Tellus (2005), 57A, 715–735
Printed in Singapore. All rights reserved

Copyright © Blackwell Munksgaard, 2005
TELLUS

A comparison of error subspace Kalman filters

By LARS NERGER*, WOLFGANG HILLER and JENS SCHRÖTER, *Alfred Wegener Institute for Polar and Marine Research, PO Box 12 0161, 27515 Bremerhaven, Germany*

(Manuscript received 6 July 2004; in final form 9 February 2005)

ABSTRACT

Three advanced filter algorithms based on the Kalman filter are reviewed and presented in a unified notation. They are the well-known ensemble Kalman filter (EnKF), the singular evolutive extended Kalman (SEEK) filter, and the less common singular evolutive interpolated Kalman (SEIK) filter. For comparison, the mathematical formulations of the filters are reviewed in relation to the extended Kalman filter as error subspace Kalman filters. The algorithms are presented in their original form and possible variations are discussed. A comparison of the algorithms shows their theoretical capabilities for efficient data assimilation with large-scale non-linear systems. In particular, problems of the analysis equations are apparent in the original EnKF algorithm due to the Monte Carlo sampling of ensembles. Theoretically, the SEIK filter appears to be a numerically very efficient algorithm with high potential for use with non-linear models. The superiority of the SEIK filter is demonstrated on the basis of identical twin experiments using a shallow-water model with non-linear evolution. Identical initial conditions for all three filters allow for a consistent comparison of the data assimilation results. These show how choices of particular state ensembles and assimilation schemes lead to significant variations of the filter performance. This is related to different qualities of the predicted error subspaces, as is demonstrated in an examination of the predicted state covariance matrices.

1. Introduction

In recent years there has been an extensive development of data assimilation algorithms based on the Kalman filter (KF; Kalman and Bucy, 1961) in the atmospheric and oceanic contexts. These filter algorithms are of special interest due to their simplicity of implementation (e.g. no adjoint operators are required) and their potential for efficient use on parallel computers with large-scale geophysical models (see, for example, Keppenne and Rienecker, 2002).

The classical KF and the extended Kalman filter (EKF; see Jazwinski, 1970) share the problems that, for large-scale models, the requirements of storage and computation time are prohibitive due to the explicit treatment of the state error covariance matrix. Furthermore, the EKF shows deficiencies with the non-linearities appearing, e.g. in oceanographic systems (Evensen, 1992). To handle these problems, there have been different working directions over the last years. One approach is based on a low-rank approximation of the state covariance matrix of the EKF to reduce the computational costs. Using finite difference approximations for the tangent linear model, these algorithms also show better abilities to handle non-linearity as compared to the EKF. Examples of low-rank filters are the reduced rank square-root (RRSQRT) algorithm (Verlaan and Heemink, 1995) and the similar singular evolutive extended Kalman (SEEK) filter (Pham et al., 1998a). An alternative direction is the use of an ensemble of model states to represent the error statistics given in the EKF by the state estimate and covariance matrix. The most widely used algorithm of this kind is the ensemble Kalman filter (EnKF; Evensen, 1994; Burgers et al., 1998) which applies a Monte Carlo method to forecast the error statistics. Several variants of the EnKF have been proposed (Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002) which can be interpreted as ensemble square-root Kalman filters (Tippett et al., 2003). For an improved treatment of non-linear error evolution, the singular evolutive interpolated Kalman (SEIK) filter (Pham et al., 1998b) was introduced as a variant of the SEEK filter. It combines the low-rank approximation with an ensemble representation of the covariance matrix. This idea has also been followed in the concept of error subspace statistical estimation (ESSE; Lermusiaux and Robinson, 1999).

Because all these recent filter developments approximate the covariance matrix by a matrix of low rank, their analysis step, the part in which observations are assimilated, operates in a low-dimensional subspace of the true error space. Despite different forecast schemes, the analysis scheme of all filters is provided by some variant of the analysis equations of the EKF. Hence, we refer to these filters as error subspace Kalman filter (ESKF) algorithms.

The major part of the computation time in data assimilation with filter algorithms is spent in the prediction of error statistics.

*Corresponding author.
e-mail: lnerger@awi-bremerhaven.de

Thus, it is of particular interest to find filter algorithms which provide a sufficiently good filter performance in terms of the state and error estimates for minimal computation time, i.e. with as few model integrations as possible. Because the prediction of the error statistics within these filters is performed by some evolution of an ensemble of model states, the required ensemble size should be as small as possible. For the EnKF as a Monte Carlo method, it has been found that ensembles of the order of 100 members are required (Evensen and van Leeuwen, 1996; Natvik and Evensen, 2003). There have been attempts to allow for smaller ensembles by applying a smoothing operator to the sampled covariance matrix (Houtekamer and Mitchell, 2001) but these likely introduce spurious modes and imbalance (Mitchell et al., 2002). Compared to the EnKF, much smaller ranks of the approximated state covariance matrix have been reported for the SEEK filter (such as a rank of 7 by Brusdal et al., 2003) as have been for the SEIK filter (e.g. a rank of 30 by Hoteit et al., 2002). These numbers are, however, hardly comparable as they all refer to different models and physical situations. For comparability, the algorithms would have to be applied to the same model configuration using the same initial state estimate and covariance matrix. In a study which applied the EnKF and RRSQRT filter to a two-dimensional advection diffusion equation (Heemink et al., 2001) the RRSQRT filter yielded comparable estimation errors to the EnKF for about half the number of model evaluations. A comparison between the SEEK algorithms and the EnKF with an OGCM (Brusdal et al., 2003) also used fewer model evaluations for the SEEK filter than for the EnKF to obtain qualitatively comparable results. However, this result is difficult to interpret because both filters were applied to slightly different model configurations and used different initial conditions for the filters. Brusdal et al. (2003) have also pointed out the strong similarity of the EnKF and SEEK algorithms. However, the algorithm denoted therein as the SEEK filter deviates from the way it was originally introduced. It corresponds to the SEEK filter using a finite difference approximation for the forecast and is not formulated with the focus on the analysed quantities used in the original SEEK filter.

The discussion about error subspace Kalman filtering is complicated by the application of different filters to different problems. Furthermore, different stabilization techniques, e.g. covariance filtering (Hamill and Whitaker, 2001) or covariance inflation, are commonly introduced. While these techniques stabilize the filter performance they make a rigorous comparison and understanding difficult. Here we compare for the first time three algorithms in their original form in the internationally accepted mathematical notation (Ide et al., 1997). For the comparison we chose the SEEK filter, representing the class of low-rank filters, and the EnKF, which is widely used and represents the pure form of an ensemble filter method. Also under consideration is the SEIK filter, which combines the strengths of both methods. Other algorithms such as the RRSQRT (Verlaan and Heemink, 1995), ESSE (Lermusiaux and Robinson, 1999) or the ensemble square-root filters (see Tippett et al., 2003), can be easily related to these algorithms. It is not our intention to discuss the various stabilization techniques which may improve filter performance in special cases but amount to the individual tuning of each algorithm. Here we wish to focus on the similarities and differences in the filter strategies.

To assess the behaviour of different filter algorithms when applied to a non-linear test model in an oceanographic context, identical twin experiments are performed. The experiments utilize shallow-water equations with strongly non-linear evolution. Synthetic observations of the sea surface height are assimilated. Using identical conditions for the algorithms permits a direct and consistent comparison of the filter performances for various ensemble sizes. The experiments are evaluated by studying the root mean square (rms) estimation error for a variety of different ensemble sizes. In addition, an examination of the quality of the sampled state covariance matrices shows how the different representations of the covariance matrix and the different analysis schemes of the filter algorithms yield varying filter performances.

2. Filter mathematics

A good approach to the filter algorithms is given by the probabilistic view similar to Cohn (1997). Here we focus on non-linear large-scale systems. For ease of comparison, the notations follow the unified notation proposed by Ide et al. (1997). First, statistical estimation is briefly presented and the EKF, which is the common basis of the following algorithms, is reviewed. Subsequently, the ESKFs are discussed.

2.1. Statistical estimation

The data assimilation problem amounts to finding an optimal estimate of the system state for a certain time interval, given a dynamical model and observations at some discrete times. We focus on filtering, i.e. the system state at time t_k is estimated on the basis of observations available up to this time.

We consider a physical system which is represented in discretized form by its true state $\mathbf{x}^{\mathrm{t}}$ of dimension n. Because the model only approximates the true physics of the system, $\mathbf{x}^{\mathrm{t}}$ is a random vector whose time propagation is given by the stochastic-dynamic time discretized model equation

$$\mathbf{x}_i^{\mathrm{t}} = M_{i,i-1}\left[\mathbf{x}_{i-1}^{\mathrm{t}}\right] + \boldsymbol{\eta}_i. \tag{1}$$

Here $M_{i,i-1}$ is a, possibly non-linear, operator describing the state propagation between the two consecutive time-steps $i-1$ and i. The vector $\boldsymbol{\eta}_i$ is the model error, which is assumed to be a stochastic perturbation with zero mean and covariance matrix $\mathbf{Q}_i$.

At discrete times $\{t_k\}$, typically each Δk time-step, observations are available as a vector $\mathbf{y}_k^{\mathrm{o}}$ of dimension m_k. The true state $\mathbf{x}_k^{\mathrm{t}}$ at time t_k is assumed to be related to the observation vector

by the forward measurement operator H_k as

$$\mathbf{y}_k^{\mathrm{o}} = H_k\left[\mathbf{x}_k^{\mathrm{t}}\right] + \boldsymbol{\epsilon}_k. \tag{2}$$

Here $H_k[\mathbf{x}_k^{\mathrm{t}}]$ describes what observations would be measured given the state $\mathbf{x}_k^{\mathrm{t}}$. The vector $\boldsymbol{\epsilon}_k$ is the observation error. It consists of the measurement error due to imperfect measurements and the representation error caused, e.g. by the discretization of the dynamics. $\boldsymbol{\epsilon}_k$ is a random vector which is assumed to be of zero mean and covariance matrix $\mathbf{R}_k$ and uncorrelated with the model error $\boldsymbol{\eta}_k$.

The state sequence $\{\mathbf{x}_i^{\mathrm{t}}\}$, prescribed by eq. (1), is a stochastic process which is fully described by its probability density function (PDF) $p(\mathbf{x}_i^{\mathrm{t}})$. Accordingly, the filtering problem is solved by the conditional PDF $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_k^{\mathrm{o}})$ of the true state given the observations $\mathbf{Y}_k^{\mathrm{o}} = \{\mathbf{y}_0^{\mathrm{o}}, \ldots, \mathbf{y}_k^{\mathrm{o}}\}$ up to time t_k. In practice, it is not feasible to compute this density explicitly for large-scale models. Therefore, one typically relies on the calculation of some statistical moments of the PDF such as the mean and the covariance matrix. In the context of KFs, usually the conditional mean $\langle \mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_k^{\mathrm{o}} \rangle$ is computed, the expectation value of $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_k^{\mathrm{o}})$, which is also the minimum variance estimate (see Jazwinski, 1970).

In the following we concentrate on sequential filter algorithms. That is, the algorithms consist of two steps. In the 'forecast step' the PDF $p(\mathbf{x}_{k-\Delta k}^{\mathrm{t}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}})$ is evolved up to the time t_k when observations are available, yielding $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}})$. Then, in the 'analysis step', the PDF $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_k^{\mathrm{o}})$ is computed from the forecast density and the observation vector $\mathbf{y}_k^{\mathrm{o}}$. Subsequently, the cycle of forecasts and analyses is repeated. To initialize the filter sequence, an initial PDF $p(\mathbf{x}_0^{\mathrm{t}} \mid \mathbf{Y}_0^{\mathrm{o}})$ is required. This PDF is in practice unknown and an estimate $p(\mathbf{x}_0)$ is used for the initialization.

2.2. Extended Kalman filter

For linear dynamic and measurement models, the KF is the minimum variance and maximum likelihood estimator if the initial PDF $p(\mathbf{x}_0^{\mathrm{t}})$ and the model error and observation error processes are Gaussian. The EKF (see Jazwinski, 1970) is a first-order extension of the KF to non-linear models as given by eqs. (1) and (2). It is obtained by linearizing the dynamic and measurement operators around the most recent state estimate. To clarify the statistical assumptions underlying the EKF, we review it in the context of statistical estimation. In addition, we discuss the approximations which are required for the derivation of the EKF. A detailed derivation of the KF in the context of statistical estimation is presented by Cohn (1997) and several approaches toward the EKF are discussed in (Jazwinski, 1970, chap. 7).

