INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2010; **62**:374–402
Published online 26 February 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.2020

Comparison of sequential data assimilation methods for the Kuramoto–Sivashinsky equation

M. Jardak[1], I. M. Navon[1,*,†] and M. Zupanski[2]

[1]*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, U.S.A.*
[2]*Cooperative Institute for Research in the Atmosphere, Colorado State University, 1375 Campus Deliver, Fort Collins, CO 80523-1375, U.S.A.*

SUMMARY

The Kuramoto–Sivashinsky equation plays an important role as a low-dimensional prototype for complicated fluid dynamics systems having been studied due to its chaotic pattern forming behavior. Up to now, efforts to carry out data assimilation with this 1-D model were restricted to variational adjoint methods domain and only Chorin and Krause (*Proc. Natl. Acad. Sci.* 2004; **101**(42):15013–15017) tested it using a sequential Bayesian filter approach. In this work we compare three sequential data assimilation methods namely the Kalman filter approach, the sequential Monte Carlo particle filter approach and the maximum likelihood ensemble filter methods. This comparison is to the best of our knowledge novel. We compare in detail their relative performance for both linear and nonlinear observation operators. The results of these sequential data assimilation tests are discussed and conclusions are drawn as to the suitability of these data assimilation methods in the presence of linear and nonlinear observation operators. Copyright © 2009 John Wiley & Sons, Ltd.

Received 29 October 2008; Revised 19 January 2009; Accepted 20 January 2009

KEY WORDS: sequential data assimilation; ensemble Kalman filter; particle filter; Kuramoto–Sivashinsky equation

1. INTRODUCTION

Sequential data assimilation is a relatively novel and versatile multidisciplinary methodology. It combines observations of the current (and possibly, past) state of a system with results from a mathematical model (the forecast) to produce an analysis, providing 'the best' estimate of the

*Correspondence to: I. M. Navon, Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, U.S.A.
†E-mail: inavon@fsu.edu

Contract/grant sponsor: National Science Foundation; contract/grant number: ATM-03727818
Contract/grant sponsor: NASA Modeling, Analysis, and Prediction Program; contract/grant number: NNG06GC67G

current state of the system. Central to the concept of sequential estimation data assimilation is the propagation of flow-dependent covariance errors.

In sequential estimation, the analysis and forecasts can be viewed of as probability distributions. The analysis step is an application of the Bayes theorem. Advancing the probability distribution in time, in the general case is done by the Chapman–Kolmogorov equation, but since it is unrealistically expensive, various approximations operating on representations of the probability distributions are used instead. If the probability distributions are normal, they can be represented by their mean and covariance, which gives rise to the Kalman filter (KF). However, it is not computationally feasible to store the covariance due to the large number of degrees of freedom in the state, so various approximations based on Monte Carlo ensemble calculations are used instead.

In the present paper we compare standard ensemble Kalman filter (EnKF) along with the particle filter (PF) using various resampling strategies to the maximum likelihood ensemble filter (MLEF) of Zupanski [1]. In addition, we compare the performance of each of the above-mentioned methods in the presence of linear and nonlinear observation operators.

While work on 4-D variational data assimilation using as model, the Kuramoto–Sivashinsky (K–S) equation has been carried out by very few research workers (see Protas *et al.* [2]), to the best of our knowledge only the work of Chorin and Krause [3] has used Bayesian filters.

We further discuss issues of computational efficiency for the above-mentioned filters in terms of central processing unit (CPU) time and memory requirements for the three aforementioned methods taking into account the additional effort required by PF due to various resampling strategies aimed at preventing filter degeneracy.

The structure of the present work is as follows. After the introduction, we present in Section 2 the K–S equation and its numerical solution. In Section 3 we present the formulation of the PF and issues related to various resampling strategies implemented. The formulation of the standard EnKF is presented as well. Then we present the formulation of the MLEF method viewed as a hybrid between sequential and 4-D variational data assimilation. Section 4 provides discussion of the numerical experiments followed by a discussion of their significance. Finally, Section 5 contains summary and conclusions along with directions for further research.

2. THE K–S EQUATION

The K–S equation is an evolution equation in one space dimension, with a Burgers nonlinearity, a fourth-order dissipation term and a second anti-dissipative term. It assumes the form

$$u_t + u_{xxxx} + u_{xx} + uu_x = 0, \quad (x,t) \in \mathbb{R} \times \mathbb{R}^+ u(x,0) = u_0(x), \quad x \in \mathbb{R} \tag{1}$$

The K–S equation models pattern formations in different physical contexts and is a paradigm of low-dimensional behavior in solutions to partial differential equations. It arises as a model amplitude equation for inter-facial instabilities in many physical contexts. It was originally derived by Kuramoto and Tsuzuki [4, 5] to model small thermal diffusive instabilities in laminar flame fronts in two space dimensions. It has also been derived in the context of angular-phase turbulence for a system of reaction–diffusion modeling the Belouzov–Zabotinskii reaction in three space dimensions. Sivashinsky [6, 7] derived it independently to model small thermal diffusive instabilities in laminar flame fronts. The equation also arises in modeling small perturbations from a reference Poiseuille flow of a film layer on an inclined plane [8], while Babchin *et al.* [9] derived (1) as

DOI: 10.1002/fld

a general mechanism modeling the nonlinear saturation of instabilities in flow films as in the Rayleigh–Taylor-type instability.

The K–S equation is nonintegrable and no explicit solutions exist. It is characterized by a second-order unstable diffusion term, responsible for an instability at large scales, a fourth-order stabilizing viscosity term, which provides damping at small scales; and a quadratic nonlinear coupling term, which stabilizes by transferring energy between large and small scales. This is readily apparent in Fourier space, where one may write (1) with periodic boundary condition as

$$\frac{\mathrm{d}\hat{u}_k}{\mathrm{d}t} = (k^2 - k^4)\hat{u}_k + \frac{\mathrm{i}}{2}\sum_{k'\in\mathbb{Z}} k'\hat{u}_{k'}\hat{u}_{k-k'} \tag{2}$$

where

$$u(x,t) = \sum_{k\in\mathbb{Z}} \hat{u}_k(t)\exp(\mathrm{i}kx), \quad k = n\frac{2\pi}{L}, \quad k' = m\frac{2\pi}{L}, \quad m,n\in\mathbb{Z}, \ \mathrm{i} = \sqrt{-1}$$

The zero solution is linearly unstable to modes with $|k|<1$; whose number is proportional to the bifurcation parameter L, are coupled with each other and to damped modes at $|k|>1$ through the nonlinear term. In an effort to characterize, understand and predict the spatially and temporally nontrivial dynamical behavior including chaos of K–S, numerical simulation has been conducted by Hyman and Nicolaenko [10].

Foias *et al.* [11], Nicoleanko *et al.* [12] and Temam [13] established the existence of a unique compact inertial manifold for the K–S equation. They also demonstrated that the K–S equation is strictly equivalent to a low-dimensional dynamical system. That is, all orbits are attracted exponentially to a finite-dimensional, bounded, compact, smooth manifold and the dynamics take a place in this 'inertial' manifold. This implies that the transition to chaos of the K–S equation can be analyzed using the tools developed for low-dimensional dynamics systems.

Desertion and Kazantzis [14] used the K–S equation to examine the performance improvement of a class of nonlinear transport processes subject to spatio-temporally varying disturbances through the employment of a comprehensive and systematic actuator activation policy. Lee and Tran [15] obtained a reduced-order system that can accurately describe the dynamics of the K–S equation by employing an approximate inertial manifold and a proper orthogonal decomposition. From this resulting reduced-order system, they designed and synthesized the feedback controller for the K–S equation. Recently, Protas *et al.* [2] came up with a comprehensive framework for the regularization of adjoint analysis in multiscale partial differential equation (PDE) systems. They examined the regularization opportunities available in the adjoint analysis and optimization of multiscale, and applied the proposed regularization strategies to the K–S equation.

2.1. Mathematical formulation

We consider the solutions of

$$u_t + u_{xxxx} + u_{xx} + uu_x = 0, \quad (x,t)\in\mathbb{R}\times\mathbb{R}^+ u(x,0) = u_0(x), \quad x\in\mathbb{R} \tag{3}$$

which are space periodic of period L, $u(x,t) = u(x+L,t)$, $L>0$.

Let $\Omega\subset\mathbb{R}$, we denote by

$$\mathbb{L}^2(\Omega) = \left\{ u | u:\Omega \longrightarrow \mathbb{R}, u \text{ measurable and } \int_\Omega |u(x)|^2\,\mathrm{d}x < \infty \right\}$$

DOI: 10.1002/fld

the Hilbert space of square integrable function over Ω endowed with the norm

$$\|u\|_2 = \left\{ \int_\Omega |u(x)|^2 \mathrm{d}x \right\}^{1/2}$$

and by

$$\mathbb{H}^4(\Omega) = \left\{ u | u : \Omega \longrightarrow \mathbb{R}, u \text{ measurable and } \frac{\mathrm{d}^i u}{\mathrm{d}x^i} \in \mathbb{L}^2(\Omega) \text{ for } i = 0, \ldots, 4 \right\}$$

the fourth-order Sobolev space with the norm, see Adams [16] and also Mazya [17],

$$\|u\|_{4,2} = \left\{ \|u\|_2^2 + \sum_{i=1}^{4} \left\| \frac{\mathrm{d}^i u}{\mathrm{d}x^i} \right\|_2^2 \right\}^{1/2}$$

Finally, we denote by $\mathbb{H}^4_{\mathrm{per}}(\Omega)$ the closure of $\mathscr{C}^\infty_{\mathrm{per}}(\Omega)$ for the $\mathbb{H}^4$-norm. $\mathscr{C}^\infty_{\mathrm{per}}(\Omega)$ is the subset of $\mathscr{C}^\infty(\mathbb{R})$ of Ω-periodic functions. We set

$$\mathscr{A} = \frac{\mathrm{d}^4}{\mathrm{d}x^4}, \quad \mathscr{H} = \left\{ u \in \mathbb{L}^2 \left(\frac{-L}{2}, \frac{L}{2} \right) \right\}$$

and

$$D(\mathscr{A}) = \mathbb{H}^4_{\mathrm{per}} \left(\frac{-L}{2}, \frac{L}{2} \right) \cap \mathscr{H}, \quad \mathscr{V} = D(\mathscr{A}^{1/2})$$

$\mathscr{H}$ is a Hilbert space, the dissipative operator $\mathscr{A}$ is a linear self-adjoint unbounded operator in $\mathscr{H}$ with domain $D(\mathscr{A})$ and dense in $\mathscr{H}$. Assuming $\mathscr{A}$ positive closed and that $\mathscr{A}^{-1}$ is compact, $\mathscr{V}$ is a Hilbert space endowed with the norm $|\mathscr{A}^{1/2} \cdot|$. Using Leray's method (see Lions [18], Temam [19]), the K–S equation (1) with initial condition $u_0 \in \mathscr{H}$ has a unique solution defined for all $t > 0$ and such that

$$u \in \mathscr{C}(\mathbb{R}^+; \mathscr{H}) \cap \mathbb{L}^2(0, T; \mathscr{V}) \quad \forall T > 0$$

For a proof of the existence and uniqueness of the K–S equation, the keen reader may be referred to Temam [13], Marion and Temam [20]. Also the appendix of Collet *et al.* [21]. In addition, if $u_0 \in \mathscr{V}$, then

$$u \in \mathscr{C}(\mathbb{R}^+; \mathscr{V}) \cap \mathbb{L}^2(0, T; D(\mathscr{A})) \quad \forall T > 0$$

Furthermore, it has been proved in Nicoleanko *et al.* [12, 22] that only the odd solutions of (1) are stable for large t. Consequently, the subspace $\mathscr{H}$ is restricted to

$$\mathscr{H} = \left\{ u \in \mathbb{L}^2 \left(\frac{-L}{2}, \frac{L}{2} \right), \ u \text{ is odd} \right\}$$

3. DATA ASSIMILATION FOR THE K–S EQUATION

Data assimilation is the process by which observational data distributed in space and time are fused with mathematical model forecast information aimed at obtaining the best initial conditions that

DOI: 10.1002/fld

are as near as possible to observations while satisfying model forecast as a strong constraint. The probabilistic state space formulation and the requirement for the updating of information when new observations are encountered are ideally suited for the Bayesian approach, and thus constitute an appropriate framework for data assimilation. The Bayesian approach and in particular ensemble or particle filtering methods are a set of efficient and flexible Monte Carlo methods to solve the optimal filtering problem. Here, one attempts to construct the posterior probability density function (PDF) of the state based on the all available information, including the set of received observations. Since this PDF embodies all available statistical information, it may be considered to be a complete solution to the estimation problem.

