scheme described in Miller et al. (1999). Subsequently white system noise with covariance $\mathbf{\Sigma}_s = \mathbf{I}\sigma_s^2\Delta t$, with Δt the discrete time step, was added. All state variables were observed at regular intervals T in time. A measurement noise with covariance $\mathbf{\Sigma}_o = \mathbf{I}\sigma_o^2$ was applied. In order to reduce the var-

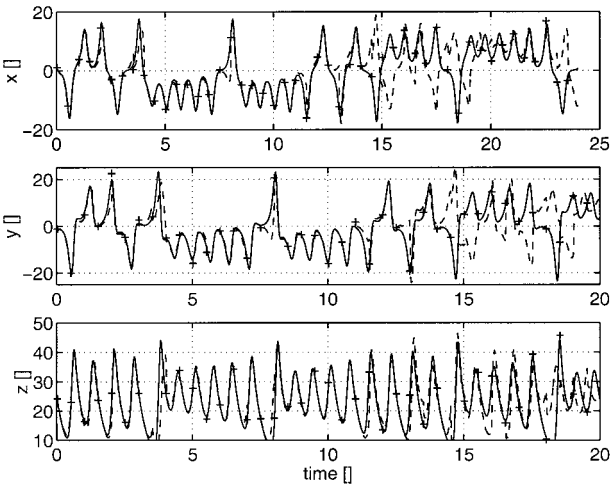

FIG. 6. True (line) and estimated (dashed) state and measurements (+) for an extended Kalman filter.

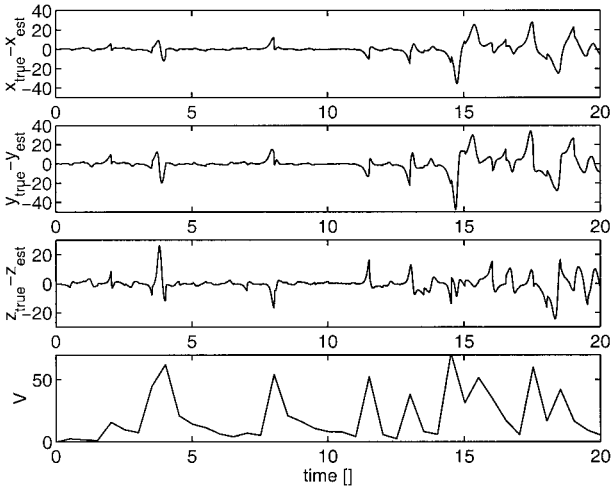

FIG. 7. Errors of the EKF estimate in *x*, *y*, *z* (line) and, in the lower panel, the nonlinearity measure *V*.

iations in the results the same random generator sequence was used for all the experiments.

c. Results

1) VARIATION OF NONLINEARITY IN TIME

The first experiment was designed to study the behavior of the nonlinearity in time. It was shown by Miller et al. (1994) that for some values of the parameters the extended Kalman filter will fail only at a few points in time because of the varying magnitude of the nonlinearities in the phase space. By correlating the measure V with the performance of the extended Kalman filter one can study the practical value of this number.

Figure 6 shows the simulated truth against a second-order RRSQRT filter run. The parameter values for this experiment were $\sigma_s = \sqrt{1/2}$, $\sigma_o = 1$, and $T = 1/2$. For these values the algorithm fails at several points in time. When comparing the errors of the estimate with the nonlinearity measure (see Fig. 7) it seems that failure of the algorithm is correlated with this measure; that is, there is a high probability of failure for $V \gg \sqrt{3}$ in this experiment. Note that since the system is stochastic the best one can do is compute the probability of a failure or detect a failure after it has occurred.

2) VARIATION OF THE NONLINEARITY WITH PARAMETER SETTINGS

In the previous experiment the parameter values (σ_s, σ_o, T) were fixed. It is expected however that variation of these parameters affects the effective nonlinearity of the data assimilation problem and therefore the performance of the data assimilation algorithm. Since in general no optimal algorithm is available the performance cannot be compared with the optimal performance. This problem can be overcome partially by studying the filter divergence PI, which is defined as the ratio between the true magnitude of the errors and their computed expected value, that is,

$$\mathrm{PI} = \sum_k \sqrt{\mathrm{trace}[\mathbf{P}_t(k|k)]} / \sqrt{\mathrm{trace}[\mathbf{P}_c(k|k)]}. \quad (76)$$

Figures 8 and 9 show the rms errors [rms $\equiv |\mathbf{e}(k|k)|$] and filter divergence for a large number of parameter settings. The folowing values were used: $\sigma_s^2 \in \{0.1, 0.2, 0.5, 1.0, 2.0\}$, $\sigma_o^2 \in \{0.1, 1.0, 2.0, 5.0, 10.0\}$, and $T \in \{0.1, 0.25, 0.5, 0.75, 1.0, 1.5\}$. One experiment was carried out for all possible combinations of these values, which resulted in 150 data assimilation runs. The simulation interval was $t \in [0, 100]$.