In the dynamic model (1) and the observation model (2) we assume that the stochastic processes $\boldsymbol{\eta}_k$ and $\boldsymbol{\epsilon}_k$ are temporal white Gaussian processes with zero mean and covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively. Further, we assume $p(\mathbf{x}_k^{\mathrm{t}})$ to be Gaussian with covariance matrix $\mathbf{P}_k$ and all three processes to be mutually uncorrelated. Denoting the expectation operator by $\langle\,\rangle$, the assumptions are summarized as

$$\boldsymbol{\eta}_i \propto \mathcal{N}(\mathbf{0}, \mathbf{Q}_i); \quad \left\langle \boldsymbol{\eta}_i \boldsymbol{\eta}_j^{\mathrm{T}} \right\rangle = \mathbf{Q}_i \delta_{ij} \tag{3}$$

$$\boldsymbol{\epsilon}_k \propto \mathcal{N}(\mathbf{0}, \mathbf{R}_k); \quad \left\langle \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_l^{\mathrm{T}} \right\rangle = \mathbf{R}_k \delta_{kl} \tag{4}$$

$$\mathbf{x}_i^{\mathrm{t}} \propto \mathcal{N}\left(\bar{\mathbf{x}}_i^{\mathrm{t}}, \mathbf{P}_i\right); \tag{5}$$

$$\left\langle \boldsymbol{\eta}_k \boldsymbol{\epsilon}_k^{\mathrm{T}} \right\rangle = 0; \quad \left\langle \boldsymbol{\eta}_i \left(\mathbf{x}_i^{\mathrm{t}}\right)^{\mathrm{T}} \right\rangle = 0; \quad \left\langle \boldsymbol{\epsilon}_k \left(\mathbf{x}_k^{\mathrm{t}}\right)^{\mathrm{T}} \right\rangle = 0. \tag{6}$$

Here $\mathcal{N}(\mathbf{a}, \mathbf{B})$ denotes the normal distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{B}$. It is $\delta_{kl} = 1$ for $k = l$ and $\delta_{kl} = 0$ for $k \neq l$. Under assumptions (3)–(5) the corresponding PDFs are fully described by their two lowest statistical moments: the mean and the covariance matrix. Applying this property, the EKF formulates the filter problem in terms of the conditional means and covariance matrices of the forecast and analysed state PDFs.

The forecast equations of the EKF require only a part of assumptions (3)–(6). Suppose the conditional PDF $p(\mathbf{x}_{k-\Delta k}^{\mathrm{t}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}})$ at time $t_{k-\Delta k}$ is given in terms of the conditional mean $\mathbf{x}_{k-\Delta k}^{\mathrm{a}} := \langle \mathbf{x}_{k-\Delta k}^{\mathrm{t}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}} \rangle$, denoted the 'analysis state', and the 'analysis covariance matrix' $\mathbf{P}_{k-\Delta k}^{\mathrm{a}} := \langle (\mathbf{x}_{k-\Delta k}^{\mathrm{t}} - \mathbf{x}_{k-\Delta k}^{\mathrm{a}})(\mathbf{x}_{k-\Delta k}^{\mathrm{t}} - \mathbf{x}_{k-\Delta k}^{\mathrm{a}})^{\mathrm{T}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}} \rangle$. In the forecast step, the EKF evolves the PDF forward up to time t_k by computing the mean and covariance matrix of $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}})$. The forecast equations are based on a Taylor expansion to eq. (1) at the last state estimate $\mathbf{x}_{i-1}^{\mathrm{a}}$

$$\mathbf{x}_i^{\mathrm{t}} = M_{i,i-1}\left[\mathbf{x}_{i-1}^{\mathrm{a}}\right] + \mathbf{M}_{i,i-1}\mathbf{z}_{i-1}^{\mathrm{a}} + \boldsymbol{\eta}_i + \mathcal{O}(\mathbf{z}^2) \tag{7}$$

where $\mathbf{z}_{i-1}^{\mathrm{a}} := \mathbf{x}_{i-1}^{\mathrm{t}} - \mathbf{x}_{i-1}^{\mathrm{a}}$ and $\mathbf{M}_{i,i-1}$ is the linearization of the operator $M_{i,i-1}$ around the estimate $\mathbf{x}_{i-1}^{\mathrm{a}}$. The 'forecast state' of the EKF is obtained as the conditional mean $\mathbf{x}_k^{\mathrm{f}} = \langle \mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}} \rangle$ while neglecting in eq. (7) terms of higher than linear order in $\mathbf{z}^{\mathrm{a}}$. Under the assumption that the model error has zero mean it is

$$\mathbf{x}_i^{\mathrm{f}} = M_{i,i-1}\left[\mathbf{x}_{i-1}^{\mathrm{a}}\right]. \tag{8}$$

This equation is iterated until time t_k to obtain $\mathbf{x}_k^{\mathrm{f}}$. The corresponding 'forecast covariance matrix' follows to first order in $\mathbf{z}^{\mathrm{a}}$ from eqs. (7) and (8) as

$$\begin{aligned}\mathbf{P}_k^{\mathrm{f}} &:= \left\langle \left(\mathbf{x}_k^{\mathrm{t}} - \mathbf{x}_k^{\mathrm{f}}\right)\left(\mathbf{x}_k^{\mathrm{t}} - \mathbf{x}_k^{\mathrm{f}}\right)^{\mathrm{T}} \middle| \mathbf{Y}_{k-\Delta k}^{\mathrm{o}} \right\rangle \\ &= \mathbf{M}_{k,k-\Delta k}\mathbf{P}_{k-\Delta k}^{\mathrm{a}}\mathbf{M}_{k,k-\Delta k}^{\mathrm{T}} + \mathbf{Q}_k,\end{aligned} \tag{9}$$

where the assumption (6) that $\mathbf{x}_k^{\mathrm{t}}$ and $\boldsymbol{\eta}_k$ are uncorrelated is used. The forecast step of the EKF is described by eqs. (8) and (9). The statistical assumptions required for the derivation of these equations are only that $\mathbf{x}_k^{\mathrm{t}}$ and $\boldsymbol{\eta}_k$ are uncorrelated processes, and that the model error is unbiased. The PDFs are not required to be Gaussian.

The analysis step of the EKF computes the mean and covariance matrix of $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_k^{\mathrm{o}})$ given the PDF $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_{k-\Delta k}^{\mathrm{o}})$ and an observation vector $\mathbf{y}_k^{\mathrm{o}}$ which is available at time T_k. Under the

assumption that $\boldsymbol{\epsilon}_k$ is white in time, the solution is given by the Bayes theorem as

$$p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{Y}_k^{\mathrm{o}}) = \frac{p\left(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}}\right) p\left(\mathbf{x}_k^{\mathrm{t}}|\mathbf{Y}_{k-\Delta k}^{\mathrm{o}}\right)}{p\left(\mathbf{y}_k^{\mathrm{o}}|\mathbf{Y}_{k-\Delta k}^{\mathrm{o}}\right)}. \tag{10}$$

This relation only implies the whiteness of $\boldsymbol{\epsilon}_k$. However, the full set of assumptions (3)–(6) is required to compute the analysis in terms of the mean and covariance matrix of $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_k^{\mathrm{o}})$. The EKF analysis equations are based on a Taylor expansion to the observation model (2) at the forecast state $\mathbf{x}_k^{\mathrm{f}}$. Neglecting in the expansion terms of higher than linear order in $\mathbf{z}_k^{\mathrm{f}} = \mathbf{x}_k^{\mathrm{t}} - \mathbf{x}_k^{\mathrm{f}}$ the analysis state $\mathbf{x}_k^{\mathrm{a}}$ and analysis covariance matrix $\mathbf{P}_k^{\mathrm{a}}$ are obtained as

$$\mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{f}} + \mathbf{K}_k\left(\mathbf{y}_k^{\mathrm{o}} - H_k\left[\mathbf{x}_k^{\mathrm{f}}\right]\right), \tag{11}$$

$$\mathbf{P}_k^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^{\mathrm{f}}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}} \tag{12}$$

$$= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^{\mathrm{f}}, \tag{13}$$

where $\mathbf{H}_k$ denotes the linearization of the measurement operator H_k around $\mathbf{x}_k^{\mathrm{f}}$. $\mathbf{K}_k$ is denoted 'Kalman gain'. It is defined by

$$\mathbf{K}_k = \mathbf{P}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\mathbf{P}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)^{-1} = \mathbf{P}_k^{\mathrm{a}}\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}, \tag{14}$$

where the latter equality requires that $\mathbf{R}_k$ is invertible. Equations (11)–(14) complete the EKF.

To apply the EKF we need to initialize the filter sequence. For this, we have to supply an initial state estimate $\mathbf{x}_0^{\mathrm{a}}$ and a corresponding covariance matrix $\mathbf{P}_0^{\mathrm{a}}$ which represent the initial PDF $p(\mathbf{x}_0^{\mathrm{t}})$.

Remark 1. The forecast of the EKF is due to linearization. The state forecast is only valid up to linear order in $\mathbf{z}$ while the covariance forecast is valid up to second order ($\mathbf{z}^2 \propto \mathbf{P}^{\mathrm{a}}$). The covariance matrix is forecast by the linearized model. For nonlinear dynamics, this neglect of higher-order terms can lead to instabilities of the filter algorithm (Evensen, 1992).

Remark 2. The covariance matrix $\mathbf{P}$ is symmetric positive semidefinite. In a numerical implementation of the KF this property is not guaranteed to be conserved, if eq. (13) is used to update $\mathbf{P}$ because the operations on this matrix are not symmetric. In contrast, eq. (12) preserves the symmetry.

Remark 3. For linear models, the KF yields the optimal minimum variance estimate if the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ as well as the initial state estimate $(\mathbf{x}_0^{\mathrm{a}}, \mathbf{P}_0^{\mathrm{a}})$ are correctly prescribed. Then the estimate is also the maximum likelihood estimate for the PDF $p(\mathbf{x}_k^{\mathrm{t}} \mid \mathbf{Y}_k^{\mathrm{o}})$ (see Jazwinski, 1970, chap. 5.3). For nonlinear systems, the EKF can only yield an approximation of the optimal estimate. For large-scale systems, as in oceanography where the state dimension can be of the order of 10^6, there are generally only estimates of the matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$ available. Also, $\mathbf{x}_0^{\mathrm{a}}$ is in general only an estimate of the initial system state. Due to this, the practical filter estimate is suboptimal.

Remark 4. For large-scale systems the largest computational cost resides in the forecast of the state covariance matrix by eq. (9). This requires $2n$ applications of the (linearized) model operator. For large-scale systems, the corresponding computational cost is not feasible. In addition, the storage of the covariance matrix is required, which contains n^2 elements. This is also not feasible for large-scale models and current size of computer memory.

2.3. Error subspace Kalman filters

The large computational cost of the EKF shows that a direct application of this algorithm to realistic models with large state dimension is not feasible. This problem has led to the development of a number of approximating algorithms from which three variants are examined here.

This work focuses on three algorithms: the EnKF (Evensen, 1994; Burgers et al., 1998), the SEEK filter (Pham et al., 1998a) and the SEIK filter (Pham et al., 1998b). As far as possible the filters are presented here in the unified notation (Ide et al., 1997) following the way they have originally been introduced by the respective authors. The relation of the filters to the EKF as well as possible variations and particular features of them are discussed.

All three algorithms use a low-rank representation of the covariance matrix either by a random ensemble or by an explicit low-rank approximation of the matrix. Thus, the filter analyses operate only in a low-dimensional subspace, denoted the error subspace, which approximates the full error space. As the three algorithms use the analysis equations of the EKF adapted to the particular method, we refer to the algorithms as ESKFs. This corresponds to the concept of ESSE (Lermusiaux and Robinson, 1999).

2.3.1. Singular evolutive extended Kalman filter. The SEEK filter (Pham et al., 1998a) is a so-called reduced rank filter. It is based on the EKF with an approximation of the covariance matrix $\mathbf{P}_0^{\mathrm{a}}$ by a singular matrix and its treatment in decomposed form.