In the field of data assimilation, there are only few contributions in sequential estimation (EnKF, PF and MLEF) using the K–S equation. Chorin and Krause [3] used PFs and proposed a strategy for reducing the size of the system of equations to be solved for evolving the particles by adaptively finding subsets of variables that do not need to be recomputed and solving only in directions in which the dynamic system is expanding. In the work of Hu and Temam [23], a robust boundary control method for the K–S equation has been proposed, and a data assimilation problem corresponding to the K–S equation has been considered.

3.1. Sequential Bayesian filter-theoretical setting

The sequential Bayesian filter employs a large number N of random samples or 'particles' advanced in time by a stochastic evolution equation, to approximate the probability densities. In order to analyze and make inference about the dynamic system, at least a model equation along with an observation operator is required. First, a model describing the evolution of the state with time, and an observation operator for noisy observations of the state. Generically, stochastic filtering problem is a dynamic system that assumes the form

$$\dot{\mathbf{x}}_t = f(t, \mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) \tag{4}$$

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{n}_t) \tag{5}$$

Equation (4) is the state equation or the system model, (5) is the observation operator equation, $\mathbf{x}_t$ is the state vector, $\mathbf{z}_t$ the observation vector and $\mathbf{u}_t$ is the system input vector serving as the driving force. $\mathbf{v}_t$ and $\mathbf{n}_t$ are the state and observation noises, respectively. In practical application, however, we are more concerned about the discrete-time filtering and, we consider the evolution of the state sequence $\{\mathbf{x}_k, k \in \mathbb{N}\}$, given by

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \tag{6}$$

where the deterministic mapping $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \longrightarrow \mathbb{R}^{n_x}$ is a possibly nonlinear function of the state $\mathbf{x}_{k-1}$, $\{\mathbf{v}_{k-1}, k \in \mathbb{N}\}$ is an independent identically distributed (i.i.d) process noise sequence, n_x, n_d are dimensions of the state and process noise vectors, respectively, and $\mathbb{N}$ is the set of the natural numbers. The objective is to recursively estimate $\mathbf{x}_k$ from observations

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{n}_k) \tag{7}$$

where $h_k : \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_n} \longrightarrow \mathbb{R}^{n_z}$ is a possibly nonlinear function, $\{\mathbf{n}_k, k \in \mathbb{N}\}$ is an i.i.d. observation noise sequence, and n_x, n_n are dimensions of the state and observation noise vectors, respectively.

We denote by $\mathbf{z}_{1:k}$ the set of all available observations $\mathbf{z}_i$ up to time $t = k$, $\mathbf{z}_{1:k} = \{\mathbf{z}_i | i = 1, \ldots, k\}$. From a Bayesian point of view, the problem is to recursively calculate some degree of belief in

DOI: 10.1002/fld

the state $\mathbf{x}_k$ at time $t=k$, taking different values, given the data $\mathbf{z}_{1:k}$ up to the time $t=k$. Then the Bayesian solution would be to calculate the PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. This density will encapsulate all the information about the state vector $\mathbf{x}_k$ that is contained in the observations $\mathbf{z}_{1:k}$ and the prior distribution for $\mathbf{x}_k$.

Suppose that the required PDF $p(\mathbf{x}|\mathbf{z}_{1:k-1})$ at time $k-1$ is available. The prediction stage uses the state Equation (6) to obtain the prior PDF of the state at time k via the Chapman–Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1})=\int p(\mathbf{x}_k|\mathbf{x}_{k-1})\,p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})\,\mathrm{d}\mathbf{x}_{k-1} \tag{8}$$

The probabilistic model of the state evolution, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, is defined by the state Equation (6) and the known statistics of $\mathbf{v}_{k-1}$.

At time $t=k$, a measurement $\mathbf{z}_k$ becomes available, and it may be used to update the prior via the Bayes rule

$$p(\mathbf{x}_k|\mathbf{z}_{1:k})=\frac{p(\mathbf{z}_k|\mathbf{x}_k)\,p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{9}$$

where the normalizing constant

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1})=\int p(\mathbf{z}_k|\mathbf{x}_k)\,p(\mathbf{x}_k|\mathbf{z}_{1:k-1})\,\mathrm{d}\mathbf{x}_k \tag{10}$$

depends on the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$, defined by the measurement Equation (7) and the known statistics of $\mathbf{n}_k$.

The relations (8) and (9) form the basis for the optimal Bayesian solution. This recursive propagation of the posterior density is only a conceptual solution. One cannot generally obtain an analytical solution. Solutions exist only in a very restrictive set of cases like that of the KFs for instance (namely, if f_k and h_k are linear and both v_k and n_k are Gaussian). PFs provide a direct approximation to the Bayes rule outlined above.

3.2. Particle filters

PFs (see [24–27]) approximate the posterior densities by population of states. These states are called 'particles'. Each of the particles has an assigned weight and the posterior distribution can then be approximated by a discrete distribution, which has support on each of the particles. The probability assigned to each particle is proportional to its weight. The different (PF) algorithms differ in the way that the population of particles evolves and assimilates the incoming observations. A major drawback of PFs is that they suffer from sample degeneracy after a few filtering steps. The PF suffers from 'the curse of dimensionality' requiring computations that increase exponentially with dimension as pointed out by Silverman [28]. This argument was enhanced and amplified by the recent work of Bengtsson *et al.* [29] and Bickel *et al.* [30] and finally explicitly quantified by Snyder *et al.* [31]. They indicated that unless the ensemble size is greater than $\exp(\tau^2/2)$, where τ^2 is the variance of the observation log-likelihood, the PF update suffers from a 'collapse' in which with high probability a single member is assigned a posterior weight close to one while all other members have vanishing small weights. This issue becomes more acute as we move to higher-spatial dimensions.

DOI: 10.1002/fld

Neverless in this 1-D problem, the common remedy is to resample the prior PDF whenever the weights focus on few members of the ensemble assuming that our experiments do not encounter the exponential collapse. Here, we use several strategies such as systematic resampling (SR), residual resampling (RR) and/or the Bayesian bootstrap filter of Gordon *et al.* [32], see also Berliner and Wikle [27, 33–37]. Multinomial resampling: The SIR algorithm generates a population of equally weighted particles to approximate the posterior at some time k. This population of particles is assumed to be an approximate sample from the true posterior at that time instant.

The PF algorithm proceeds as follows:

- *Initialization*: The filter is initialized by drawing a sample of size N from the prior at the initial time. The algorithm is then started with the filtering step.
- *Preliminaries*: Assume that $\{\mathbf{x}_{k-1}^i\}_{i=1,\ldots,N}$ is a population of N particles, approximately distributed as in an independent sample from $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$.
- *Prediction*: Sample N values, $\{q_k^1, \ldots, q_k^N\}$, from the distribution of $\mathbf{v}_k$. Use these to generate a new population of particles, $\{\mathbf{x}_{k|k-1}^1, \mathbf{x}_{k|k-1}^2, \ldots, \mathbf{x}_{k|k-1}^N\}$ via the equation

$$\mathbf{x}_{k|k-1}^i = f_k(\mathbf{x}_{k-1}^i, \mathbf{v}_k^i) \tag{11}$$

- *Filtering*: Assign each $\mathbf{x}_{k|k-1}^i$, a weight q_k^i. This weight is calculated by

$$q_k^i = \frac{p(\mathbf{z}_k|\mathbf{x}_{k|k-1}^i)}{\sum_{j=1}^N p(\mathbf{z}_k|\mathbf{x}_{k|k-1}^j)} \tag{12}$$

  This defines a discrete distribution which, for $i \in \{1, 2, \ldots, N\}$, assigns probability mass q_k^i to element $\mathbf{x}_{k|k-1}^i$.
- *Resampling*: Resample independently N times, with replacement from the distribution obtained in the filtering stage. The resulting particles, $\{\mathbf{x}_k^i\}_{i=1,\ldots,N}$, form an approximate sample from $p(\mathbf{x}_k|\mathbf{z}_{1:k})$.

The method outlined above can be justified as follows. If the particles at time $t = k-1$ were an i.i.d. sample from the posterior at time $t = k-1$, then the predictive stage just produces an i.i.d. sample from the prior at time $t = k$. The filtering stage can be viewed as an importance sampling approach to generate an empirical distribution, which approximates the posterior.

The proposal density is just the prior $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$, and as a result of Bayes formula, we obtain

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{z}_k) \propto p(\mathbf{x}_k|\mathbf{z}_{1:k-1})\, p(\mathbf{z}_k|\mathbf{x}_k) \tag{13}$$

the weights are proportional to the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$. As N tends to infinity, the discrete distribution, which has probability mass q_i at point $\mathbf{x}_{k|k-1}^i$, converges weakly to the true posterior. The resampling step is a crucial and computationally expensive part in a PF. It is used to generate equally weighted particles aimed at avoiding the problem of degeneracy of the algorithm, that is, avoiding the situation that all but one of the weights are close to zero. The resampling step modifies the weighted approximate density $p(\mathbf{x}_k|\mathbf{z}_k)$ to an unweighted density $\hat{p}(\mathbf{x}_k|\mathbf{z}_k)$ by eliminating particles having low importance weights and by multiplying particles having highly importance weights. Formally

$$p(\mathbf{x}_k|\mathbf{z}_k) = \sum_{i=1}^N q_i\, \delta(\mathbf{x}_k - \mathbf{x}_k{}^i) \tag{14}$$

DOI: 10.1002/fld

is replaced by

$$\hat{p}(\mathbf{x}_k|\mathbf{z}_k)=\sum_{i=1}^{N}\frac{1}{N}\delta(\mathbf{x}_k-\mathbf{x}_k{}^{\star})=\sum_{i=1}^{N}\frac{n_i}{N}\delta(\mathbf{x}_k-\mathbf{x}_k{}^{i}) \tag{15}$$

where n_i is the number of copies of particle $\mathbf{x}_k{}^{i}$ in the new set of particles $\{\mathbf{x}_k{}^{\star}\}$.

Generically, it is implemented as follows:

- Draw N particles $\{\tilde{\mathbf{x}}_k^i\}_{i=1,\ldots,N}$ from the uniform distribution.
- Assign the resampled particles $\{\tilde{\mathbf{x}}_k^i\}_{i=1,\ldots,N}$ to $\{\mathbf{x}_k^i\}_{i=1,\ldots,N}$ and assign equal weights $1/N$ to each particle.