The nonlinearity V varies with all three parameters: σ_s, σ_o, and T. The value of V grows both with system noise and observation noise as is expected of the influence of nonlinearity on a data assimilation problem.

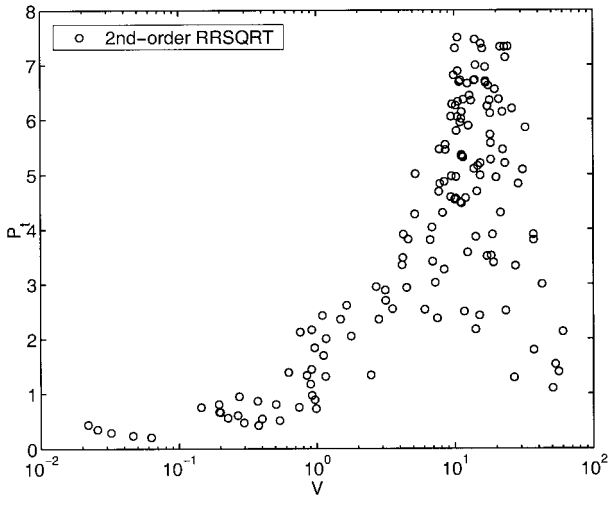

FIG. 8. Rms errors vs nonlinearity V for various parameter values.

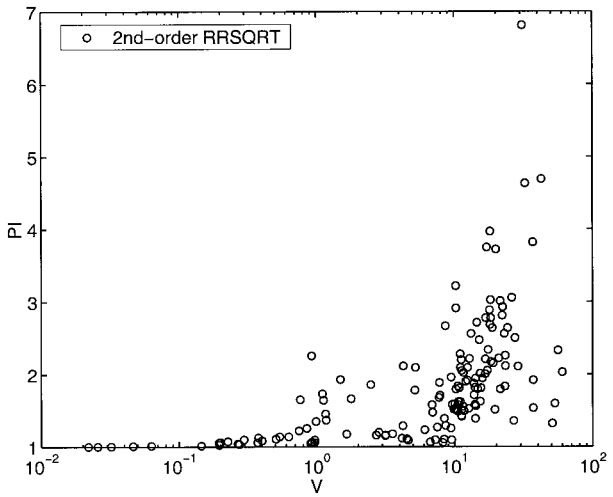

FIG. 9. Filter divergence PI vs nonlinearity V for various parameter values.

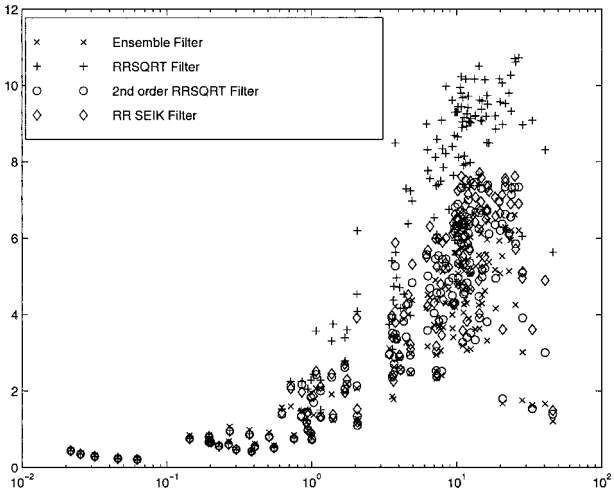

FIG. 11. Rms error and nonlineariy V for various algorithms.

Figure 10 shows that the nonlinearity also grows with the sampling interval T. These results are consistent with the experiments of Miller et al. (1994), where the observation intervals could be increased for more accurate measurements.

Assuming that stronger nonlinear data assimilation problems are more difficult to solve, one can expect that a second-order filter will fail for larger values of V. Figures 8 and 9 show that this indeed happens. The algorithm performs well for $V < ½$, for stronger nonlinear problems the algorithm may fail. It should be noted that failure of the algorithm depends on the trajectory and is thus subject to random effects. To eliminate random effects from V, rms, and PI their values were averaged over the full simulation interval, which has been taken quite long for this purpose (300 nondimensional time units). For smal intervals the variation between experiments started with different seed values for the random number generator may be significant.