From the point of view of statistics, the rank reduction is motivated by the fact that the PDF $p(\mathbf{x}_0^{\mathrm{t}})$ is not isotropic in state space. If the PDF is Gaussian it can be described by a probability ellipsoid, whose centre is given by the mean $\mathbf{x}_0^{\mathrm{a}}$ and the shape is described by $\mathbf{P}_0^{\mathrm{a}}$. The principal axes of the ellipsoid are found by an eigenvalue decomposition of $\mathbf{P}_0^{\mathrm{a}}$: $\mathbf{P}\mathbf{v}_{(l)} = \lambda_{(l)}\mathbf{v}_{(l)}$, $l = 1, \ldots, n$, where $\mathbf{v}_{(l)}$ is the lth eigenvector and $\lambda_{(l)}$ the corresponding eigenvalue. Hence, the principal vectors are $\{\tilde{\mathbf{v}}_{(l)} = \lambda_{(l)}^{1/2}\mathbf{v}_{(l)}\}$. Approximating $\mathbf{P}_0^{\mathrm{a}}$ by the r ($r \ll n$) largest eigenmodes take into account only the most significant principal axes of the probability ellipsoid. Mathematically, this provides the best rank-r approximation of $\mathbf{P}_0^{\mathrm{a}}$ (see Golub and van Loan, 1989). The retained principal directions define a tangent space at the state space point $\mathbf{x}_0^{\mathrm{a}}$. This error subspace approximates the full error space given by the full covariance matrix. The error subspace is evolved up to the next analysis time of the filter by forecasting the basis vectors $\{\mathbf{v}_{(l)}\}$. In the analysis step the filter operates only in the most significant directions of uncertainty given by the error subspace.

The SEEK filter is described by the following equations.

Initialization. The initial PDF $p(\mathbf{x}_0^{\mathrm{t}})$ is provided by the initial state estimate $\mathbf{x}_0^{\mathrm{a}}$ and a rank-r approximation ($r \ll n$) of the covariance matrix $\mathbf{P}_0^{\mathrm{a}}$ given in decomposed form

$$\mathbf{x}_0^{\mathrm{a}} = \langle \mathbf{x}_0^{\mathrm{t}} \rangle; \quad \hat{\mathbf{P}}_0^{\mathrm{a}} := \mathbf{V}_0 \mathbf{U}_0 \mathbf{V}_0^{\mathrm{T}} \approx \mathbf{P}_0^{\mathrm{a}}. \tag{15}$$

Here the diagonal matrix $\mathbf{U}_0$ holds the r largest eigenvalues. Matrix $\mathbf{V}_0$ is of dimension $n \times r$ and contains in its columns the corresponding eigenmodes of $\hat{\mathbf{P}}_0^{\mathrm{a}}$, where we denote with the hat symbol ($\hat{}$) quantities that are particular for the SEEK filter.

A popular choice for $\mathbf{V}_0$ is the matrix of empirical orthogonal functions (EOFs) of a sequence of model states sampled from a model integration over some period. However, this is not necessary when better estimates of $\mathbf{P}_0^{\mathrm{a}}$ exist.

Forecast. The forecast equations of the SEEK filters are derived from the EKF by treating the covariance matrix in decomposed form as provided by the initialization:

$$\mathbf{x}_i^{\mathrm{f}} = M_{i,i-1}\left[\mathbf{x}_{i-1}^{\mathrm{a}}\right] \tag{16}$$

$$\mathbf{V}_k = \mathbf{M}_{k,k-\Delta k} \mathbf{V}_{k-\Delta k}. \tag{17}$$

Analysis. The analysis equations are a reformulation of the EKF analysis equations for a covariance matrix given in decomposed form. To maintain the rank r of $\hat{\mathbf{P}}_0^{\mathrm{a}}$ the model error covariance matrix $\mathbf{Q}_k$ is projected onto the error subspace by

$$\hat{\mathbf{Q}}_k := \left(\mathbf{V}_k^{\mathrm{T}} \mathbf{V}_k\right)^{-1} \mathbf{V}_k^{\mathrm{T}} \mathbf{Q}_k \mathbf{V}_k \left(\mathbf{V}_k^{\mathrm{T}} \mathbf{V}_k\right)^{-1}. \tag{18}$$

With this, the analysis equations of the SEEK filter are for an invertible matrix $\mathbf{R}_k$

$$\mathbf{U}_k^{-1} = (\mathbf{U}_{k-\Delta k} + \hat{\mathbf{Q}}_k)^{-1} + (\mathbf{H}_k \mathbf{V}_k)^{\mathrm{T}} \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{V}_k, \tag{19}$$

$$\mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{f}} + \hat{\mathbf{K}}_k\left(\mathbf{y}_k^{\mathrm{o}} - H_k\left[\mathbf{x}_k^{\mathrm{f}}\right]\right), \tag{20}$$

$$\hat{\mathbf{K}}_k = \mathbf{V}_k \mathbf{U}_k \mathbf{V}_k^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} {\mathbf{R}_k}^{-1}. \tag{21}$$

The analysis covariance matrix is implicitly given by $\hat{\mathbf{P}}_k^{\mathrm{a}} := \mathbf{V}_k \mathbf{U}_k \mathbf{V}_k^{\mathrm{T}}$.

Re-initialization. The mode matrix $\mathbf{V}_k$ can be directly used to evaluate the next forecast step. However, to avoid that the modes $\{\mathbf{v}_{(i)}\}$ become large and more and more aligned, a re-orthonormalization of these vectors is useful. This can be performed by computing the eigenvalue decomposition of the $r \times r$ matrix

$$\mathbf{B}_k := \mathbf{A}_k^{\mathrm{T}} \mathbf{V}_k^{\mathrm{T}} \mathbf{V}_k \mathbf{A}_k, \tag{22}$$

where $\mathbf{A}_k$ is obtained from a Cholesky decomposition $\mathbf{A}_k \mathbf{A}_k^{\mathrm{T}} = \mathbf{U}_k$. The eigenvalues of $\mathbf{B}_k$ are the same as the non-zero eigenvalues of $\hat{\mathbf{P}}_k^{\mathrm{a}}$. Let $\mathbf{B}_k = \mathbf{C}_k \mathbf{D}_k \mathbf{C}_k^{\mathrm{T}}$ be the eigenvalue decomposition of $\mathbf{B}_k$ where $\mathbf{C}_k$ contains in its columns the eigenvectors and the diagonal matrix $\mathbf{D}_k$ the corresponding eigenvalues. Then the re-orthonormalized error subspace basis $\hat{\mathbf{V}}$ and corresponding eigenvalue matrix $\hat{\mathbf{U}}$ are given by

$$\hat{\mathbf{V}}_k = \mathbf{V}_k \mathbf{A}_k \mathbf{C}_k \mathbf{D}_k^{-1/2}; \quad \hat{\mathbf{U}}_k = \mathbf{D}_k. \tag{23}$$

Remark 5. The algorithm is designed to treat the covariance matrix in the decomposed form $\hat{\mathbf{P}} = \mathbf{V}\mathbf{U}\mathbf{V}^{\mathrm{T}}$. Using a truncated eigenvalue decomposition of a prescribed matrix $\mathbf{P}_0^{\mathrm{a}}$ yields mathematically the best approximation of this matrix. $\mathbf{P}_0^{\mathrm{a}}$ can also be given in implicit form, e.g. as the perturbation matrix of a state trajectory. In this case the rank reduction and decomposition of $\mathbf{P}_0^{\mathrm{a}}$ can be computed by a singular value decomposition of the perturbation matrix without explicitly computing the matrix $\mathbf{P}_0^{\mathrm{a}}$. However, we like to stress once more that the matrix $\mathbf{V}_0$ need not be derived from an EOF analysis.

Remark 6. The covariance forecast is computed by forecasting the r modes of $\hat{\mathbf{P}}$. With typically $r < 100$ this brings the forecast step toward acceptable computation times.

Remark 7. The SEEK filter is a reformulation of the EKF. It focuses on the analysed state estimate and covariance matrix. The SEEK filter, however, inherits the stability problem of the EKF by considering only the two lowest statistical moments of the PDF. If r is too small, this problem is even amplified, as $\hat{\mathbf{P}}^{\mathrm{a}}$ systematically underestimates the variance prescribed by the full covariance matrix $\mathbf{P}^{\mathrm{a}}$.

Remark 8. In practice it can be difficult to specify the linearized dynamic model operator $\mathbf{M}_{i,i-1}$. Alternatively, one can apply a finite difference approximation. That is, the forecast of column α of $\mathbf{V}_{i-1}^{\mathrm{a}}$, denoted by $\mathbf{V}_{i-1,\alpha}^{\mathrm{a}}$, is given by

$$\mathbf{M}_{i,i-1} \mathbf{V}_{i-1,\alpha}^{\mathrm{a}} \approx \frac{M_{i,i-1}\left[\mathbf{x}_{i-1}^{\mathrm{a}} + \epsilon \mathbf{V}_{i-1,\alpha}^{\mathrm{a}}\right] - M_{i,i-1}\left[\mathbf{x}_{i-1}^{\mathrm{a}}\right]}{\epsilon}. \tag{24}$$

For a finite difference approximation the coefficient ϵ needs to be a small positive number ($\epsilon \ll 1$). Some authors (Voorrips et al., 1999; Heemink et al., 2001) report the use of $\epsilon \approx 1$. This can bring the algorithm beyond a purely tangent-linear forecast, but it is no longer defined as a finite difference approximation and would require an ensemble interpretation.

Sometimes the use of the gradient approximation (24) is denoted as the interpolated variant of the SEEK filter (i.e. as SEIK). However, this should not be confused with the SEIK algorithm by Pham et al. (1998b) which involves many more steps (see below).

Remark 9. The increment for the analysis update of the state estimate in eq. (20) is computed as a weighted average of the mode vectors in $\mathbf{V}_k$ which belong to the error subspace. This becomes visible when the definition of the Kalman gain (eq. 21) is inserted into eq. (20):

$$\mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{f}} + \mathbf{V}_k \left[\mathbf{U}_k \mathbf{V}_k^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} {\mathbf{R}_k}^{-1} \left(\mathbf{y}_k^{\mathrm{o}} - H_k\left[\mathbf{x}_k^{\mathrm{f}}\right]\right)\right]. \tag{25}$$

The term in brackets represents a vector of weights for combining the modes $\mathbf{V}$.

Remark 10. Equation (19) for the matrix $\mathbf{U}_k$ can be modified by multiplying with a so-called forgetting factor ρ, ($0 < \rho \leq 1$)

(Pham et al., 1998a):

$$\mathbf{U}_k^{-1} = \left(\rho^{-1}\mathbf{U}_{k-\Delta k} + \hat{\mathbf{Q}}_k\right)^{-1} + (\mathbf{H}_k\mathbf{V}_k)^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{V}_k. \tag{26}$$

The forgetting factor can be used as a tuning parameter of the analysis step to downweight the state forecast relative to the observations. This can increase the filter stability as the systematic underestimation of the variance is reduced.

Remark 11. In eq. (17) the modes $\mathbf{V}$ of $\hat{\mathbf{P}}$ are evolved with initially unit norm. However, it is also possible to use modes scaled by the square root of the corresponding eigenvalue and matrix $\mathbf{U}$ being the identity matrix. Then the rediagonalization should be performed after each analysis step, replacing eq. (23) by $\hat{\mathbf{V}}_k = \mathbf{V}_k\mathbf{C}_k$ and $\hat{\mathbf{U}}_k = \mathbf{I}_{r\times r}$. This scaled algorithm is equivalent to the RRSQRT algorithm by Verlaan and Heemink (1995).

2.3.2. Ensemble Kalman filter. The EnKF (Evensen, 1994; Burgers et al., 1998) has been introduced as a Monte Carlo method to sample and forecast the PDF. The initial density $p(\mathbf{x}_0^{\mathrm{t}})$ is sampled by a finite random ensemble of state realizations. Each ensemble state is forecast with the stochastic model (1) and updated in the analysis step.