3.3. Merging particle filter

To avoid the high cost inherent in recursive resampling (used in order to remove numbers with very small weights and replenish the ensemble), we test here a method put forward by Nakano *et al.* [38] called the merging particle filter (MPF), which is a modification to the PF. The MPF is similar to the genetic algorithm of Goldberg [39]. A filtered ensemble is constructed based on the samples from a forecast ensemble as in the regular PF with resampling. However, each particle of a filtered ensemble is generated as an amalgamation of multiple particles (i.e. ensembles of various sizes). The merging of several particles of a prior ensemble allows the degeneracy problem to be alleviated. For further details, the reader is referred to the paper of Nakano *et al.* [38].

3.4. The ensemble Kalman filter

The EnKF was first proposed by Evensen [40] and further developed by Burgers *et al.* [41] and Evensen [42, 43]. It is related to PFs in the context that a particle is identical to an ensemble member. EnKF is a sequential filter method, which means that the model is integrated forward in time and, whenever observations are available, these are used to reinitialize the model before the integration continues. The EnKF originated as a version of the extended Kalman filter [44, 45] for large problems. The classical KF [46] method is optimal in the sense of minimizing the variance only for linear systems and Gaussian statistics. Similar to the PF method, the EnKF stems from a Monte Carlo integration of the Fokker–Planck equation governing the evolution of the PDF that describes the prior, forecast and error statistics. In the analysis step, each ensemble member is updated according to the KF scheme and replaces the covariance matrix by the sample covariance computed from the ensemble. However, the EnKF presents two potential problems, namely:

(1) Even though the EnKF uses full nonlinear dynamics to propagate the forecast error statistics, the EnKF assumes that all probability distributions involved are Gaussian.
(2) The updated ensemble preserves only the first two moments of the posterior.

Let $p(\mathbf{x})$ be denote the Gaussian prior probability density distribution of the state vector $\mathbf{x}$ with mean μ and covariance $\mathscr{Q}$

$$p(\mathbf{x})\propto\exp(\tfrac{-1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\mathscr{Q}^{-1}(\mathbf{x}-\mu))$$

We assume the data $\mathbf{z}$ to have a Gaussian PDF with covariance $\mathscr{R}$ and mean $\mathscr{H}\mathbf{x}$, where $\mathscr{H}$ is the so-called the observation matrix, is related to h of Equation (5), and where the value $\mathscr{H}\mathbf{x}$

DOI: 10.1002/fld

assumes the value of the data $\mathbf{z}$ would be for the state $\mathbf{x}$ in absence of observation errors. Then, the conditional probability or likelihood $p(\mathbf{z}|\mathbf{x})$ assumes the form

$$p(\mathbf{z}|\mathbf{x}) \propto \exp(\tfrac{-1}{2}(\mathbf{z}-\mathscr{H}\mathbf{x})^{\mathrm{T}}\mathscr{R}^{-1}(\mathbf{z}-\mathscr{H}\mathbf{x}))$$

According to the Bayes theorem, the posterior probability density follows from the relation:

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})\, p(\mathbf{x}) \tag{16}$$

There are many variants of implementing the EnKF of various computational efficiency and in what follows we employ standard formulation of the EnKF for linear and nonlinear observation operators with covariance localization. See [40, 41, 47–49]. The implementation of the standard EnKF may be divided into three steps, as follows:

- *Setting and matching*
  - Define the ensemble

$$\mathscr{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \tag{17}$$

    be an $n_x \times N$ matrix, whose columns are a sample from the prior distribution. N being the number of the ensemble members.
  - Form the ensemble mean

$$\bar{\mathscr{X}} = \mathscr{X} \cdot \mathbf{1}_N \tag{18}$$

    where $\mathbf{1}_N \in \mathbb{R}^{N\times N}$ is the matrix where each element is equal to $1/N$.
  - Define the ensemble perturbation matrix $\mathscr{X}'$ and set the $\mathbb{R}^{n_x \times n_x}$ ensemble covariance matrix $\mathscr{C}$

$$\mathscr{X}' = \mathscr{X} - \frac{1}{N}\bar{\mathscr{X}} \tag{19}$$

$$\mathscr{C} = \frac{\mathscr{X}'\mathscr{X}'^{\mathrm{T}}}{N-1} \tag{20}$$

- *Sampling*
  - Generate

$$\mathscr{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \tag{21}$$

    be an $n_z \times N$ matrix, whose columns are a replicate of the measurement vector $\mathbf{z}$ plus a random vector from the normal distribution $\mathscr{N}(0, \mathscr{R})$.
  - Form the $\mathbb{R}^{n_z \times n_z}$ measurement error covariance

$$\mathscr{R} = \frac{\mathscr{Z}\mathscr{Z}^{\mathrm{T}}}{N-1} \tag{22}$$

- *Updating*: Obtain the posterior $\mathscr{X}^p$ by the linear combinations of members of the prior ensemble

$$\mathscr{X}^p = \mathscr{X} + \mathscr{C}\mathscr{H}^{\mathrm{T}}(\mathscr{H}\mathscr{C}\mathscr{H}^T + \mathscr{R})^{-1}(\mathscr{Z} - \mathscr{H}\mathscr{X}) \tag{23}$$

DOI: 10.1002/fld

The matrix

$$\mathscr{K} = \mathscr{C}\mathscr{H}^{\mathrm{T}}(\mathscr{H}\mathscr{C}\mathscr{H}^{\mathrm{T}} + \mathscr{R})^{-1} \tag{24}$$

is the Kalman gain matrix. Since $\mathscr{R}$ is always positive definite (i.e. covariance matrix), the inverse $(\mathscr{H}\mathscr{C}\mathscr{H}^{\mathrm{T}} + \mathscr{R})^{-1}$ exists. An easy computation shows that the mean and covariance of the posterior or updated ensemble are given by

$$\bar{\mathscr{X}}^p = \mathscr{X}^p + \mathscr{K}[\mathbf{z} - (\mathscr{H}\mathscr{X}^p] \tag{25}$$

and

$$\mathscr{C}^p = \mathscr{C} - \mathscr{K}[\mathscr{H}\mathscr{C}\mathscr{H}^{\mathrm{T}} + \mathscr{R}]\mathscr{K}^{\mathrm{T}} \tag{26}$$

In the case of nonlinear observation operators, a modification to the above algorithm is advised. As presented in Evensen [42], let $\hat{\mathbf{x}}$ be the augmented state vector made up of the state vector and the predicted observation vector (nonlinear in this case).

$$\hat{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathscr{H}(\mathbf{x}) \end{pmatrix} \tag{27}$$

Define the linear observation operator $\hat{\mathscr{H}}$ by

$$\hat{\mathscr{H}} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{y} \tag{28}$$

and carry out the steps of the EnKF formulation in augmented state space $\hat{\mathbf{x}}$ and $\hat{\mathscr{H}}$ instead of $\mathbf{x}$ and $\mathscr{H}$. Superficially, this technique appears to reduce the nonlinear problem to the previous linear observation operator case. However, while the augmented problem, involving linear observation problem, is a reasonable way of formulating the EnKF, it is not as well-founded as the linear case, which can be justified as an approximation to the exact and optimal KF.

To prevent the occurrence of filter divergence usually due to the background-error covariance estimates from the small number of ensemble members as pointed out in Houtekamer and Mitchell [50], the use of covariance localization was suggested. Mathematically, the covariance localization increases the effective rank of the background error covariances. See the work of Gaspari and Cohn [51] also Hamill and Snyder [52], Hamill [53] and Ehrendorfer [54]. The covariance localization consists of multiplying point by point the covariance estimate from the ensemble with a correlation function that is 1.0 at the observation location and zero beyond some prescribed distance. Mathematically, to apply covariance localization, the Kalman gain

$$\mathscr{K} = \mathscr{C}\mathscr{H}^{\mathrm{T}}(\mathscr{H}\mathscr{C}\mathscr{H}^{\mathrm{T}} + \mathscr{R})^{-1}$$

is replaced by a modified gain

$$\hat{\mathscr{K}} = [\rho \circ \mathscr{C}]\mathscr{H}^{\mathrm{T}}(\mathscr{H}[\rho \circ \mathscr{C}]\mathscr{H}^{\mathrm{T}} + \mathscr{R})^{-1} \tag{29}$$

where $\circ$ denotes the Schur product (the Schur product of matrices $\mathscr{A}$ and $\mathscr{B}$ is a matrix $\mathscr{D}$ of the same dimension, where $d_{ij} = a_{ij}b_{ij}$) of a matrix $\mathscr{S}$ with local support with the covariance model generated by the ensemble. Various correlation matrices have been used for horizontal localization

DOI: 10.1002/fld

Gaspari and Cohn [51] constructed a Gaussian-shaped function that is actually a fourth-order piece-wise polynomial. Houtekamer and Mitchell [50] and Evensen [42] used a cut-off radius so that observations are not assimilated beyond a certain distance from the grid points.

3.5. The maximum likelihood ensemble filter

The MLEF proposed by Zupanski [1], and Zupanski and Zupanski [55] is a hybrid filter combining the 4-D variational method with the EnKF. It maximizes the likelihood of posterior probability distribution, which justifies its name. The method comprises of three steps, a forecast step that is concerned with the evolution of the forecast error covariances, an analysis step based on solving a nonlinear cost function and updating step.

- *Forecasting*: It consists of evolving the square root analysis error covariance matrix through the ensembles. The starting point is from the evolution equation of the discrete KF described in Jazwinski [44]

$$P_f^k = \mathscr{M}_{k-1,k} P_a^k \mathscr{M}_{k-1,k}^{\mathrm{T}} + Q_{k-1} \tag{30}$$

where $P_f^{(k)}$ is the forecast error covariance matrix at time k, $\mathscr{M}_{k-1,k}$ the nonlinear model evolution operator from time $k-1$ to time k and Q_{k-1} is the model error matrix, which is assumed to be normally distributed. Since P_a^{k-1} is positive matrix for any k, Equation (30) could be factorized and written as

$$P_f^k = \overbrace{(\mathscr{M}_{k-1,k}(P_a^k)^{1/2})}^{(P_f^k)^{1/2}} (\mathscr{M}_{k-1,k}(P_a^k)^{1/2})^{\mathrm{T}} + Q_{k-1} \tag{31}$$

where $(P_a^k)^{1/2}$ is of the form

$$(P_a^k)^{1/2} = \begin{pmatrix} p_{(1,1)}^k & p_{(2,1)}^k & \cdots & p_{(N,1)}^k \\ p_{(1,2)}^k & p_{(2,2)}^k & \cdots & p_{(N,2)}^k \\ p_{(1,n)}^k & p_{(2,n)}^k & \cdots & p_{(N,n)}^k \end{pmatrix} \tag{32}$$

as usual N is the number of ensemble members and n the number of state variables. The lower case $p_{i,j}^k$ are obtained by calculating the square root of (P_a^k). Using Equation (32), the square root forecast error covariance matrix $(P_f^k)^{1/2}$ can then be expressed as

$$(P_f^k)^{1/2} = \begin{pmatrix} b_{(1,1)}^k & b_{(2,1)}^k & \cdots & b_{(N,1)}^k \\ b_{(1,2)}^k & b_{(2,2)}^k & \cdots & b_{(N,2)}^k \\ b_{(1,n)}^k & b_{(2,n)}^k & \cdots & b_{(N,n)}^k \end{pmatrix} \tag{33}$$

DOI: 10.1002/fld

where for each $1 \leqslant i \leqslant N$

$$\mathbf{b}_i^k = \begin{pmatrix} b_{(i,1)}^k \\ b_{(i,2)}^k \\ \vdots \\ b_{(i,n)}^k \end{pmatrix} = \mathscr{M}_{k-1,k} \begin{pmatrix} x_1^k + p_{(i,1)}^k \\ x_2^k + p_{(i,2)}^k \\ \vdots \\ x_n^k + p_{(i,n)}^k \end{pmatrix} - \mathscr{M}_{k-1,k} \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_n^k \end{pmatrix} \tag{34}$$

The vector $\mathbf{x}^k = (x_1^k x_2^k \ldots x_n^k)^{\mathrm{T}}$ is the analysis state from the previous assimilation cycle, which is found from the posterior analysis pdf as presented in Lorenc [56].