3) A COMPARISON OF SEVERAL DATA ASSIMILATION ALGORITHMS

In the previous experiments it was shown that experiments with the Lorenz model for different values of the system noise, observation noise, and sampling interval can be structured when studying nonlinearity of the data assimilation problem and failure of a data assimilation algorithm. This structure is also helpful when comparing different data assimilation algorithms for nonlinear systems. In the next experiment several algorithms will be compared for the Lorenz model. Using

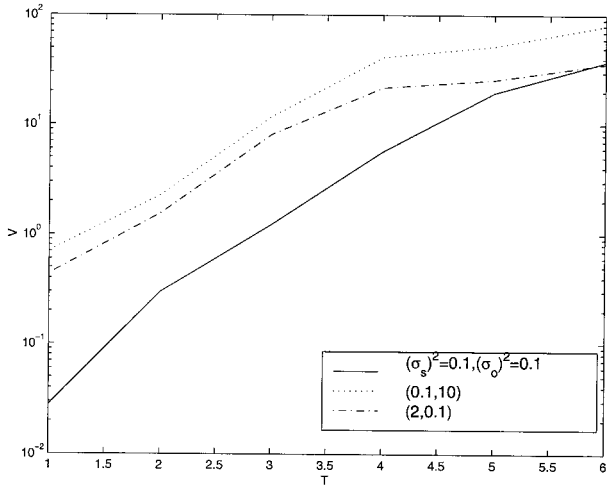

FIG. 10. Nonlinearity V vs sampling interval T.

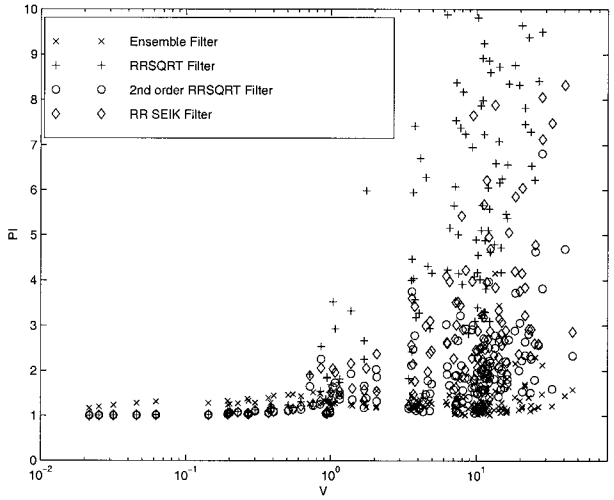

FIG. 12. Filter divergence (PI) and nonlinearity V for various algorithms.

TABLE 1. A comparison of several suboptimal Kalman filter algorithms.

Name of the algorithm	Order of accuracy	Index of computational requirements	Accuracy in experiments
RRSQRT (EKF)	1	4 (q + 1)	Least accurate
SEIK	2	5 (q + 2)	
RRSQRT (second order)	2	7 (2q + 1)	
EnKF	∞	30 (q)	Most accurate

the measure V the experiments can be organized against one axis instead of three.

Figures 11 and 12 show the rms error and filter divergence for the same set of parameters as in the previous experiments.

It can be seen that for nearly linear problems, which amounts here to $V \ll \sqrt{3}$, all the algorithms perform well. Thus it suffices to use the computationally least expensive method, which in this case is the EKF version of the RRSQRT filter. For highly nonlinear data assimilation problems (here $V \gg \sqrt{3}$) the EnKF is the most accurate. Thus if the computational requirements of this algorithm can be met, it is preferable to apply this or a similar algorithm. In between these two extremes there is a trade-off between accuracy and computational requirements. The computational requirements and accuracy of the algorithms studied here are shown in Table 1. The computational requirements are shown as the number of model integrations needed per forecast. For the reduced rank Kalman filters q denotes the rank, which is here the full state dimension $n = 3$, whereas for the EnKF q denotes the number of ensemble members. In the experiments a reasonable convergence of the EnKF occurred for 30 ensemble members. The relative computational requirements of the reduced rank Kalman filters and the EnKF depends on the application; thus, the numbers shown here may not be indicative for other applications.

However, due to the predictability time limit, caused by the inherent nonlinear instabilities of the system, all methods perform worse for larger intervals between succesive observations. In addition this is influenced by the magnitude of the errors in the model and in the observations. When a forecast is started with a more accurate initial condition and small system errors the tangent linear approximation will be more accurate. But even for a linear model the rms errors will be larger for larger model errors and larger observation errors. This implies that there exists a lower bound for the rms errors shown in Fig. 7; that is, with too few or too inaccurate observations there exists no algorithm that can compute an accurate estimate of the state of the system. For the performance index shown in Fig. 12 the situation is quite different. For an infinite ensemble size and infinite simulation time interval the EnKF should still compute the exact manitude of the (large) errors; that is, PI = 1.