From the statistical point of view, the EnKF solves, for sufficiently large ensembles, the Fokker–Planck–Kolmogorov equation for the evolution of the PDF $p(\mathbf{x}^{\mathrm{t}})$ by a Monte Carlo method. In contrast to the SEEK algorithm, where the rank reduction directly uses the assumption that the PDF is Gaussian and thus can be described by a probability ellipsoid, the EnKF samples the PDF by a random ensemble of N model states $\{\mathbf{x}_0^{\mathrm{a}(\alpha)}, \alpha = 1, \ldots, N\}$. Denoting by d$N$ the number of ensemble states lying within some volume element in state space, the PDF $p(\mathbf{x}^{\mathrm{t}})$ is approximated by the ensemble member density dN/N in state space. This sampling of $p(\mathbf{x}_0^{\mathrm{t}})$ converges rather slowly (proportional to $N^{-1/2}$), but it is valid for any kind of PDF, not just Gaussian PDFs. Forecasting each $\mathbf{x}_0^{\mathrm{a}(\alpha)}$ with the stochastic-dynamic model (1) evolves the sampled PDF with the non-linear model up to the next analysis time. In the analysis step, the EKF analysis, which implies that the PDFs are Gaussian, is applied to each of the ensemble states. The covariance matrix $\mathbf{P}$ is approximated for the analysis by the ensemble covariance matrix $\tilde{\mathbf{P}}$. Because the rank of $\tilde{\mathbf{P}}$ is at most $N-1$, the EnKF also operates in an error subspace which is determined by the random sampling. To ensure that the ensemble analysis represents the combination of two PDFs, a random ensemble of observations is required in the analysis step (Burgers et al., 1998). Each ensemble state is then updated using a vector from this observation ensemble. This implicitly updates the state covariance matrix.

The EnKF algorithm according to Evensen (1994) is described by the following equations.

Initialization. The initial PDF $p(\mathbf{x}_0^{\mathrm{t}})$ is sampled by a random ensemble

$$\left\{\mathbf{x}_0^{\mathrm{a}(l)}, l = 1, \ldots, N\right\} \tag{27}$$

of N state realizations: The statistics of this ensemble approximate the initial state estimate and the corresponding covariance matrix. Thus, for $N \to \infty$

$$\overline{\mathbf{x}_0^{\mathrm{a}}} = \frac{1}{N}\sum_{l=1}^{N}\mathbf{x}_0^{\mathrm{a}(l)} \to \left\langle\mathbf{x}_0^{\mathrm{t}}\right\rangle, \tag{28}$$

$$\tilde{\mathbf{P}}_0^{\mathrm{a}} := \frac{1}{N-1}\sum_{l=1}^{N}\left(\mathbf{x}_0^{\mathrm{a}(l)} - \overline{\mathbf{x}_0^{\mathrm{a}}}\right)\left(\mathbf{x}_0^{\mathrm{a}(l)} - \overline{\mathbf{x}_0^{\mathrm{a}}}\right)^{\mathrm{T}} \to \mathbf{P}_0^{\mathrm{a}}, \tag{29}$$

where the tilde is used to characterize quantities which are particular for the EnKF algorithm.

Forecast. Each ensemble member is evolved up to time t_k with the non-linear stochastic-dynamic model (1) as

$$\mathbf{x}_i^{\mathrm{a}(l)} = M_{i,i-1}[\mathbf{x}_{i-1}^{\mathrm{a}(l)}] + \boldsymbol{\eta}_i^{(l)}, \tag{30}$$

where each ensemble state is subject to individual noise $\boldsymbol{\eta}_i^{(l)}$.

Analysis. For the analysis a random ensemble of observation vectors $\{\mathbf{y}_k^{\mathrm{o}(l)}, l = 1, \ldots, N\}$ is generated. The ensemble statistics approximate the observation error covariance matrix $\mathbf{R}_k$. Each ensemble member is updated analogously to the EKF analysis by

$$\mathbf{x}_k^{\mathrm{a}(l)} = \mathbf{x}_k^{\mathrm{f}(l)} + \tilde{\mathbf{K}}_k\left(\mathbf{y}_k^{\mathrm{o}(l)} - H_k\left[\mathbf{x}_k^{\mathrm{f}(l)}\right]\right), \tag{31}$$

$$\tilde{\mathbf{K}}_k = \tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)^{-1}, \tag{32}$$

$$\tilde{\mathbf{P}}_k^{\mathrm{f}} = \frac{1}{N-1}\sum_{l=1}^{N}\left(\mathbf{x}_k^{\mathrm{f}(l)} - \overline{\mathbf{x}_k^{\mathrm{f}}}\right)\left(\mathbf{x}_k^{\mathrm{f}(l)} - \overline{\mathbf{x}_k^{\mathrm{f}}}\right)^{\mathrm{T}}. \tag{33}$$

The analysis state and covariance matrix are then defined by the ensemble mean and covariance matrix as

$$\mathbf{x}_k^{\mathrm{a}} := \frac{1}{N}\sum_{l=1}^{N}\mathbf{x}_k^{\mathrm{a}(l)}, \tag{34}$$

$$\tilde{\mathbf{P}}_k^{\mathrm{a}} := \frac{1}{N-1}\sum_{l=1}^{N}\left(\mathbf{x}_k^{\mathrm{a}(l)} - \mathbf{x}_k^{\mathrm{a}}\right)\left(\mathbf{x}_k^{\mathrm{a}(l)} - \mathbf{x}_k^{\mathrm{a}}\right)^{\mathrm{T}}, \tag{35}$$

which complete the analysis equations of the EnKF.

An efficient implementation of this analysis is formulated in terms of 'representers' (Bennett, 1992; Evensen and van Leeuwen, 1996). This formulation also allows us to handle the situation when $\mathbf{H}_k\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}$ is singular, which will occur if $m_k > N$. The state analysis eq. (31) is rewritten as

$$\mathbf{x}_k^{\mathrm{a}(l)} = \mathbf{x}_k^{\mathrm{f}(l)} + \tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}\mathbf{b}_k^{(l)}. \tag{36}$$

The columns of the matrix $\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}$ are called representers and constitute influence vectors for each of the measurements. Amplitudes for the influence vectors are given by the vectors $\{\mathbf{b}_k^{(l)}\}$ which are obtained as the solution of

$$\left(\mathbf{H}_k\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)\mathbf{b}_k^{(l)} = \mathbf{y}_k^{\mathrm{o}(l)} - H_k\left[\mathbf{x}_k^{\mathrm{f}(l)}\right]. \tag{37}$$

In addition, explicit computation of $\tilde{\mathbf{P}}_k^{\mathrm{f}}$ by eq. (33) is not needed. It suffices to compute (see, for example, Houtekamer and Mitchell, 1998):

$$\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}} = \frac{1}{N-1}\sum_{l=1}^{N}\left(\mathbf{x}_k^{\mathrm{f}(l)} - \overline{\mathbf{x}_k^{\mathrm{f}}}\right)\left[\mathbf{H}_k\left(\mathbf{x}_k^{\mathrm{f}(l)} - \overline{\mathbf{x}_k^{\mathrm{f}}}\right)\right]^{\mathrm{T}}, \tag{38}$$

$$\mathbf{H}_k\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}} = \frac{1}{N-1}\sum_{l=1}^{N}\mathbf{H}_k\left(\mathbf{x}_k^{\mathrm{f}(l)} - \overline{\mathbf{x}_k^{\mathrm{f}}}\right)\left[\mathbf{H}_k\left(\mathbf{x}_k^{\mathrm{f}(l)} - \overline{\mathbf{x}_k^{\mathrm{f}}}\right)\right]^{\mathrm{T}}. \quad (39)$$

The EnKF comprises some particular features due to the use of a Monte Carlo method in all steps of the filter.

Remark 12. Using a Monte Carlo sampling of the initial PDF non-Gaussian densities can also be represented. As the sampling converges slowly with $N^{-1/2}$, rather large ensembles ($N \geq 100$) are required (Evensen, 1994; Evensen and van Leeuwen, 1996) to avoid too large sampling errors.

Remark 13. The forecast step evolves all N ensemble states with the non-linear model. This also allows for non-Gaussian densities. Algorithmically, the ensemble evolution has the benefit that a linearized model operator is not required.

Remark 14. The analysis step is derived from the EKF. Thus, it assumes Gaussian PDFs and only accounts for the two lowest statistical moments of the PDF. Using the mean of the forecast ensemble as the state forecast estimate leads for sufficiently large ensembles to a more accurate estimate than in the EKF. From the Taylor expansion (eq. 7), it is obvious that this takes into account higher-order terms than the EKF does. In contrast to the EKF and SEEK filters $\mathbf{P}$ is only updated implicitly by the analysis of the ensemble states.

Remark 15. The generation of an observation ensemble is required to ensure consistent statistics of the updated state ensemble (Burgers et al., 1998; Houtekamer and Mitchell, 1998). With the observation ensemble the covariance matrix $\mathbf{R}_k$ in eq. (12) is represented as $\tilde{\mathbf{R}}_k$. However, this introduces additional sampling errors to the ensemble which are largest when the ensemble is small compared to the rank of $\mathbf{R}_k$ (e.g. if $\mathbf{R}_k$ is diagonal). Furthermore, it is likely that the state and observation ensembles have spurious correlations. This introduces an additional error term in eq. (12) (see Whitaker and Hamill, 2002).

Remark 16. While for sufficiently large ensembles the EnKF can be considered as solving the Fokker–Planck–Kolmogorov equation by a Monte Carlo method, this is not valid for very small ensembles. In this case, the EnKF needs to be regarded as an error subspace method.

Remark 17. Combining eqs. (31), (32) and (38) it becomes obvious that the analysis increments for the ensemble states are computed as weighted means of the error subspace vectors $\{\mathbf{x}_k^{\mathrm{f}(l)} - \overline{\mathbf{x}_k^{\mathrm{f}}}\}$. Alternatively, the analysis can also be interpreted as a weakly non-linear combination of the ensemble states (Evensen, 2003). However, the latter interpretation hides the error subspace character of the algorithm.

Remark 18. If the number of observations is larger than the ensemble size, it will be costly to compute the matrix $\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}$ explicitly according to eq. (38). In this case, it is more efficient to change the order of matrix computations such that $\tilde{\mathbf{P}}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}$ is not computed explicitly.

Remark 19. In eqs. (32) and (37) it is possible to use, instead of the prescribed matrix $\mathbf{R}_k$, the matrix $\tilde{\mathbf{R}}_k$ as sampled by the observation ensemble $\{\mathbf{y}_k^{\mathrm{o}(l)}\}$. This allows for a computationally very efficient analysis scheme as proposed by Evensen (2003). However, due to the sampling problems of $\mathbf{R}_k$ this can lead to a further degradation of the filter quality.

2.3.3. Singular evolutive interpolated Kalman filter. The SEIK filter (Pham et al., 1998b) has been derived as a variant of the SEEK algorithm using interpolation instead of linearization for the forecast step. Alternatively, the SEIK filter can be interpreted as an ensemble KF using a preconditioned ensemble and a computationally very efficient analysis formulation. The SEIK algorithm should not be mixed up with other interpolated variants of the SEEK filter, as in Verron et al. (1999), which typically correspond to the SEEK filter with finite difference approximation (eq. 24).

Statistically, the initialization of the SEIK filter is analogous to that of the SEEK algorithm: the PDF $p(\mathbf{x}_0^{\mathrm{t}})$ is again represented by the principal axes of $\mathbf{P}_0^{\mathrm{a}}$ and approximated by the r largest eigenmodes. However, the SEIK algorithm does not evolve the eigenmodes directly but generates a stochastic ensemble of $r+1$ state realizations. This ensemble exactly represents the mean and covariance matrix of the approximated PDF. The PDF is forecast by evolving each of the ensemble members with the non-linear model as in the EnKF. The evolved error subspace is determined by computing the state forecast estimate and covariance matrix from the ensemble. The analysis is performed analogously to the SEEK filter followed by a re-initialization.

The SEIK filter is described by the following equations.