- *Analyzing*: The analysis step for the MLEF involves solving a nonlinear minimization problem. As in Lorenc [56], the associated cost function is defined in terms of the forecast error covariance matrix and is given as

$$\mathscr{J}(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}} (P_f^k)^{-1} (\mathbf{x} - \mathbf{x}_b) + \tfrac{1}{2}[\mathbf{y} - h(\mathbf{x})]^{\mathrm{T}} \mathscr{R}^{-1} [\mathbf{y} - h(\mathbf{x})] \tag{35}$$

where $\mathbf{y}$ is the vector of observations, h is the nonlinear observation operator, $\mathscr{R}$ is the observational covariance matrix and $\mathbf{x}_b$ is a background state given by

$$\mathbf{x}_b = \mathscr{M}_{k-1,k}(\mathbf{x}^k) + Q_{k-1} \tag{36}$$

Through a Hessian preconditioner, we introduce the change of variable

$$(\mathbf{x} - \mathbf{x}_b) = (P_f^k)^{1/2} (\mathscr{I} + \mathscr{O})^{-T/2} \boldsymbol{\xi} \tag{37}$$

where $\boldsymbol{\xi}$ is vector of control variables, $\mathscr{O}$ is referred to as the observation information matrix and $\mathscr{I}$ is the identity matrix. The matrix $\mathscr{O}$ is provided by

$$\mathscr{O} = (P_f^k)^{T/2} \tilde{\mathscr{H}}^{\mathrm{T}} \tilde{\mathscr{H}} \mathscr{R}^{-1} \tilde{\mathscr{H}} (P_f^k)^{T/2} = (\mathscr{R}^{-1/2} \tilde{\mathscr{H}} (P_f^k)^{1/2})^{\mathrm{T}} (\mathscr{R}^{-1/2} \tilde{\mathscr{H}} (P_f^k)^{1/2}) \tag{38}$$

here, $\tilde{\mathscr{H}}$ is the Jacobian matrix of the nonlinear observation operator h evaluated at the background state $\mathbf{x}_b$.

Let $\mathscr{Z}$ be the matrix defined by

$$\mathscr{Z} = \begin{pmatrix} z_{(1,1)} & z_{(2,1)} & \cdots & z_{(N,1)} \\ z_{(1,2)} & z_{(2,2)} & \cdots & z_{(N,2)} \\ z_{(1,n)} & z_{(2,n)} & \cdots & z_{(N,n)} \end{pmatrix}, \quad \mathbf{z}_i = \begin{pmatrix} z_{(i,1)} \\ z_{(i,2)} \\ \vdots \\ z_{(i,n)} \end{pmatrix} = \mathscr{R}^{-1/2} \tilde{\mathscr{H}} \mathbf{b}_i^k \tag{39}$$

using the following approximations:

$$\mathbf{z}_i \approx \mathscr{R}^{-1/2} [h(\mathbf{x} + \mathbf{b}_i^k) - h(\mathbf{x})] \tag{40}$$

and

$$\mathscr{O} \approx \mathscr{Z} \mathscr{Z}^{\mathrm{T}} \tag{41}$$

DOI: 10.1002/fld

one can use an eigenvalue decomposition of the symmetric positive definite matrix $\mathscr{I}+\mathscr{O}$ to calculate the inverse square root matrix necessary to the updating step. Its worth mentioning that the approximation (40) is not necessary and a derivation of the MLEF not involving (40) has been recently developed in Zupanski *et al.* [57].

- *Updating*: The final point about MLEF is to update the square root analysis error covariance matrix. In order to estimate the analysis error covariance at the optimal point, the optimal state $\mathbf{x}_{\text{opt}}$ minimizing the cost function $\mathscr{J}$ given by (35) is substituted

$$(P_a^k)^{T/2}=(P_f^k)^{T/2}(\mathscr{I}+\mathscr{O}(\mathbf{x}_{\text{opt}}))^{-T/2} \tag{42}$$

4. NUMERICAL EXPERIMENTS

The novelty of this work consists in comparing for the first time, the above described and commonly used filters in the framework of the K–S equation model focusing on the performance in the presence of nonlinear observation operators.

4.1. Numerical solution of the K–S equation

We consider the one-dimensional PDE with initial data as used in [58, 59]

$$\begin{aligned} &u_t=-u_{xxxx}-u_{xx}-uu_x, \quad x\in[0,32\pi]\\ &u(x+L,t)=u(x,t), \quad L=32\pi \quad \forall t>0\\ &u(x,0)=\cos\left(\frac{x}{16}\right)\left(1+\sin\left(\frac{x}{16}\right)\right) \end{aligned} \tag{43}$$

The system (43) is known to be stiff. In fact, the stiffness is due to rapid exponential decay of some modes (the dissipative part), the stiffness is also due to rapid oscillations of some modes (the dispersive part).

As the equation is periodic, a Fourier spectral method is used for spatial discretization. Despite the remarkable success of the spectral and pseudo-spectral methods for a wide range of applications [60, 61], the set of ordinary differential equations for the mode amplitudes is stiff, due to the time scale associated with the nth mode scales as $O(n^{-m})$ for large n, where m is the order of the highest spatial derivative, so that the highest modes evolve on short time scales.

In order to carry out numerical solution of K–S, a modification of the exponential time-differencing fourth-order Runge–Kutta method (ETDRK4) has been used. This method has been proposed by Cox and Matthews [62] and further modified by Kassam and Trefethen [58]. A short review of the ETDRK4 is as follows:

First, we transform (43) to Fourier space

$$\hat{u}_t=-\frac{\mathrm{i}k}{2}\hat{u}^2+(k^2-k^4)\hat{u} \quad \text{with } \hat{u}^2=\sum_{k'\in\mathbb{Z}}\hat{u}_{k'}\hat{u}_{k-k'} \tag{44}$$

set

$$\mathscr{L}\hat{u}(k)=(k^2-k^4)\hat{u}(k), \quad \mathscr{N}(\hat{u},t)=\mathscr{N}(\hat{u})=-\frac{\mathrm{i}k}{2}(\mathscr{F}(\mathscr{F}^{-1}(\hat{u}^2))) \tag{45}$$

DOI: 10.1002/fld

$\mathscr{L}$ and $\mathscr{N}$ stand for linear and nonlinear operators, respectively. $\mathscr{F}$ denotes the discrete Fourier transform. Write (44) in an operational form

$$\hat{u}_t = \mathscr{L}\hat{u} + \mathscr{N}(\hat{u}, t) \tag{46}$$

Define $v = \mathrm{e}^{-\mathscr{L}t}\hat{u}$, where $\mathrm{e}^{-\mathscr{L}t}$ the integrating factor to obtain

$$v_t = \mathrm{e}^{-\mathscr{L}t}\mathscr{N}(\mathrm{e}^{\mathscr{L}t}v) \tag{47}$$

Let h be denote the time step length, then integrating (47) over a single time step of length h, we obtain

$$\hat{u}_{n+1} = \mathrm{e}^{\mathscr{L}h}\hat{u}_n + \mathrm{e}^{\mathscr{L}h}\int_0^h \mathrm{e}^{-\mathscr{L}\tau}\mathscr{N}(\hat{u}(t_n+\tau), t_n+\tau)\,\mathrm{d}\tau \tag{48}$$

where $\hat{u}_n$ is the solution at the time $t = nh$ and $0 < \tau < h$.

Equation (48) is exact and the various order EDT schemes differ only on the way one approximates the integral in (48). Cox and Matthews [62] proposed the generating formula

$$\hat{u}_{n+1} = \mathrm{e}^{\mathscr{L}h}\hat{u}_n + h\sum_{m=0}^{s-1} g_m \sum_{k=0}^{m}(-1)^k \binom{m}{k}\mathscr{N}_{n-k} \tag{49}$$

where s is the order of the scheme. The coefficients g_m are provided by the recurrence relation

$$\begin{gathered}\mathscr{L}(hg_0) = \mathrm{e}^{\mathscr{L}h} - I \\ \mathscr{L}(hg_{m+1}) + \mathbf{I} = g_m + \frac{1}{2}g_{m-1} + \frac{1}{3}g_{m-2} + \cdots + \frac{1}{m+1}g_0, \quad m \geqslant 0\end{gathered} \tag{50}$$

We solve the K–S equation employing 64 Fourier spectral modes and integrate from $t = 0$ to $t = 250$ (nondimensional time units) using the EDTRK4 time stepping. The time evolution for the K–S equation is depicted in Figure 1, while the time evolution for the K–S at two different locations is presented in Figure 2.

4.2. Design of the numerical experiments

The following table summarizes the logic of the numerical experiments conducted in this paper:

Filter	Features	Observation operator type
PF	Resampling and merging methods	Linear and nonlinear
EnKF	Standard and with localization and inflation	Linear and nonlinear
MLEF	Standard	Linear and nonlinear

In order to compare the above-mentioned EnKF, PF and MLEF sequential DA methods, we use the K–S equation model with both linear and nonlinear observation operators for the same number of ensemble members. During the numerical experiments, the following set-up for each of the above-mentioned DA methods was used. The observations were provided at a frequency consisting of one observation every 30 time steps. Each run was carried out for 250 nondimensional time units with a time step of $\Delta t = \frac{1}{10}$. Thus 80 DA cycles were used for each experiment. The number

DOI: 10.1002/fld

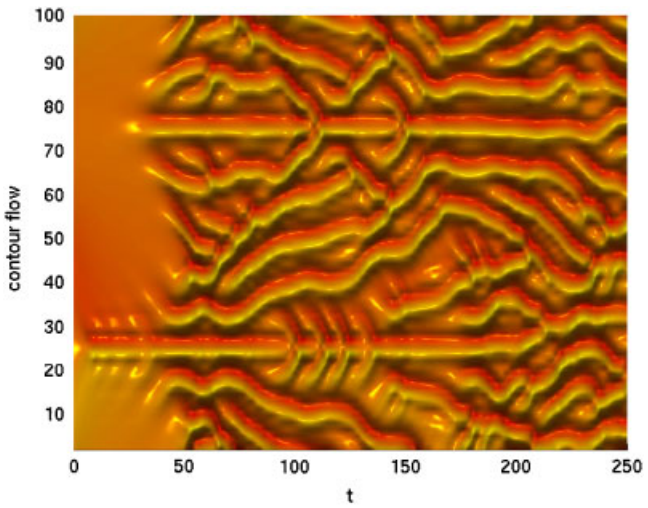

Figure 1. Time evolution for the K–S equation from $t=0$ to $t=250$ time units.