Finally it should be noted that only Kalman filter algorithms were studied in this experiment. Kalman smoother algorithms, which also propagate information backward in time, generally give a better performance (see, e.g., Todling et al. 1998; Evensen 1997). However, a minimum data requirement also exists for Kalman smoother algorithms, albeit a less restrictive one.

6. Conclusions

In this work a measure for the nonlinearity of a data assimilation problem is proposed. The measure has several nice properties; that is, it is invariant under linear transformation of the state vector and nearly invariant under grid refinement. Moreover, it can be computed at low cost when a second-order Kalman filter is applied.

The measure was computed for a data assimilation problem with a simple but highly nonlinear model, known as the Lorenz model. The results indicate that the measure can detect failure of Kalman filtering type algorithms. It was shown that the nonlinearity of the data assimilation problem depends not only on the numerical model used, but also on on the accuracy of the measurements, the sampling frequency, and the variance of the system noise.

The experiments indicate that the method can be used for classifying the nonlinearity of data assimilation problems and for comparing data assimilation algorithms. The different Kalman filter algorithms showed almost the same performance for weakly nonlinear problems. For strongly nonlinear problems the more advanced methods, which are computationally more expensive, are also more accurate. The optimal choice seems to depend on the available computational resources and the nonlinearity of the problem. These preliminary results indicate that for weakly nonlinear problems the EKF version of the RRSQRT filter is to be preferred for its computational efficiency. For strongly nonlinear problems the ensemble Kalman filter is preferable for its higher accuracy.

Future research will focus on extending the comparison to include other data assimilation methods and on application of the method to other models.

REFERENCES

Budgell, W. P., 1986: Nonlinear data assimilation for shallow water equations in branched channels. *J. Geophys. Res.,* **91** (C9), 10 633–10 644.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: On the analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 1719–1724.

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

Dee, D. P., and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

Evensen, G., 1992: Using the extended kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.,* **97** (C11), 17 905–17 924.

——, 1994: Sequential data assimilation with a nonlinear quasi-geo-

strophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.
——, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.,* **125,** 1342–1354.
——, and P. J. van Leeuwen, 1995: Advanced data assimilation based on ensemble statistics. *Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Tokyo, Japan, WMO/TD-No. 651, 153–158.
——, and ——, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.,* **124,** 85–96.
——, and ——, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.,* **128,** 1852–1867.
Fukumori, I., and P. Melanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation; an example with an idealized Gulf Stream model. *J. Geophys. Res.,* **100,** 6777–6793.
Ghil, M., 1997: Advances in sequential estimation for atmospheric and oceanic flows. *J. Meteor. Soc. Japan,* **75,** 179–194.
Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Vol. 64, Mathematics in Science and Engineering, Academic Press, 376 pp.
Kalman, R. E., 1960: A new aproach to linear filter and prediction theory. *J. Basic Eng.,* **82D,** 35–45.
Kushner, H. J., 1967a: Approximations to optimal nonlinear filters. *IEEE Trans. Automatic Control,* **AC-12,** 546–556.
——, 1967b: Dynamical equations for optimum nonlinear filtering. *J. Differential Equations,* **3,** 179–190.
Lorenz, E., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.,* **20,** 130–141.
Miller, R. N., M. Ghil, and F. Ghautiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1055.
——, E. Carter, and S. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus,* **51A,** 167–194.
Pedlosky, J., 1987: *Geophysical Fluid Dynamics.* 2d ed. Springer-Verlag, 710 pp.
Pham, D. T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter with EOF initialization for data assimilation in oceanography. *J. Mar. Sci.,* **16,** 323–340.
Segers, A., A. Heemink, M. Verlaan, and M. van Loon, 1999: Nonlinear Kalman filters for atmospheric chemistry models. *Inverse Methods in Global Biogeochemical Cycles, Geophys. Monogr.,* Vol. 114, Amer. Geophys. Union, 139–146.
Sparrow, C., 1982: *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors.* Springer-Verlag, 269 pp.
Tennekes, H., and J. Lumley, 1972: *A First Course in Turbulence.* MIT Press, 300 pp.
Todling, R., S. E. Cohn, and N. Sivakumaran, 1998: Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Mon. Wea. Rev.,* **126,** 2274–2286.
van Leeuwen, P. J., and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.,* **124,** 2898–2913.
Verlaan, M., 1998: Efficient Kalman filtering algorithms for Hydrodynamic models. Ph.D. thesis, Delft University of Technology, Delft, Netherlands, 201 pp.
——, and A. W. Heemink, 1995: Reduced rank square root filters for large scale data assimilation problems. *Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Tokyo, Japan, WMO TD/-No.651, 247–252.