Initialization. The initial PDF $p(\mathbf{x}_0^{\mathrm{t}})$ is provided by eq. (15) as the initial state estimate $\mathbf{x}_0^{\mathrm{a}}$ and a rank-r approximation of $\mathbf{P}_0^{\mathrm{a}}$ given in decomposed form. From this information an ensemble

$$\left\{\mathbf{x}_0^{\mathrm{a}(l)},\ l = 1, \ldots, r+1\right\} \quad (40)$$

of $r+1$ state realizations is generated which fulfils

$$\overline{\mathbf{x}_0^{\mathrm{a}}} \equiv \mathbf{x}_0^{\mathrm{a}}, \quad (41)$$

$$\check{\mathbf{P}}_0^{\mathrm{a}} := \frac{1}{r+1}\sum_{l=1}^{r+1}\left(\mathbf{x}_0^{\mathrm{a}(l)} - \overline{\mathbf{x}_0^{\mathrm{a}}}\right)\left(\mathbf{x}_0^{\mathrm{a}(l)} - \overline{\mathbf{x}_0^{\mathrm{a}}}\right)^{\mathrm{T}} \equiv \hat{\mathbf{P}}_0^{\mathrm{a}} \quad (42)$$

where the check symbol ($\check{}$) is used to characterize quantities particular to the SEIK filter.

To ensure that eqs. (41) and (42) hold, the ensemble is generated in a procedure called minimum second-order exact sampling (see, for example, Pham, 2001). For this, let $\mathbf{C}_0$ contain in its diagonal the square roots of the eigenvalues of $\hat{\mathbf{P}}_0^{\mathrm{a}}$, such that $\mathbf{U}_0 = \mathbf{C}_0^{\mathrm{T}}\mathbf{C}_0$. Then $\check{\mathbf{P}}_0^{\mathrm{a}}$ is written as

$$\check{\mathbf{P}}_0^{\mathrm{a}} = \mathbf{V}_0\mathbf{C}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{\mathrm{T}}\boldsymbol{\Omega}_0\mathbf{C}_0\mathbf{V}_0^{\mathrm{T}}, \quad (43)$$

where $\boldsymbol{\Omega}_0$ is an $(r+1)\times r$ random matrix whose columns are orthonormal and orthogonal to the vector $(1, \ldots, 1)^{\mathrm{T}}$ which can be obtained by Householder reflections (see, for example, Hoteit et al., 2002). The state realizations of the ensemble are then given by

$$\mathbf{x}_0^{\mathrm{a}(l)} = \mathbf{x}_0^{\mathrm{a}} + \sqrt{r+1}\,\mathbf{V}_0\mathbf{C}_0^{\mathrm{T}}\boldsymbol{\Omega}_{0,l}^{\mathrm{T}}, \quad (44)$$

where $\boldsymbol{\Omega}_{0,l}^{\mathrm{T}}$ denotes the lth column of $\boldsymbol{\Omega}_0^{\mathrm{T}}$.

The formulation of the SEIK filter is based on an efficient description of $\check{\mathbf{P}}_0^{\mathrm{a}}$ in terms of the ensemble states. Denoting $\mathbf{X}_0^{\mathrm{a}} = [\mathbf{x}_0^{\mathrm{a}(1)}, \ldots, \mathbf{x}_0^{\mathrm{a}(r+1)}]$ the matrix whose columns are the ensemble state vectors, we have

$$\check{\mathbf{P}}_0^{\mathrm{a}} = \frac{1}{r+1}\mathbf{X}_0^{\mathrm{a}}\mathbf{T}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\left(\mathbf{X}_0^{\mathrm{a}}\right)^{\mathrm{T}}. \tag{45}$$

Here $\mathbf{T}$ is an $(r+1) \times r$ matrix with zero column sums. A possible choice for $\mathbf{T}$ is

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_{r\times r} \\ \mathbf{0}_{1\times r} \end{pmatrix} - \frac{1}{r+1}\left[\mathbf{1}_{(r+1)\times r}\right]. \tag{46}$$

Here $\mathbf{0}$ is the matrix holding only zeros and $\mathbf{1}$ is the matrix with only unit entries. Matrix $\mathbf{T}$ fulfils the purpose of implicitly subtracting the ensemble mean when computing $\check{\mathbf{P}}_0^{\mathrm{a}}$. Equation (45) can be written in a form analogous to the covariance matrix in eq. (15) as

$$\check{\mathbf{P}}_0^{\mathrm{a}} = \mathbf{L}_0\mathbf{G}\mathbf{L}_0^{\mathrm{T}} \tag{47}$$

with

$$\mathbf{L}_0 := \mathbf{X}_0^{\mathrm{a}}\mathbf{T}\,; \quad \mathbf{G} := \frac{1}{r+1}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}. \tag{48}$$

Forecast. Each ensemble member is evolved up to time t_k with the non-linear dynamic model equation

$$\mathbf{x}_i^{\mathrm{f}(l)} = M_{i,i-1}\left[\mathbf{x}_{i-1}^{\mathrm{a}(l)}\right]. \tag{49}$$

Analysis. The analysis equations are analogous to the SEEK filter, but here the state forecast estimate is given by the ensemble mean $\overline{\mathbf{x}_k^{\mathrm{f}}}$. To maintain the rank r of $\check{\mathbf{P}}_k$ the matrix $\mathbf{Q}_k$ is projected onto the error subspace, analogously to the SEEK filter, by

$$\check{\mathbf{Q}}_k := \left(\mathbf{L}_k^{\mathrm{T}}\mathbf{L}_k\right)^{-1}\mathbf{L}_k^{\mathrm{T}}\mathbf{G}\mathbf{L}_k\left(\mathbf{L}_k^{\mathrm{T}}\mathbf{L}_k\right)^{-1}. \tag{50}$$

Then, the analysis equations are

$$\mathbf{U}_k^{-1} = [\mathbf{G} + \check{\mathbf{Q}}_k]^{-1} + (\mathbf{H}_k\mathbf{L}_k)^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{L}_k, \tag{51}$$

$$\mathbf{x}_k^{\mathrm{a}} = \overline{\mathbf{x}_k^{\mathrm{f}}} + \check{\mathbf{K}}_k\left(\mathbf{y}_k^{\mathrm{o}} - H_k\left[\overline{\mathbf{x}_k^{\mathrm{f}}}\right]\right), \tag{52}$$

$$\check{\mathbf{K}}_k = \mathbf{L}_k\mathbf{U}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}. \tag{53}$$

The analysis covariance matrix is implicitly given by $\check{\mathbf{P}}_k^{\mathrm{a}} := \mathbf{L}_k\mathbf{U}_k\mathbf{L}_k^{\mathrm{T}}$.

Re-initialization. To proceed with the filter sequence, the ensemble has to be transformed to represent the analysis state and covariance matrix at time t_k. The procedure is analogous to the initial ensemble generation but here a Cholesky decomposition is applied to obtain $\mathbf{U}_k^{-1} = \mathbf{C}_k\mathbf{C}_k^{\mathrm{T}}$. Then $\check{\mathbf{P}}_k^{\mathrm{a}}$ can be written in analogy to eq. (43) as

$$\check{\mathbf{P}}_k^{\mathrm{a}} = \mathbf{L}_k\left(\mathbf{C}_k^{-1}\right)^{\mathrm{T}}\boldsymbol{\Omega}_k^{\mathrm{T}}\,\boldsymbol{\Omega}_k\mathbf{C}_k^{-1}\mathbf{L}_k^{\mathrm{T}}, \tag{54}$$

where $\boldsymbol{\Omega}_k$ has the same properties of orthonormality and orthogonality to $(1, \ldots, 1)^{\mathrm{T}}$ as in the initialization. Accordingly, the ensemble members are given by

$$\mathbf{x}_k^{\mathrm{a}(l)} = \mathbf{x}_k^{\mathrm{a}} + \sqrt{r+1}\,\mathbf{L}_k\left(\mathbf{C}_k^{-1}\right)^{\mathrm{T}}\boldsymbol{\Omega}_{k,l}^{\mathrm{T}}. \tag{55}$$

The SEIK algorithm shares features of both the SEEK filter and the EnKF.

Remark 20. Operating with an ensemble method in an error subspace given by the most significant directions of uncertainty the SEIK filter is similar to the concept of ESSE (Lermusiaux and Robinson, 1999). The analysis is also similar to variants of the EnKF denoted ensemble square-root Kalman filters (Tippett et al., 2003), which do not use an ensemble of observations but perform a transformation of the ensemble after updating the ensemble mean (Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002; Ott et al., 2004). The difference between these filters and the SEIK filter lies in the fact that they compute the analysis update in the observation space rather than the error subspace.

Remark 21. Using second-order exact sampling of the low-rank approximated covariance matrix leads to smaller sampling errors of the ensemble covariance matrix compared to the Monte Carlo sampling in the EnKF.

Remark 22. The ensemble members are evolved with the non-linear model. Thus, as algorithmic benefit, the linearized model operator is not required.

Remark 23. Equation (49) does not include model errors. However, it is possible to extend the equation to treat model errors as a stochastic forcing as in the EnKF (eq. 30).

Remark 24. The forecast state estimate is computed as the mean of the ensemble forecast. Analogous to the EnKF this leads to a forecast accounting for higher-order terms in the Taylor expansion (eq. 7).

Remark 25. As for the SEEK filter, the increment for the analysis update of the state estimate (eq. 52) is computed as a weighted average of the error subspace vectors in $\mathbf{L}_k$. In addition, this is also true for the increments of the ensemble states (eq. 55).

Remark 26. Because the structure of $\mathbf{T}$ is known, the SEIK filter can be implemented without explicitly computing matrix $\mathbf{L}_k$ according to eq. (48) but with $\mathbf{T}$ as an operator.

Remark 27. The analysis and re-initialization steps of the SEIK filter can be reformulated to obtain an ensemble analysis analogous to that of the EnKF. For this, each ensemble member is updated according to eq. (31) with a Kalman gain given by eq. (53) and an observation from a generated ensemble of observations. No re-initialization of the ensemble is required. This algorithm will be less costly than the analysis of the EnKF. However, this variant of the SEIK filter will show the problem of sampling errors which is also inherent to the EnKF (see Remark 15).

3. Comparison of the EnKF, SEEK and SEIK algorithms

For the application of filter algorithms to geophysical modelling problems, we are concerned with the search for filter algorithms

for large-scale non-linear systems. The algorithmic formulations of all three filters look quite similar, which is due to their relation to the EKF. However, the small algorithmic differences can be expected to lead to significantly different filter performances, in particular with non-linear models. Brusdal et al. (2003) compared the EnKF and SEEK filters, but focused rather on the similarity of the algorithms. We compare the filters here with a focus on their differences and we discuss consequences for their application to large-scale non-linear systems.

3.1. Representation of initial error subspaces

The initialization of the algorithms implies their different representation of the error subspace approximating the PDF $p(\mathbf{x}_0^t)$. The Monte Carlo sampling used in the EnKF represents $p(\mathbf{x}_0^t)$ by a random ensemble of model state realizations. This approach allows us to sample arbitrary PDFs but converges rather slowly because the relative importances of the error space directions are not taken into account. The SEEK and SEIK algorithms represent the error subspace by the r major principal axes of the error ellipsoid described by the covariance matrix $\mathbf{P}_0^a$. This implies that the PDF is Gaussian, or at least well described by $\mathbf{P}_0^a$. The SEEK filter treats the covariance matrix directly in its decomposed form given by eigenvectors and a matrix of eigenvalues. The SEIK filter uses a stochastic ensemble of minimum size, generated by minimum second-order exact sampling, whose ensemble statistics exactly represent the approximated $\mathbf{P}_0^a$. For SEEK and SEIK filters, the convergence of the approximation with increasing r depends on the eigenvalue spectrum of $\mathbf{P}_0^a$. Typically, the sampling errors in the SEEK and SEIK filters will be much smaller than in the EnKF.

Despite their different representations of the error subspace, all three filters can be initialized from the same PDF or covariance matrix. For a consistent comparison of the filter performance of different algorithms it is even necessary to use the same initial conditions. Furthermore, the EnKF and SEIK algorithms are in fact independent from the method by which the state ensembles are generated. Thus, the initialization methods of Monte Carlo sampling and second-order exact sampling can be interchanged between the EnKF and the SEIK filter.