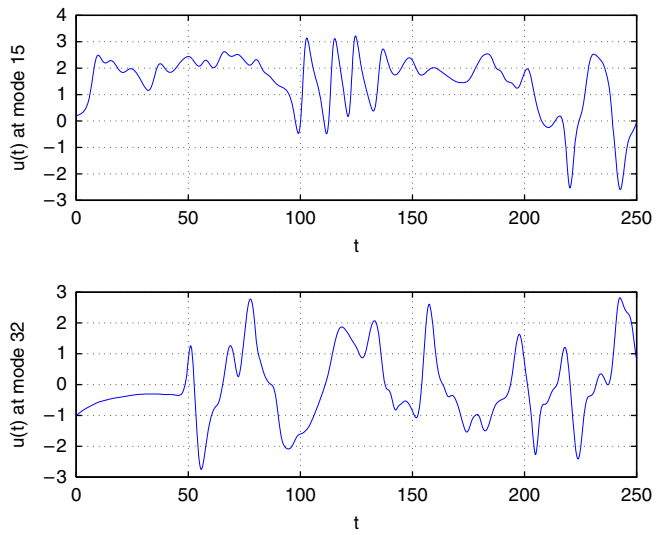

Figure 2. K–S solution at different spatial positions.

of Fourier modes in (2) was set to 64 modes. The standard deviation of each of the state and observation noise was taken to be $\sqrt{2}$.

For each type of the observation operator, the conclusion on the performance of each of the above-mentioned three methods from the numerical experiments will based on:

- amount of discrepancy between the time evolution of the forecast and the filter solutions;
- reduction of the analysis covariance error and

DOI: 10.1002/fld

- reduction of the root mean squared error (RMSE) and the χ^2 statistics between the forecast and data assimilation solutions.

4.3. Implementation details

- *Particle filter*: The observations were provided at a frequency consisting of one observation every 30 time steps. Each run was carried out for 250 nondimensional time units. Thus, 80 DA cycles were used for each experiment. The standard deviation of each of the Gaussian state noise and observation Gaussian noise was taken to be $\sqrt{2}$. Different numbers of particles were used to illustrate the effect of the number of particles on the respective analysis solutions. However, 300 particles were employed during the comparison of the three filters. Three different resampling methods were tested, all yielding similar results but only the SR method was retained since it exhibited the fastest performance.
- *Ensemble Kalman filter*: The observations were provided at a frequency consisting of one observation every 30 time steps. Each run was carried out for 250 nondimensional time units. Thus, 80 DA cycles were used for each experiment. The standard deviation of each of the Gaussian state noise and observation Gaussian noise was taken to be $\sqrt{2}$. Different numbers of ensemble members were tested for the EnKF method, but for actual experiments 200 ensemble members were found to be sufficient to conduct the numerical comparison. The features of localization and inflation were used. The additive covariance inflation of Anderson and Anderson [63] with $r = 1.001$ has been employed. The Gaussian correlation function

$$\rho(D) = \exp\left[-\left[\frac{D}{l}\right]^2,\ l = 450 \text{ length units}\right] \tag{51}$$

  was found to yield better results in the construction of the covariance localization. Finally, the singular value decomposition has been used whenever the calculation of the Kalman gain was needed.
- *The maximum likelihood ensemble filter*: Only 20 ensemble members were sufficient during the MLEF runs, while the number of cycles were set to 50. The iterative minimization employed the Fletcher–Reeves nonlinear conjugate gradient algorithm. In each of the MLEF DA cycles, three minimization iterations are performed to obtain the analysis. The standard deviation of each of the Gaussian state noise and observation Gaussian noise was taken to be $\sqrt{2}$.

4.4. Sequential data assimilation with a linear observation operator

Before comparing the performance of above-mentioned DA methods, we examine the performance of each method separately starting with the PF method in the presence of linear observation operator.

In Figure 3 we present the time evolution of the weight distribution of particles (ensemble members) using the SR method with 300 particles. The results obtained yielded a similar behavior to that observed when multinomial or RR methods were employed. We also provide in Figure 4 a detailed representation of the weight distribution for a fixed time, here $t = T_{\text{final}} = 250$. Figure 4 shows clearly that despite resampling most of the particles are degenerated and only few have significant weights. It illustrates the fact discussed by Snyder *et al.* [31] that an exponential growth

DOI: 10.1002/fld

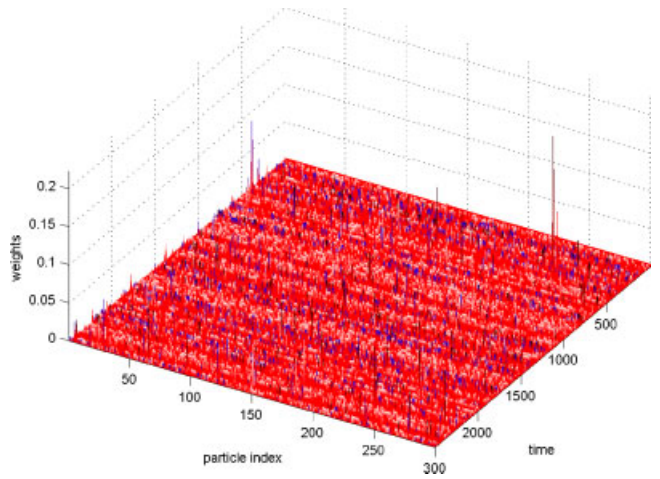

Figure 3. Time evolution of the weights distribution. 300 particles over 250 time units, using the systematic resampling method, linear observation operator case.

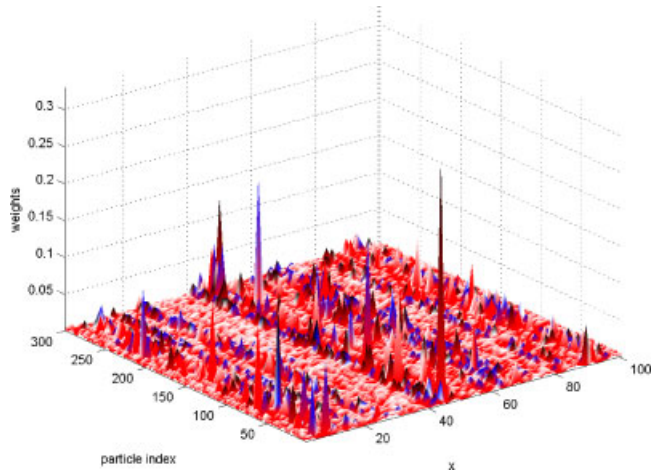

Figure 4. Weights distribution at a fixed time. 300 particles at the final time $t=250$, using the systematic resampling method, linear observation operator case.

(i.e. $\exp(\tau^2/2)$, where τ^2 is the variance of the observation log-likelihood) in the number of particles is required for the PF to perform satisfactorily.

Figure 5 illustrates two different (modes) locations for the above-mentioned PF experiment. The trends show an almost perfect matching between the forecast and the PF filter data assimilation results. This fact being further confirmed by the small value of the analysis error covariance depicted in Figure 6. Moreover, Figure 6 shows how the analysis error covariance decreases and becomes localized (concentration around the diagonal) with time. One thousand particles were employed and the plots were taken after $t=20$, $t=100$ and $t=250$ time units, respectively. We

DOI: 10.1002/fld

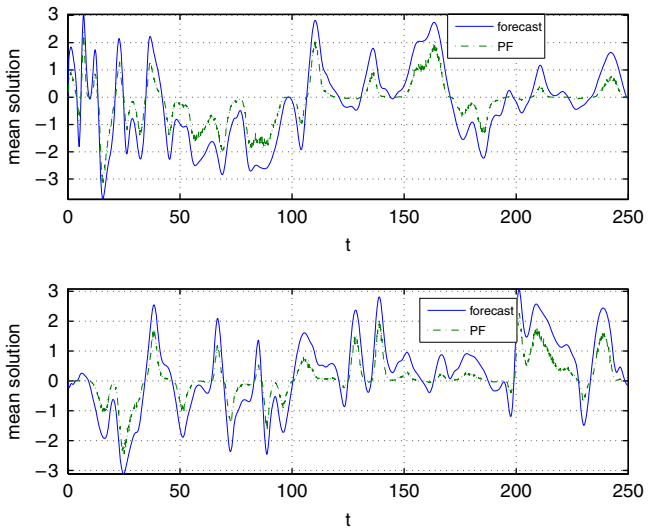

Figure 5. Forecast and PF filter mean K–S solution at two different locations, linear observation operator case.

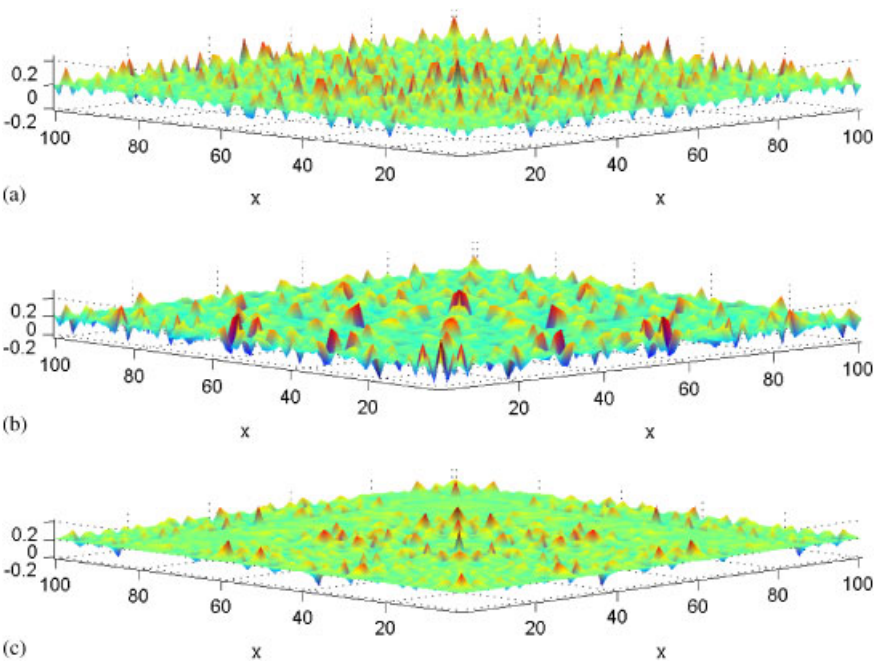

Figure 6. Analysis error covariance using 200 particles: (a) after $t = 20$ time units; (b) after $t = 100$ time units; and (c) after $t = 250$ time units, linear observation operator case.

DOI: 10.1002/fld

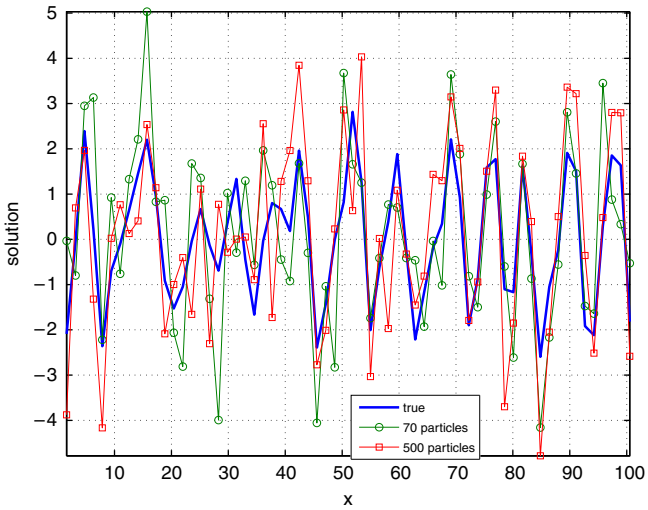

Figure 7. PF solutions for different number of particles (ensemble members) taken at $t = 250$ time units, linear observation operator case.

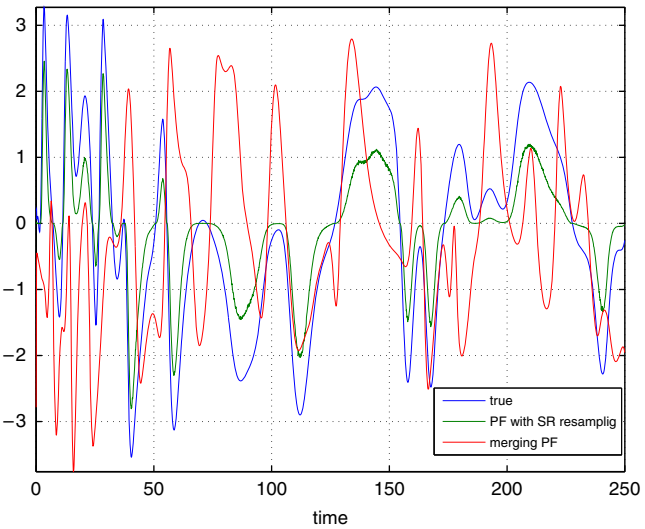

Figure 8. PF with the SR and MPF solutions for the K–S equation model, linear observation operator case.

present in Figure 7, the true and the PF solutions for different number of particles and for fixed time $t = 250$ time units. From Figure 7 one can deduce that as the number of particles increases, the PF solution tends to converge toward the true solution.

Even though the purpose of the paper is not to compare the efficiency of resampling methods, we have examined the residual, systematic and the multinomial resampling methods. They all

DOI: 10.1002/fld

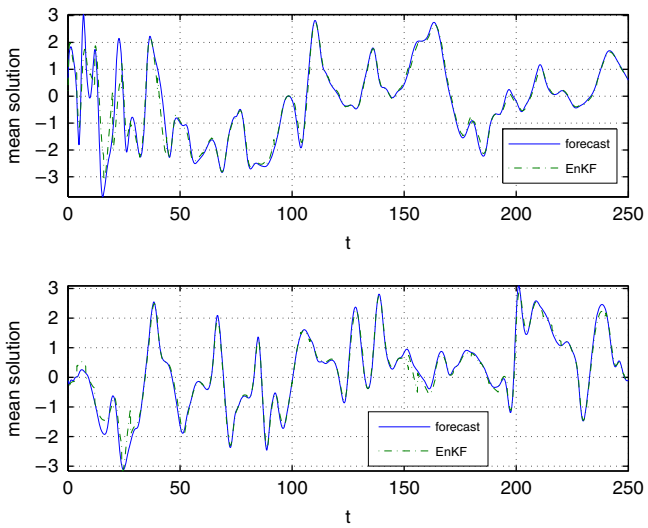

Figure 9. Forecast and EnKF filter mean K–S solution at two different locations, linear observation operator case.

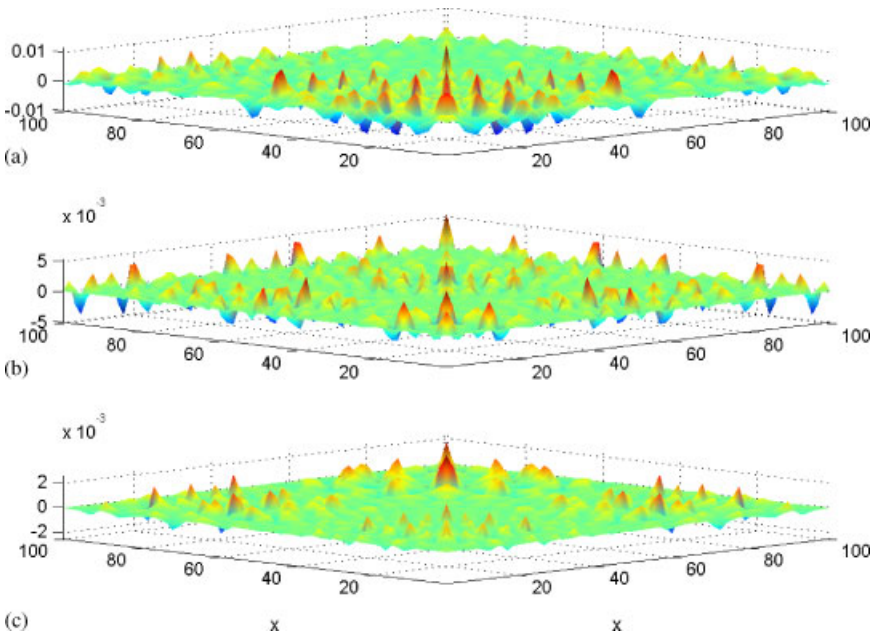

Figure 10. Analysis error covariance using 100 ensemble members: (a) after $t = 20$ time units; (b) after $t = 100$ time units; and (c) after $t = 250$ time units, linear observation operator case.

DOI: 10.1002/fld

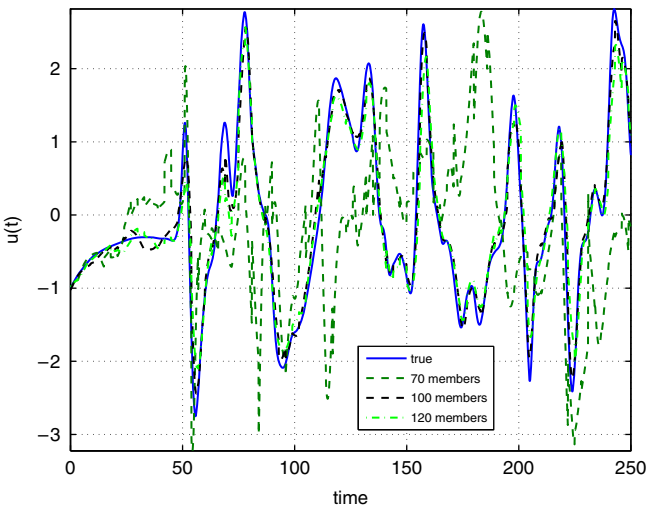

Figure 11. EnKF solutions for different number of ensemble members, linear observation operator case.

yield the similar results. In terms of CPU time, the SR method was the fastest followed by the RR method. This is in good agreement with the finding of Arulampalam *et al.* [26] and also Doucet *et al.* [25].

We also examined and compared the performance of the MPF of Nakano *et al.* [38] to the standard PF with resampling. Indeed, the MPF method was faster than the PF method, but for the K–S equation model, the PF provides more accurate results as it can be easily seen from the Figure 8.

We present in Figure 9, a perfect matching between the forecast and the EnKF solution at two different locations. The analysis error covariance for different times is depicted in Figure 10. Similar to the PF case, the analysis error covariance decreases and becomes local as time increases. The effect on the number of ensemble members is illustrated in Figure 11. Indeed, as the number of ensemble members increases, the EnKF data assimilation solution converges towards the true solution. It should be also pointed out that there exists a certain number of ensemble members, say N_{em}, such that for $\forall n > N_{\text{em}}$, the discrepancy between the EnkF solution and the true solution is virtually the same. This claim was observed to be valid throughout our numerical tests.

The effect of localization and inflation in EnKF for the case of the K–S equation model is presented in Figure 12; however, this effect has no sizable impact for our model in comparison to the numerical weather prediction models employed by Hamill and Snyder [52], Houtekamer and Mitchell [50] and Jardak *et al.* [64]. The additive covariance inflation of Anderson and Anderson [63] with $r = 1.001$ has been employed. The covariance localization was also used to prevent filter divergence. For the present model, we tested the fifth-order correlation function of Gaspari and Cohn [51] and the usual Gaussian correlation function

$$\rho(D) = \exp\left[-\left[\frac{D}{l}\right]^2,\ l = 450 \text{ length units}\right] \tag{52}$$

was found to yield better results.

DOI: 10.1002/fld

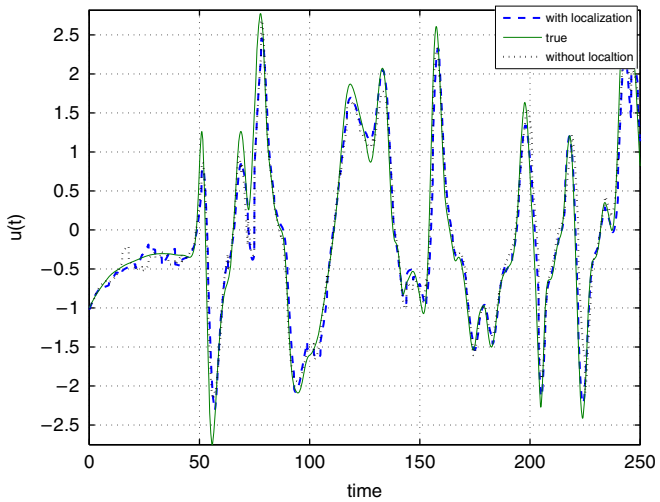

Figure 12. EnKF solutions with and without the inflation and localization covariance feature, linear observation operator case.

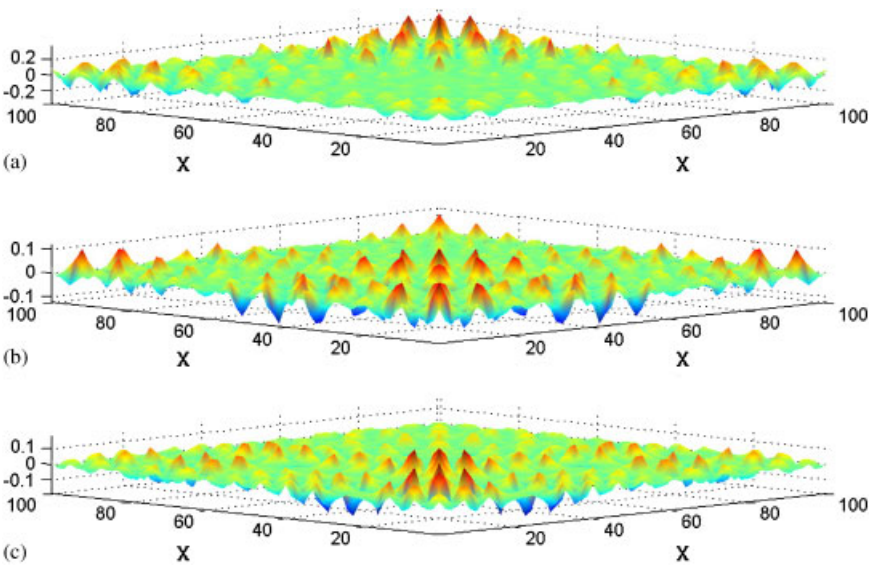

Figure 13. Analysis error covariance: (a) after $t = 20$ time units; (b) after $t = 100$ time units; and (c) after $t = 250$ time units, linear observation operator case, MLEF method.

Our numerical results using the MLEF are presented in Figure 13. Again, the analysis error covariance decreases and becomes localized (concentration around the diagonal) with time as shown in Figure 13.

DOI: 10.1002/fld

Table I. Root mean square error for different number of particle/ensemble members, linear observation operator case.