3.2. Prediction of error subspaces

The forecast step of the filter algorithms computes a prediction of the state estimate $\mathbf{x}_k^f$ and the error subspace at the next observation time t_k. The SEEK filter evolves the state estimate $\mathbf{x}_{k-\Delta k}^a$ with the non-linear model to predict $\mathbf{x}_k^f$. This only approximates the mean of the evolved PDF. To evolve the basis of the error subspace, the eigenmodes of $\mathbf{P}_{k-\Delta k}^a$ are integrated with the linearized model or a finite difference approximation of it. In contrast, the EnKF and SEIK algorithms rely on non-linear ensemble forecasting. This accounts for higher-order terms in the Taylor expansion (eq. 7).

Due to the different forecast schemes, the EnKF and SEIK algorithms have the potential to provide much more realistic predictions of the error subspace with non-linear models than the SEEK filter. Also, it can be dangerous to directly evolve the modes of $\mathbf{P}_{k-\Delta k}^a$, because this does not represent non-linear interactions between different modes and the increasingly finer scales of higher modes can lead to forecasts which do not provide meaningful directions of the error subspace.

3.3. Treatment of model errors

The SEEK and SEIK filters consider model errors by adding the model error covariance matrix $\mathbf{Q}$ to the forecast state covariance matrix. This is analogous to the EKF, with the difference that the SEEK/SEIK algorithms neglect parts of $\mathbf{Q}$ which are orthogonal to the error subspace to conserve the dimension of the error subspace. Alternatively, a simplified treatment is possible by applying the forgetting factor. This increases the variance in all directions of the error subspace by the same factor.

The EnKF applies a stochastic forcing during the ensemble forecast to account for model errors. It is also possible to use a forgetting factor with the EnKF (see, for example, Hamill and Whitaker, 2001, where it is denoted as 'covariance inflation'). Because the SEIK filter also uses an ensemble forecast, it is possible to apply stochastic forcing in this algorithm, too.

In the context of a non-linear system, the addition of $\mathbf{Q}$ at observation times is only an approximation because over finite time the additive stochastic forcing in eq. (1) will result in non-additive effects. Thus, applying stochastic forcing to the ensemble evolution will, in general, yield a more realistic representation of model errors than the addition of a matrix $\mathbf{Q}$. However, this requires the model errors to be known or, at least, to be well estimated. When the model errors are only poorly known, the forgetting factor provides a simple and numerically very efficient way to account for them and to stabilize the filtering process.

3.4. Analysis step

The analysis step of all three algorithms is based on the EKF analysis. Hence, only the first two statistical moments of the predicted PDF, the mean and covariance matrix, are taken into account. This implies that the analysis step will provide only reasonable and approximately variance minimizing results if the non-Gaussian parts of the PDFs of the state and the observations have negligible influence for the analysis update. For non-linear models this can occur, for example, if the analysed covariance is small (see van Leeuwen, 2001). If this is not fulfilled, the forecast density for non-linear systems will contain non-Gaussian parts which limit the applicability of the analysis equations. However, Brusdal et al. (2003) argued that in large-scale oceanographic applications the state density will usually be close to Gaussian if a sufficient number of observations with Gaussian errors is

assimilated. This will at least be fulfilled, if no strong nonlinearity is present.

The analysis of the SEEK filter is just a reformulation of the EKF update equations for a mode-decomposed covariance matrix $\hat{\mathbf{P}}^{\mathrm{a}}$. The state estimate, given by the central forecast, is updated using a Kalman gain computed from $\hat{\mathbf{P}}^{\mathrm{a}}$, which itself is obtained by updating only the matrix $\mathbf{U}$ of size $r \times r$. The analyses of the EnKF and SEIK algorithms use the ensemble mean as the forecast state estimate $\mathbf{x}^{\mathrm{f}}$ and a covariance matrix $(\tilde{\mathbf{P}}_k,\ \check{\mathbf{P}}_k)$ computed from the ensemble statistics. While the SEIK filter updates the single $\mathbf{x}_k^{\mathrm{f}}$ and the matrix $\mathbf{U}$, the EnKF updates each ensemble member using for each update an observation vector from an ensemble of observations which needs to be generated. $\tilde{\mathbf{P}}_k^{\mathrm{f}}$ is updated implicitly by this ensemble analysis.

The requirement for an observation ensemble points to the drawback of the EnKF that for finite ensembles the observation ensemble will introduce additional sampling errors in the analysed state ensemble (see Whitaker and Hamill, 2002). This is particularly pronounced if the rank of $\mathbf{R}_k$ is larger than the ensemble size, which occurs if the dimension of the observation vector is large compared to the ensemble size and if $\mathbf{R}_k$ is diagonal.

For linear dynamic and measurement operators the predicted error subspace in the SEEK and SEIK algorithms will be identical if the same rank r is used and model errors are treated in the same way. Because also the analysis steps are equivalent, both filters will yield identical results for linear dynamics. The filter results of the EnKF will differ from those of the SEEK/SEIK filters even for linear dynamics and $N = r + 1$. This is due to the introduction of sampling noise by the Monte Carlo ensembles.

3.5. Re-initialization

Because the EnKF updates the whole ensemble of model states in the analysis step, the algorithm can proceed directly to the next ensemble forecast without the need for a re-initialization algorithm. In contrast to this, in the SEIK filter a new state ensemble representing $\check{\mathbf{P}}_k^{\mathrm{a}}$ and $\mathbf{x}_k^{\mathrm{a}}$ has to be generated, which can be done by a transformation of the forecast ensemble. In the SEEK filter, the forecast modes of the covariance matrix can be used directly in the next forecast step. In general, these are no longer the basis vectors of the error subspace, because they are not orthonormal. A re-orthonormalization of the modes is recommended and can be performed occasionally to stabilize the mode forecast. The choice of whether an algorithm with or without re-initialization is used has no particular implications for the performance of the filter algorithm.

4. Computational cost

In most realistic filtering applications, the major amount of computing time is spent for the model evolution. This time is proportional to the size of the ensemble to be evolved. It is equal for all three algorithms if $r + 1 = N$ where r is the rank of the approximated covariance matrix in the SEEK and SEIK filters and N is the ensemble size in the EnKF. For efficient data assimilation, it is thus of highest interest to find that algorithm which yields the best filtering performance, in terms of estimation error reduction, with the smallest ensemble size. The forecast step consists of N independent model evaluations. This is also true for the SEEK filter if a finite difference approximation of the linearized model is used. Thus, the forecast step of the filters can easily be parallelized for all three filters. The computation time spent in the analysis and re-initialization steps can also be non-negligible, especially if observations are frequently available. The three filter algorithms can show significant differences in these steps. Below we assume $n \gg m > N$, as will occur typically with large-scale models. Also, m can be significantly larger than N, e.g. if data from satellite altimetry are used. Under this assumption, operations on arrays involving the dimension n are most expensive, followed by operations on arrays involving the dimension m. Table 1 shows the scaling of the computational cost for the three filter algorithms. Because the interest is only in the scaling, the numbers neglect the difference between r and N. N is used if some operation is proportional to the ensemble size or the rank of the covariance matrix.

Without the explicit treatment of the model error covariance matrix $\mathbf{Q}$, the SEEK filter is the least costly algorithm. All

Table 1. Computational complexity of the filter algorithms. The scaling numbers only show the dependence on the three dimensions. The first column shows the number of the corresponding equation. The difference between the ensemble size N and the rank r is neglected. Thus, the complexity is given in terms of N also for the SEEK filter. h denotes the cost of the observation operator

Equation	$\mathcal{O}$(operations)
SEEK analysis	
19	$m^2N + mN^2 + m + Nh$
20/21	$nN + n + mN + m + h + N^3 + N^2$
18	$n^2N + nN^2 + N^3$
SEEK re-initialization	
22	$nN^2 + N^3$
23	$nN^2 + nN + N^3 + N^2$
SEIK analysis	
51	$m^2N + mN^2 + mN + N^2 + Nh$
52/53	$nN + n + mN + m + h + N^3 + N^2 + N$
50	$n^2N + nN^2 + N^3$
SEIK re-initialization	
55	$nN^2 + nN + N^3 + N^2 + N$
EnKF analysis	
39	$m^2N + mN + Nh$
37	$m^3 + m^2N + mN$
36	$nN^2 + nN + mN^2$

operations which depend on the state dimension n scale linear with n. The matrix of weights for the state update is computed in the error subspace. Thus, the complexity of the related operations depends on $\mathcal{O}(N)$. Most costly will be here the inversion of $\mathbf{U}^{-1}$ which scales with $\mathcal{O}(N^3)$. The product $\mathbf{R}^{-1}\mathbf{HV}$, which is required to update $\mathbf{U}^{-1}$ according to eq. (19), is the only operation which can scale non-linearly in m. In the worst case, the cost will scale with $\mathcal{O}(m^2N)$, which will only occur if the matrix $\mathbf{R}$ is non-diagonal (i.e. if different measurements are correlated). If the measurements are independent, $\mathbf{R}$ is diagonal. Hence, the product will scale with $\mathcal{O}(mN)$. Because the product can be implemented with $\mathbf{R}^{-1}$ as an operator, it can always be computed in an optimal way depending on the structure of $\mathbf{R}^{-1}$. The re-initialization of the SEEK filter requires extensive operations on the matrix $\mathbf{V}$ which holds the modes of the covariance matrix. The cost of the computation of the product $\mathbf{V}^{\mathrm{T}}\mathbf{V}$ in eq. (22) and the initialization of the new orthonormal modes in eq. (23) scales proportional to $\mathcal{O}(nN^2)$. Because it is only occasionally required to compute the re-initialization, this operation will not affect the overall numerical efficiency of the SEEK filter.

The computational cost of the analysis step of the SEIK filter is very similar to that of the SEEK algorithm. The computation of the ensemble mean state in eq. (52) will produce an overhead in comparison to the SEEK algorithm. Its cost scales with $\mathcal{O}(nN + n)$. Other additional operations in comparison to the SEEK filter are applications of the matrix $\mathbf{T}$. Because the multiplication with $\mathbf{T}$ can be implemented as operators, these products require $2mN + m + 2N$ floating point operations. Finally, the initialization of matrix $\mathbf{G}$ (eq. 48) is required. This will require N^2 operations, because it can be performed directly without computing the product $\mathbf{T}^{\mathrm{T}}\mathbf{T}$. The re-initialization step of the SEIK filter (eq. 55) is significantly less costly compared to that of the SEEK algorithm, because no diagonalization of $\check{\mathbf{P}}^{\mathrm{a}}$ is performed. The weight matrix for the ensemble transformation is computed entirely in the error subspace. Most costly will be the Cholesky decomposition of $\mathbf{U}^{-1}$ and the inversion of $\mathbf{C}$. The cost of these operations scales with $\mathcal{O}(N^3)$ and can have a significant cost for rather large ensembles. The cost of the initialization of the matrix $\mathbf{\Omega}$ can be neglected. For each re-initialization, the same matrix $\mathbf{\Omega}$ can be used in eq. (55). Thus, $\mathbf{\Omega}$ can be stored. Operations on matrices involving the dimension n occur finally in the computation of the new ensemble states. The cost of these operations scales linearly with $\mathcal{O}(n)$.

The computational cost of the SEEK and SEIK algorithms will increase strongly if the model error covariance matrix $\mathbf{Q}$ is taken into account. This is due to the amount of operations required for the projection of $\mathbf{Q}$ onto the error subspace (eqs. 18 and 50). If $\mathbf{Q}$ has full rank, this projection requires $\mathcal{O}(n^2N + nN^2 + N^3)$ operations. Due to the part scaling with $\mathcal{O}(n^2N)$, it is impractical to apply this projection. The computational cost is significantly smaller if $\mathbf{Q}$ has a low rank of $k \ll n$ and is stored in square-root form $\mathbf{Q} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ with $\mathbf{A}$ being some n by k matrix. In this case, the projection requires $\mathcal{O}(nN^2 + nkN + N^2k + N^3)$ operations. Hence, the cost of the projection is comparable to the cost of the re-initialization steps of the SEEK and SEIK filters if $k \approx N$. If the model errors are only poorly known, it would probably be too costly to apply this projection. Alternatively, the forgetting factor could be used, which requires N^2 operations. In the SEIK filter it is also possible to apply model errors as a stochastic forcing during the forecast step. If this forcing is applied at every time-step to each element of all ensemble states, its cost scales with $\mathcal{O}(nN)$ for each time-step.