	PF	EnKF	MLEF
70	1.3558	2.0961	1.38189133
100	1.4313	0.3297	1.43104518
150	1.3635	1.4603	1.7722971
250	1.3903	1.3388	1.5577693

Table II. χ^2 statistics for different number of particle/ensemble members, linear observation operator case.

	PF	EnKF	MLEF
70	2.5346690	2.345658	1.9612679
100	1.9854678	1.136783	2.0121631
150	3.0457823	1.783459	2.6414165
250	1.934876	1.655631	1.8208084

Comparing linear observation operator results of covariances, RMSE and χ^2 for the three above-mentioned sequential data assimilation filters leads to the conclusion that for the model at hand, the EnKF provides better results followed by the MLEF and then the PF filter with SR. These results are of course a function of our particular implementation of the EnKF, PF and MLEF and are function of the number of particles (number of ensemble members) and the number of observations as well as the particular model tested. In as far as computational efficiency, the EnKF was found to be faster than the MLEF, while the PF filter proved to be the most time consuming in as far as the K–S equation model with linear observation operator is concerned. The performance of the above-mentioned methods is assessed and encapsulated in Tables I and II.

4.5. Sequential data assimilation with nonlinear observation operator

A similar setting to the linear observation case in terms of the number of ensemble member and particles has been used to carry out numerical experiments with nonlinear observation operator. The EnKF version tailored to nonlinear observation operator as described in Section 3 has been implemented and an observation operator assuming the form $\mathscr{H}(u)=u^2$ has been considered. As depicted in Figure 14, an intense discrepancy between the forecast and EnKF mean solutions is easily seen. We join many authors, Nakano *et al.* [38] among them to draw the conclusion that even with the modification suggested by Evensen [42], the EnKF does not provide good estimates in the case of nonlinear observation operator applied to the K–S equation model. It came recently to our attention the work of Kalnay [65] dealing with the issue of nonlinear observation operators in the case of EnKF. We intend to implement and test Kalnay's approach in our forthcoming paper [64].

Having the previous conclusion in mind, the comparison reduces to comparing the PF and the MLEF methods in the presence of nonlinear observation operator. Both the PF and the MLEF exhibited good performances in terms of sequential data assimilation as measured in different error metrics such as error covariance matrix, RMSE and χ^2 statistics. The PF method with SR appears

DOI: 10.1002/fld

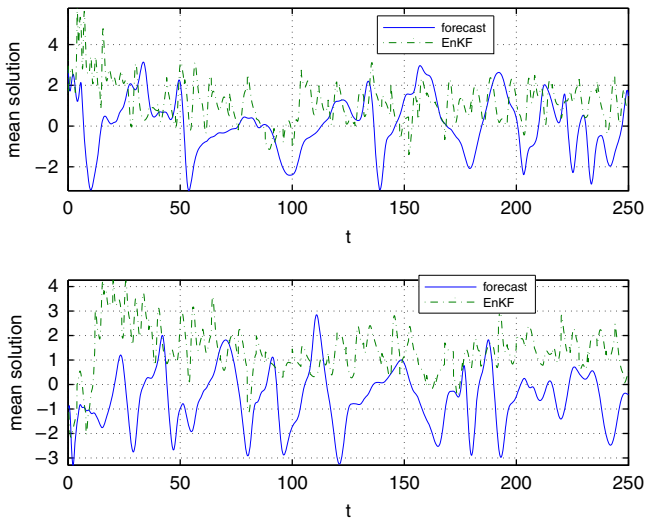

Figure 14. Means of the forecast and the EnKF filtered K–S solution in the case of a nonlinear observation operator $\mathscr{H}(u)=u^2$, nonlinear observation operator case.

Table III. Root mean square error for different number of particle/ensemble members, nonlinear observation operator case.

	PF for $\mathscr{H}(u)=u^2$	MLEF for $\mathscr{H}(u)=u^2$	PF for $\mathscr{H}(u)=\exp(u)$	MLEF for $\mathscr{H}(u)=\exp(u)$
70	0.35639178	1.3781648	0.27696167	1.43203425
100	0.3244502203	1.4070521	0.30416423	1.40807023
250	0.366861214	1.43104518	0.31519531	1.5476146

Table IV. χ^2 statistics for different number of particle/ensemble members, nonlinear observation operator case.

	PF for $\mathscr{H}(u)=u^2$	MLEF for $\mathscr{H}(u)=u^2$	PF for $\mathscr{H}(u)=\exp(u)$	MLEF for $\mathscr{H}(u)=\exp(u)$
70	2.6344987	3.7372720	4.3238997	10.1312923
100	2.8783455	4.2008870	3.7356629	9.37517203
250	1.9873456	2.0121631	2.1099875	9.52631038

to have an edge in terms of CPU time, while in other metrics, the PF has a slight edge over the MLEF method as seen in Tables III and IV.

This result is further confirmed, as it can also be assessed from Tables III and IV, for the case when the nonlinear observation operator assumes the form $\mathscr{H}(u)=\exp(u)$.

Our numerical experimentations in the case of nonlinear observation operators are depicted in Figures 15 and 16, where the forecast mean is contrasted against the PF mean for both the square and the exponential observation operators. In Figures 17 and 18, we present the analysis error covariance at different times also for both the square and the exponential observation operators.

DOI: 10.1002/fld

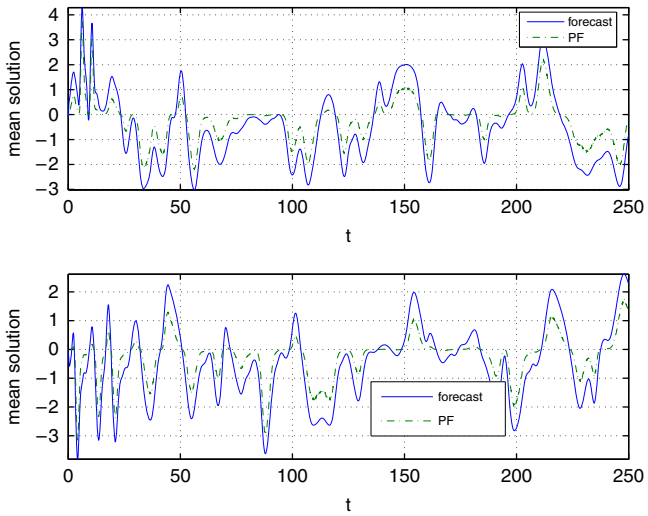

Figure 15. Means of the K–S solution using SIR particle filter, nonlinear observation operator case $\mathscr{H}(u)=u^2$.

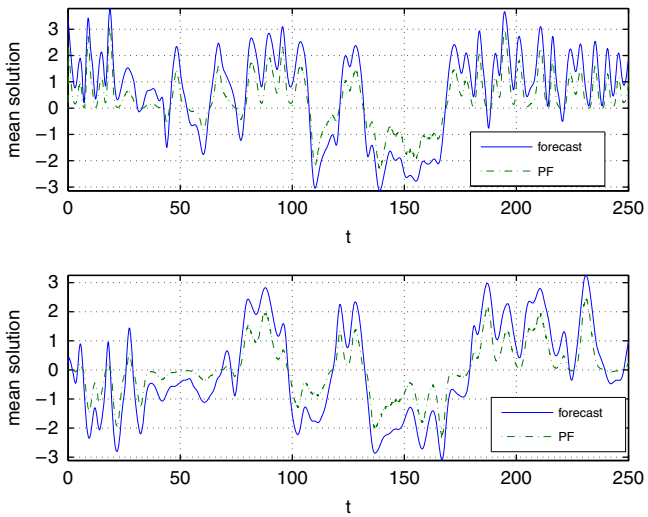

Figure 16. Means of the K–S solution using particle filter, nonlinear observation operator case, $\mathscr{H}(u)=\exp(u)$.

It appears that the high CPU time required by the MLEF is due to the minimization component of this method. This remark remains to be elucidated in further research. The results of the performance of the PF and MLEF methods with $\mathscr{H}(u)=\exp(u)$ as nonlinear observation operator are presented in Tables III and IV.

DOI: 10.1002/fld

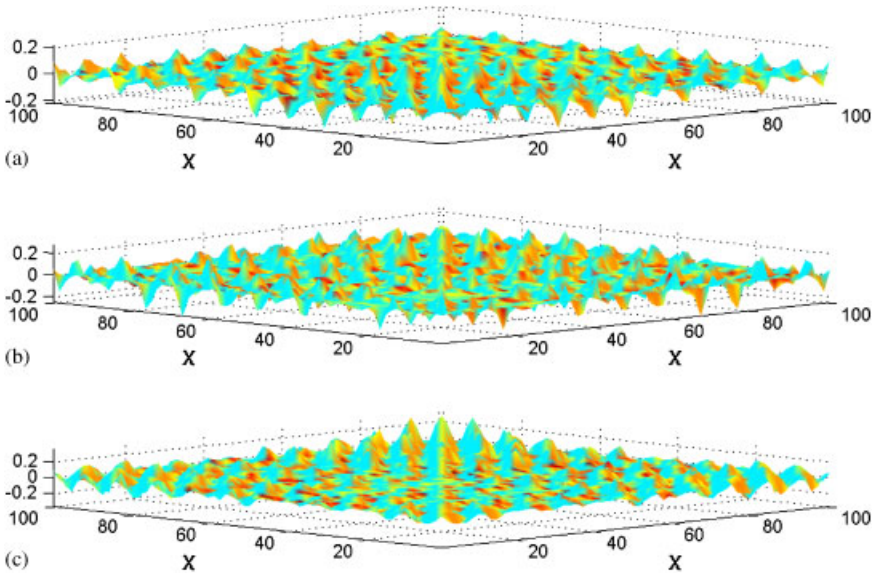

Figure 17. Analysis error covariance: (a) after $t=20$ time units; (b) after $t=100$ time units; and (c) after $t=250$ time units, nonlinear observation operator case $\mathscr{H}(u)=u^2$, MLEF method.

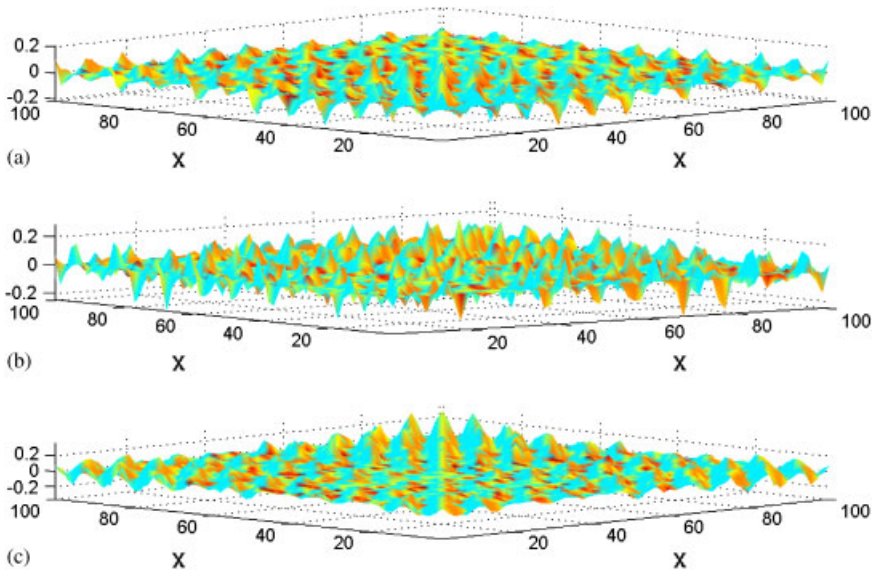

Figure 18. Analysis error covariance: (a) after $t=20$ time units; (b) after $t=100$ time units; and (c) after $t=250$ time units, nonlinear observation operator case $\mathscr{H}(u)=\exp(u)$, MLEF method.

DOI: 10.1002/fld

5. SUMMARY AND CONCLUSIONS

In this work we have presented a comparison of three sequential data assimilation methods, PF, EnKF and MLEF applied to the K–S equation model. For each of the aforementioned methods, both linear and nonlinear observation operators have been employed and extensive numerical experiments have been carried out to assess their relative performances.

It appears that the standard EnKF implementation with localization and inflation works well for the case of linear observation operator. However, it has a major drawback dealing with nonlinear observation operators. The PF and the MLEF perform satisfactory for both the types of observation operators. Our experience suggests that the PF with resampling overall performed somewhat better than the MLEF for the particular set-up tested. Additional future research will focus on alternative PF resampling such as the one proposed in Xiong *et al.* [66] and Salman [67]. Also, new options of using hybrid PF EnKF filter as proposed by Brockmeier [68] will be considered to alleviate the degeneracy catastrophe of PF.

ACKNOWLEDGEMENTS

The research of Prof. I. M. Navon and Dr M. Jardak is supported by the National Science Foundation (NSF), grant ATM-03727818. The authors also acknowledge the support by NASA Modeling, Analysis, and Prediction Program under Award NNG06GC67G. The first author is thankful to W. H. Hanya for the continuous help.

REFERENCES

1. Zupanski M. Maximum likelihood ensemble filter. Part1: the theoretical aspect. *Monthly Weather Review* 2005; **133**:1710–1726.
2. Protas B, Bewley TR, Hagen G. A computational framework for the regularization of adjoint analysis in multiscale PDE systems. *Journal of Computational Physics* 2004; **195**(1):49–89.
3. Chorin AJ, Krause P. Dimensional reduction for a Bayesian filter. *Proceedings of the National Academy of Sciences* 2004; **101**(42):15013–15017.
4. Kuramoto Y, Tsuzuki T. On the formation of dissipative structures in reaction–diffusion systems. *Progress of Theoretical Physics* 1975; **54**:687–699.
5. Kuramoto Y, Tsuzuki T. Persistent propagation of concentration waves in dissipative media far from thermal equilibrium. *Progress of Theoretical Physics* 1976; **55**:356–369.
6. Sivashinsky GI. Nonlinear analysis of hydrodynamic instability in laminar flames I. Derivation of basic equations. *Acta Astronautica* 1977; **4**:1177–1206.
7. Sivashinsky GI. On irregular wavy flow of a liquid film down a vertical plane. *Progress of Theoretical Physics* 1980; **63**:2112–2114.
8. Benny DJ. Long waves in liquid film. *Journal of Mathematics and Physics* 1966; **45**:150–155.
9. Babchin AJ, Frenkel AL, Levich BG, Sivashinsky GI. Nonlinear saturation of Rayleigh–Taylor instability in thin films. *Physics of Fluids* 1983; **26**:3159–3161.
10. Hyman JM, Nicolaenko B. The Kuramoto–Sivashinsky equation: a bridge between PDE'S and dynamical systems. *Physica D* 1986; **18**:113–126.
11. Foias C, Nicolaenko B, Sell GR, Temam R. Inertial manifolds for the Kuramoto–Sivashinski equation and an estimate of their lowest dimension. *Journal de Mathmatiques Pures et Appliques* 1988; **67**:197–226.
12. Nicoleanko B, Scheurer B, Temam R. Some global dynamical properties of the Kuramoto–Sivashinsky equations: nonlinear stability and attractors. *Physica D* 1985; **16**:155–185.
13. Temam R. *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, *Applied Mathematical Sciences*, vol. 68. Springer: Berlin, 1997.
14. Desertion KA, Kazantzis N. Improved performance of the controlled Kuramoto–Sivashinsky equation via actuator and controller switching. *Proceedings of the 2004 American Control Conference*, Boston, MA, 30 June–2 July 2004; 267–272.

DOI: 10.1002/fld

15. Lee CH, Tran HT. Reduced-order-based feedback control of the Kuramoto–Sivashinsky equation. *Journal of Computational and Applied Mathematics* 2005; **173**:1–19.
16. Adams RA. *Sobolev Spaces*. Academic Press: New York, 1975.
17. Mazya VG. *Sobolev Spaces*. Springer: Berlin, 1985.
18. Lions JL. *Quelques méthodes de résolution de problèmes aux limites non linéaires*. Dunod: Paris, 1969.
19. Temam R. *Navier–Stokes Equations* (3rd edn). North-Holland: Amsterdam, New York, 1984.
20. Marion M, Temam R. Nonlinear Galerkin methods. *SIAM Journal on Numerical Analysis* 1989; **26**(5):1139–1157.
21. Collet P, Eckmann JP, Epstein H, Stubbe J. A global attracting set for the Kuramoto–Sivashinsky equation. *Communications in Mathematical Physics* 1993; **152**(1):203–214.
22. Nicoleanko B, Scheurer B, Temam R. Some global dynamical properties of a class of pattern formation equations. *Communications in Partial Differential Equations* 1989; **14**:245–297.
23. Hu C, Temam R. Robust boundary control for the Kuramoto–Sivashinsky equation. *Dynamics of Continuous Discrete and Impulsive Systems—Series B—Applications and Algorithms* 2001; **8**(3):315–338.
24. Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 2000; **10**(30):197–208.
25. Doucet A, de Freitas JFG, Gordon NJ. An introduction to sequential Monte Carlo methods, In *Sequential Monte Carlo Methods in Practice*, Doucet A, de Freitas JFG, Gordon NJ (eds). Springer: New York, 2001.
26. Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 2002; **150**(2):174–188.
27. Berliner ML, Wikle CK. A Bayesian tutorial for data assimilation. *Physica D* 2007; **30**:1–16.
28. Silverman BW. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall: London, 1986.
29. Bengtsson T, Bickel P, Li B. Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. *Probability and Statistics*: *Essays in Honor of David A. Freedman* 2008; **2**:316–334.
30. Bickel P, Li B, Bengtsson T. Sharp failure rates for the bootstrap particle filter in high dimensions. *IMS Collections Pushing the Limits of Contemporary Statistics*: *Contributions in Honor of Jayanta K. Ghosh* 2008; **3**:318–329.
31. Snyder C, Bengtsson T, Bickel P, Anderson J. Obstacles to high-dimensional particle filtering. *Monthly Weather Review* 2008; **136**(12):4629–4640.
32. Gordon NJ, Salmond DJ, Smith AFM. Novel approach to nonlinear non-Gaussian Bayesian state estimate. *IEEE Proceedings F* 1993; **140**:107–113.
33. Berliner ML, Wikle CK. Approximate importance sampling Monte Carlo for data assimilation. *Physica D* 2007; **30**:37–49.
34. Hol JD, Schön TB, Gustafsson F. On resampling algorithms for particle filters. *Nonlinear Statistical Signal Processing Workshop*, Cambridge, U.K., 2006.
35. Liu JS, Chen R. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association* 1998; **93**(442):1032–1044.
36. Kotecha G, Djurić PM. Gaussian particle filtering. *IEEE Transactions on Signal Processing* 2003; **51**:2592–2601.
37. Kitagawa G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* 1996; **5**(1):1–25.
38. Nakano S, Ueno G, Higuchi T. Merging particle filter for sequential data assimilation. *Nonlinear Processes in Geophysics* 2007; **14**:395–409.
39. Goldberg DE. *Genetic Algorithms in Search*, *Optimization and Machine Learning*. Addison-Wesley: Reading, MA, 1989.
40. Evensen G. Sequential data assimilation in non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 1994; **99**(C510):108–129.
41. Burgers G, van Leeuwen PJ, Evensen G. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* 1998; **126**:1719–1724.
42. Evensen G. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics* 2003; **53**:343–367.
43. Evensen G. *Data Assimilation*: *The Ensemble Kalman Filter*. Springer: Berlin, 2007.
44. Jazwinski AH. *Stochastic Processes and Filtering Theory*. Academic Press: New York, 1970.
45. Bucy RS. Recurrent sets. *Annals of Mathematical Statistics* 1965; **36**:535–545.
46. Kalman RE. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering* 1960; **82**(Series D):35–45.
47. Mandel J. Efficient implementation of the ensemble Kalman filter. *CCM Report*, vol. 231, 2006.
48. Mandel J. A brief tutorial on the ensemble Kalman filter manuscript, 2007.

DOI: 10.1002/fld

49. Lewis JL, Lakshivarahan S, Dhall SK. *Dynamic Data Assimilation*: *A Least Squares Approach*. Series on Encyclopedia of Mathematics and its Applications, vol. 104. Cambridge University Press: Cambridge, 2006.
50. Houtekamer P, Mitchell M. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review* 1998; **126**:796–811.
51. Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society* 1999; **125**:723–757.
52. Hamill TM, Snyder C. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review* 2000; **128**:2905–2919.
53. Hamill TM. Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*. Cambridge Press: Cambridge, 2006; 124–156.
54. Ehrendorfer E. A review of issues in ensemble-based Kalman filtering. *Meteorologische Zeitschrift* 2007; **16**: 795–818.
55. Zupanski D, Zupanski M. Model error estimation employing ensemble data assimilation approach. *Monthly Weather Review* 2006; **134**:1337–1354.
56. Lorenc AC. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society* 1986; **112**:1177–1194.
57. Zupanski M, Navon IM, Zupanski D. The maximum likelihood ensemble filter as a non-differentiable minumization algorithm. *Quarterly Journal of the Royal Meteorology* 2008; **134**:1039–1050.
58. Kassam AK, Trefethen LN. Fourth-order time-stepping for stiff PDS's. *SIAM Journal on Scientific Computing* 2005; **26**(4):1214–1233.
59. Krogstad S. Generalized integrating factor methods for stiff PDE's. *Journal of Computational Physics* 2005; **203**:72–88.
60. Canuto C, Hussaini MY, Quarteroni A, Zang TA. *Spectral Methods in Fluids Dynamics*. Springer: Berlin, 1988.
61. Trefethen NL. *Spectral Methods in Matlab*. Software, Environments, Tools. SIAM: Philadelphia, 2000.
62. Cox SM, Matthews PC. Exponential time differencing for stiff systems. *Journal of Computational Physics* 2002; **176**:430–455.
63. Anderson JL, Anderson SL. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review* 1999; **127**:2741–2785.
64. Jardak M, Navon IM, Zupanski D. Ensemble data assimilation for the shallow water equation. *Journal of Geophysical Research* 2008; submitted.
65. Kalnay E. Some Ideas for Ensemble Kalman Filtering. *US National Oceanic and Atmospheric Administration*, *Climate Test Bed Joint Seminar Series*, IGES/COLA, 2008.
66. Xiong X, Navon IM, Uzunoglu B. A note on the particle filter with posterior Gaussian resampling. *Tellus* 2006; **58A**:456–460.
67. Salman H. A hybrid grid/particle filter for Lagrangian data assimilation: formulating the passive scalar approximation. *Quarterly Journal of the Royal Meteorological Society* 2008; **134**:1539–1550.
68. Brockmeier M. Using the particle filter for reservoir model updating. *Master of Science in Applied Mathematics*, Delft University of Technology, 2007.

DOI: 10.1002/fld