The EnKF appears appealing as it does not require an explicit re-initialization of the ensemble because the ensemble states are updated during the analysis step of the filter. The cost of the ensemble update according to eq. (36) scales with $\mathcal{O}(nN^2 + nN + mN^2)$. Hence, the cost of this operation is larger than the re-initializations in the SEEK and SEIK filters. In fact, the computation of new modes or ensemble states amounts for all three filters to the calculation of weighted averages of the prior ensembles or modes. Because the EnKF uses the representer formulation, which operates in the observation space rather than the error subspace, it requires more operations than the SEIK and SEIK filters if $m > N$. In addition, all other operations in the analysis algorithm are dependent on m. The cost of the solver step for the representer amplitudes in eq. (37) scales with $\mathcal{O}(m^3 + m^2N)$. Thus, this operation will be very costly if large observational data sets are assimilated. Also costly will be the computation of the matrix $\mathbf{H}\tilde{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}$, which scales with $\mathcal{O}(m^2N)$. Another costly operation can be the generation of an ensemble of observations. This can be implemented using a transformation of independent random numbers (see, for example, Fukumori, 2002). The transformation requires the eigenvalue decomposition of the matrix $\mathbf{R}$, which scales with $\mathcal{O}(m^3)$. The cost of the subsequent initialization of the ensemble vectors is proportional to $\mathcal{O}(m^2N)$. Hence, the generation of the observation ensemble is of comparable complexity to the solver step for the representer amplitudes. Computationally, it is most efficient to perform the ensemble update in matrix form. That is, the residuals $\{\mathbf{d}^{(l)} = \mathbf{y}_k^{\mathrm{o}(l)} - H_k[\mathbf{x}_k^{\mathrm{f}(l)}]\}$ are stored in the columns of a matrix $\mathbf{D}$, then all amplitudes $\{\mathbf{b}^{(l)}\}$ are computed at once according to eq. (37). Finally, all ensemble states are updated at once according to eq. (36). This procedure requires more computer memory, but it can be better optimized by compilers than a serial version where for each single residual the amplitudes and finally a single updated ensemble state are computed. Overall, the EnKF analysis requires more operations than the SEEK and SEIK filters if the observation dimension m is larger than the ensemble size N. This is caused by the representer formulation used in the EnKF. Due to this, the EnKF operates on the observation space rather than the error subspace, which is directly taken into account by the SEEK and SEIK filters.

To optimize the performance of the EnKF and its ability to handle very large observational data sets, Houtekamer and Mitchell (2001) discussed the use of an iterated analysis update. In this case, the observations are subdivided into batches of independent

observations. Each iteration uses one batch of observations to update the ensemble states. This reduces the effective dimension of the observation vector. It is also common practice to follow this approach and perform a set of local analyses (see, for example, Evensen, 2003). Because the EnKF contains several operations which scale with $\mathcal{O}(m^3)$ or $\mathcal{O}(m^2)$, this technique diminishes the computational cost of the algorithm. In addition, the memory requirements are reduced. The iterative analysis update can also be applied with the SEEK and SEIK filters. As has been discussed above, only the cost of the product $\mathbf{R}^{-1}\mathbf{HV}$ can scale with $\mathcal{O}(m^2)$ if $\mathbf{R}^{-1}$ is not diagonal while all other operations in the analysis step of the SEEK and SEIK filters scale at most linearly with m. Hence, no particular gain in the computational cost can be expected for SEEK and SEIK when using batches of observations. The memory requirements are, however, reduced also for these filters.

Recently, Evensen (2003) proposed an efficient analysis scheme for the EnKF which is based on a factorization of the term in parentheses in the Kalman gain eq. (32). This relies on an ensemble representation of the observation error covariance matrix $\mathbf{R}$ and requires that the state and observation ensembles are independent. As has been discussed in the remarks on the EnKF, this scheme can lead to a further degradation of the filter quality. In this scheme the computational cost is reduced to be linear in m. An exception from this is the generation of the observation ensemble which remains unchanged. Thus, apart from the generation of the observation ensemble, the computational cost of the newly proposed EnKF analysis scheme will be similar to that of the SEEK and SEIK filters.

5. Set-up of assimilation experiments

To assess the filter behaviour of the different filter algorithms, identical twin experiments with a simple model are performed. Non-linear shallow-water equations are used, which are discretized in potential enstrophy conserving form (Sadourny, 1975). We extended these to include the Coriolis force. It is assumed that the model is exact, and thus no model error is simulated. This choice is motivated by the fact that it is not possible to apply the model error in exactly the same way in all three filters. However, while relevant for the practical application of the filters, the model error is not important for an intercomparison of the algorithms, because it is consistently missing in all three filters applied here. The model domain is chosen as a box measuring 950 km per side with a flat bottom at 1000 m depth on an f-plane with a Coriolis parameter of 10^{-4} s^{-1}. Periodic boundary conditions are applied in zonal and meridional directions. The experiments were performed with 30×30 grid points and a time-step of 100 s using a leapfrog scheme with a Robert–Asselin filter. The model state vector $\mathbf{x}$ consists of the surface elevation $\mathbf{h}$ and the horizontal velocity components $\mathbf{u}$ and $\mathbf{v}$ at the grid points. The state dimension amounts to $n = 2700$. This number is sufficiently large to obtain meaningful filter results also for the low-rank algorithms, but it is still small enough to allow for a direct study of the filter-represented covariance matrices.

For the twin experiments, the 'true' state trajectory of the system is generated by a model integration initialized with the state shown in Fig. 1a. It is in geostrophic balance and has a shape that ensures non-linear evolution with the shallow-water equations. Synthetic observations of the surface elevation at each grid point are generated by adding normally distributed random numbers of constant variance of 10^{-4} m^2 to the true surface elevation. Using only the surface elevation as observations, the dimension of the observation vector is $m = 900$. The synthetic observations are quite accurate. However, this is useful because the dependence of filter performance on ensemble size can be better assessed for large ensembles with accurate observations.

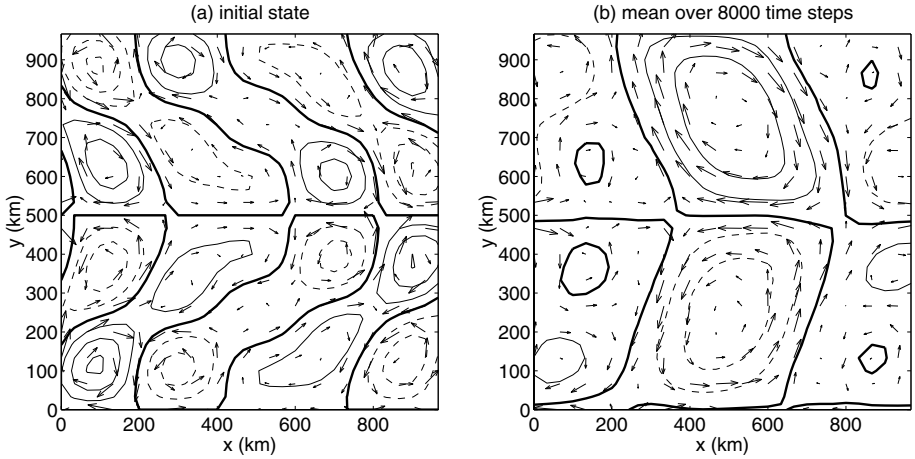

Fig. 1. Surface elevation and velocity field (subsampled for displaying) for (a) initial state used for generating the 'true' model trajectory and the synthetic observations and (b) mean state over 8000 time-steps (8×10^5 s) using each tenth step. The contour interval is 1 m. The thick line denotes the mean sea level, and dashed lines negative deviations.

Two types of experiments are performed. For the first, referred to as experiment 'A', the initialization of the model state estimate $\mathbf{x}_0^{\mathrm{a}}$ and the corresponding covariance matrix $\mathbf{P}_0^{\mathrm{a}}$ is performed for all three filter algorithms by applying the EOF procedure described by Pham et al. (1998a), which uses a sequence of model states. The initial state estimate $\mathbf{x}_0^{\mathrm{a}}$, shown in Fig. 1b, is chosen as the mean state of the true model simulation over 8000 time-steps using the state vector at each tenth time-step. The covariance matrix $\mathbf{P}_0^{\mathrm{a}}$ is computed as the variation of the true model trajectory about this mean. This matrix does not reflect the estimated error of the initial state but the estimated mean temporal variability of the model state. However, the procedure yields a consistent and simple way to obtain variance estimates together with estimates of the covariances.

This mean and covariance matrix serves as a baseline. However, it soon turned out that all algorithms can improve this 'state of large ignorance'. A much more enlightening setting would be to use a model state and covariance matrix that are already quite accurate and difficult to improve. To this end, the initialization of the second type of experiment, referred to as experiment 'B', is conducted with the estimated state and covariance matrix after the second analysis update from an assimilation experiment of type A with the EnKF using a very large ensemble of $N = 5000$ members. This is a very accurate state estimate whose rms deviation from the true state is two orders smaller than the initial estimate of type A. The structure of this state is thus very similar to that of the true initial state displayed in Fig. 1a. In addition, the covariance matrix of type B is an estimated error covariance matrix of the state estimate. It has a strongly different structure compared to the covariance matrix of type A. This is obvious from the eigenvalue spectrum, displayed in Fig. 2. For type A, the covariance matrix has a very large condition number and the 10 largest eigenmodes already explain 99% of the variance. In contrast to this, 371 eigenmodes are required to explain 90% of the variance for type B.

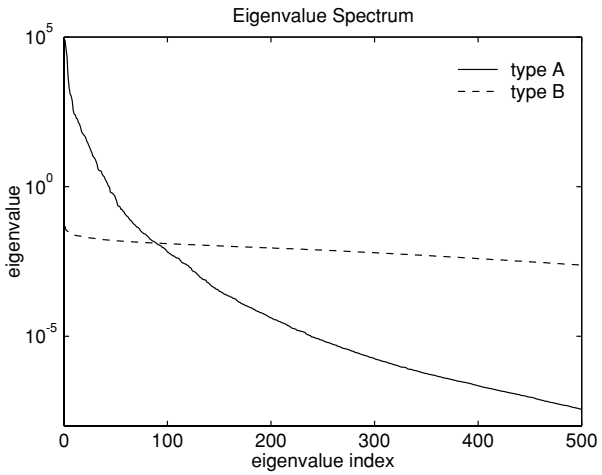

Fig. 2. Eigenvalues for the covariance matrices for experiments of types A and B up to eigenvalue index 500. While in type A the spectrum is highly red and 10 modes already explain 99% of the variance, in the more realistic setting B 371 modes are required to explain 90% of the variance.

To initialize the SEEK and SEIK filters, decomposed low-rank approximations $\hat{\mathbf{P}}_0^{\mathrm{a}} = \mathbf{V}_0\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}}$ of the covariance matrix $\mathbf{P}_0^{\mathrm{a}}$ are required. These are computed by incomplete eigenvalue decompositions of $\mathbf{P}_0^{\mathrm{a}}$ retaining only the r largest eigenmodes. The N ensemble states required for the EnKF algorithm have been generated from the state estimate $\mathbf{x}_0^{\mathrm{a}}$ and the covariance matrix $\mathbf{P}_0^{\mathrm{a}}$ by a transformation of independent random numbers. For this, the eigenvalue decomposition of $\mathbf{P}_0^{\mathrm{a}}$ is computed, yielding $\mathbf{P}_0^{\mathrm{a}} = \mathbf{VUV}^{\mathrm{T}}$. The eigenvectors are scaled by the square root of the corresponding eigenvalue as $\mathbf{L} = \mathbf{VU}^{1/2}$. For each ensemble state $\{\mathbf{x}_0^{\mathrm{a}(l)}, l = 1, \ldots, N\}$, each scaled eigenvector $\mathbf{L}_{(i)}$ is multiplied by a random number $b_i^{(l)}$ from a normal distribution of zero mean and unit variance and added to the state estimate $\mathbf{x}_0^{\mathrm{a}}$:

$$\mathbf{x}_0^{\mathrm{a}(l)} = \mathbf{x}_0^{\mathrm{a}} + \sum_{i=1}^{q} b_i^{(l)}\,\mathbf{L}_{(i)}; \quad l = 1, \ldots, N. \tag{56}$$

Because the prescribed covariance matrix has in type A a maximum rank of 799, we use here only $q = 799$ eigenmodes for the ensemble generation.

The assimilation experiments are performed over an interval of 8000 time-steps for type A and 7600 time-steps for type B with an analysis step each 200 time-steps. For a particular ensemble size N, filter configurations are used in which the rank in the SEEK and SEIK filters is $r = N - 1$. In this case, the number of model evaluations is equal for all three algorithms and the filter performances can be directly related to computing time. Below, the expression 'ensemble size' is used to denote the number of different model states to be evolved.

The filter algorithms are applied in their 'pure' form as discussed in Section 2. In particular, no techniques to stabilize the filters, e.g. by localization of the analysis (see, for example, Houtekamer and Mitchell, 2001) or by forgetting factors (see, for example, Pham, 2001; Whitaker and Hamill, 2002), are used. While these techniques are expected to prevent the filters from filter divergence, they would amount to the individual tuning of each filter for maximum filter performance. As a consequence, the basic abilities of the filters and the dependence of the filter performance on the sampling quality of the covariance matrices, to be discussed below, would be less visible. Accordingly, we apply the filter algorithms without these techniques, bearing in mind that there are techniques which are able to improve the filter performance for all three algorithms.

6. Comparison of filter performances

To evaluate the filter performance of the three algorithms, the estimation error, given by the rms deviation of the assimilated state from the true state and denoted as E_1, is considered separately for the three state fields, **h**, **u** and **v**. For the EnKF, the leftmost panels of Fig. 3 show estimation errors for experiments of type A with the three ensemble sizes, $r = 30$, 100 and 500. In

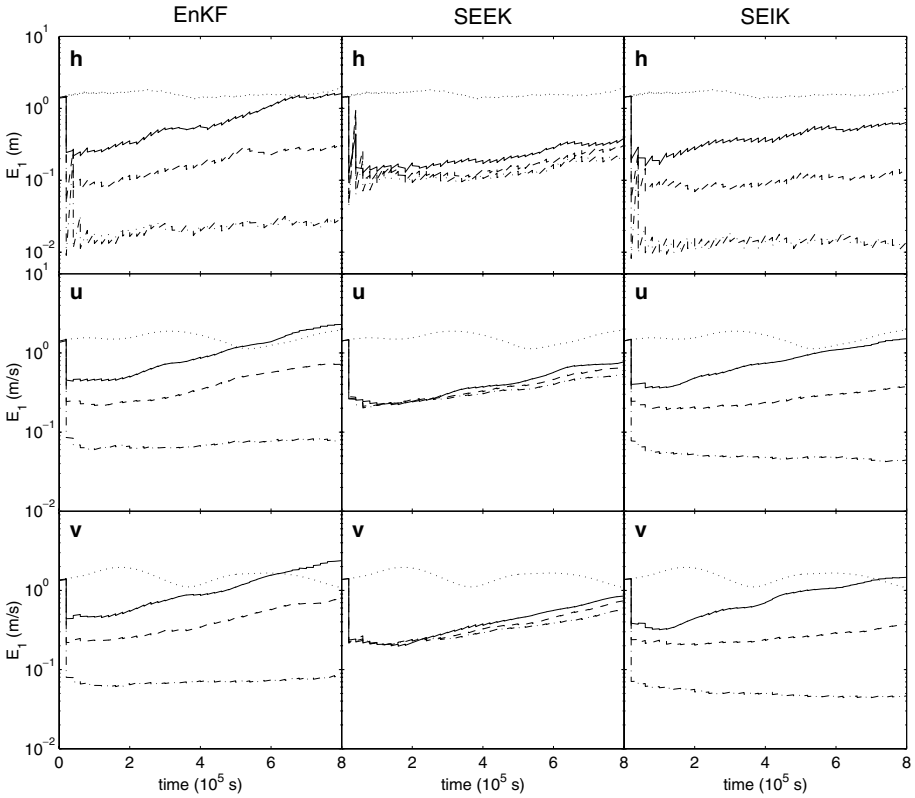

Fig. 3. The rms estimation errors for experiments of type A with poor initial guess. Shown is the time development of the error of the assimilated state for three ensemble sizes ($N = 30$, solid line; $N = 100$, dashed line; $N = 500$, dash-dotted line) and for a model simulation without assimilation (dotted line). Each column shows the results for one particular filter algorithm. The three rows show the different model fields. For the EnKF and SEIK algorithms, divergence can be avoided by large enough ensemble size.

addition, E_1 for an experiment without assimilation is displayed which shows only small variations over time.

The temporal development of E_1 with assimilation is characterized by a large reduction at the first analysis step. This is due to an initially large error in the state estimate in connection with quite accurate observations. Subsequent analysis steps have significantly smaller influence. The filter performs better with increasing ensemble size where E_1 is strongly diminished. For small ensembles, such as $N = 30$, E_1 increases with assimilation time, showing that the filter is unstable. As is visible in Fig. 3, the state estimate of the assimilation after 8000 timesteps with 40 analysis cycles is even worse than without assimilation. For larger ensembles, the assimilated state remains close to the true state. Because only observations of the height field **h** are assimilated, the velocity components are merely updated via cross-covariances between the height field and the velocity components. The representation of these covariances is generally worse than that of the height field variances and covariances, as discussed in the following section. Due to this, the values of E_1 for the velocity components **u**, **v** are larger than for the height field.

The centred and rightmost panels of Fig. 3 show that for the SEEK and SEIK filters the general behaviour of E_1 as a function of assimilation time is analogous to that of the EnKF. In order to compare the performance of all three filter algorithms in a compact way, we define the normalized time-integrated state estimation error by

$$E_2 := \frac{1}{3} \sum_{f=\mathbf{h},\mathbf{u},\mathbf{v}} \left[\sum_{k=k_{\min}}^{40} \frac{E_1^{\text{assim}}(f, t_k)}{E_1^{\text{free}}(f, t_k)} \right], \tag{57}$$

where $E_1^{\text{assim}}(f, t_k)$ denotes the value of E_1 at time t_k for the state field $f \in \{\mathbf{h}, \mathbf{u}, \mathbf{v}\}$ from an assimilation experiment and $E_1^{\text{free}}(f, t_k)$ is the corresponding value from an experiment without assimilation. Dependent on the type of experiment, it is $k_{\min} = 1$ for type A and $k_{\min} = 3$ for type B. This excludes for type A the initial state estimate because it would dominate the value of E_2 due to the large error decrease at the first analysis step. E_2 provides a rms measure of the decrease in estimation error due to data assimilation, which respects a possible different scaling of the state fields.

In Fig. 4 E_2 for the three filter algorithms is shown as a function of ensemble size N for experiments of type A. For the EnKF and SEIK algorithms, mean results and standard deviations over 20 experiments using different random numbers in the initialization step are shown. There are significant variations of the filter performance depending on the used set of random numbers

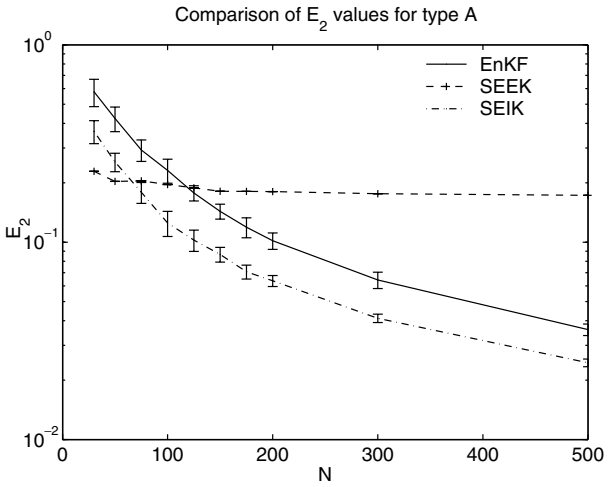

Fig. 4. Normalized time-integrated estimation errors for the three filter algorithms as a function of the ensemble size N ($N = r + 1$ for SEEK and SEIK) for experiments of type A with poor initial guess. Error bars describe the spread due to the use of random numbers.

because the computer generated random numbers in fact do not represent the prescribed statistics exactly and do determine in which directions of the state space the ensemble vectors point. For small N, the latter will likely lead to different qualities of the forecast ensemble. The SEEK filter is deterministic in its initialization, and hence only the result of a single simulation per ensemble size is shown. As the observations are also generated using computer generated random numbers, they will also determine the filter performance. This is of no concern here, because the observation error is quite small here and all three algorithms use the same observations.

Overall, E_2 converges in the same manner for the EnKF and SEIK filters. A different convergence for the SEIK filter which should be expected due to the second-order exact sampling is not visible. This is caused by the eigenvalue spectrum of the covariance matrix $\mathbf{P}_0^{\mathrm{a}}$, which shows that the number of significant eigenvalues is extremely small. For the EnKF and the SEIK filter the convergence in the interval $100 < N < 500$ can be approximated by $E_2 \propto N^{-x}$ with $x \approx 1.2$ for the EnKF and $x \approx 1.0$ for the SEIK filter. In general, the mean E_2 values for the EnKF are, depending on the ensemble size, between 1.5 and 1.85 times larger than those for the SEIK filter. In order to achieve the same filter performance, the ensemble for the EnKF needs to be between about 1.5 and 1.8 times larger than for the SEIK filter. These numbers are of course specific for the configuration of these experiments. However, variations of the assimilation interval and strong increase of the observation errors preserved the relative performances of the algorithms. The behaviour for the SEEK filter deviates significantly from that of the EnKF and SEIK algorithms. For $N < 70$ the SEEK filter shows the best filter performance of the three algorithms. However, with further increasing N, E_2 stagnates at a rather large value.

For experiments of type B, the estimation errors E_1 over time are displayed in Fig. 5. The initial state approximates the true state quite well, but over time the rms deviation for the evolution without assimilation increases by about two orders of magnitude. Thus, the conditions for this experiment are quite different from those of type A in which the initial state estimate was strongly deviating from the true state and remained over simulation time at an almost constant rms deviation from the true state. In experiments of type B, the assimilation of height field observations keeps the estimates of all state fields much nearer to the true state compared to the simulation without assimilation. Similar to type A, the estimation error of the velocity components is higher than for the sea level.

Comparing the filter behaviours in experiments of type B with those of type A, it is obvious that the filter results depend on the initialization. This is due to the fact that in the experiments no model error is present which could extend the initial error subspace in a stochastic way. Thus, the initially chosen error subspace determines the error statistics at later times and different error subspaces will result in different filter results. While it is known for the KF with a linear autonomous model operator that the estimated error statistics become independent from the initialization after a sufficient period, this cannot necessarily be expected for error subspace algorithms. Even, when model error is simulated, the length of the initial period will depend on the relative size of the model error to the error estimates in the error subspace. Accordingly, a good initialization method, such as that applied in the SEIK filter, can be beneficial if the error estimates are not dominated by the model error.

E_2 as a function of ensemble size is displayed in Fig. 6 for the experiments of type B. Here, mean results and standard deviations over 20 experiments with different random numbers in the initialization are only shown for the EnKF. The dependence of the SEIK filter on the random numbers used in the initialization is negligible for this type of experiment (not shown). The performance of the SEEK and SEIK filters is almost indistinguishable, with a relative difference of the values of E_2 below 6×10^{-3}. The values of E_2 are smaller for type B than for type A, which is due to the normalization by E_1^{free} when computing E_2. As in type A, the value of E_2 converges similarly for the EnKF and the SEIK filter. However, for small ensembles ($N \leq 75$) the SEIK filter converges faster than the EnKF. Again, the dependence of E_2 on N can be approximated in the interval $100 < N < 500$ to be $E_2 \propto N^{-x}$. We find $x \approx 0.42$ for the EnKF and $x \approx 0.44$ for the SEIK filter. Thus, the convergence rate with ensemble size is much smaller for type B than for type A. It requires more effort to improve a good solution than a poor solution estimate. To obtain the same filter performance, the ensemble in the EnKF would need to be between about 1.6 and 2.2 times larger than for the SEEK and SEIK filters. This result corresponds to that reported by Heemink et al. (2001). There the RRSQRT filter, which is similar to the SEEK algorithm, yielded

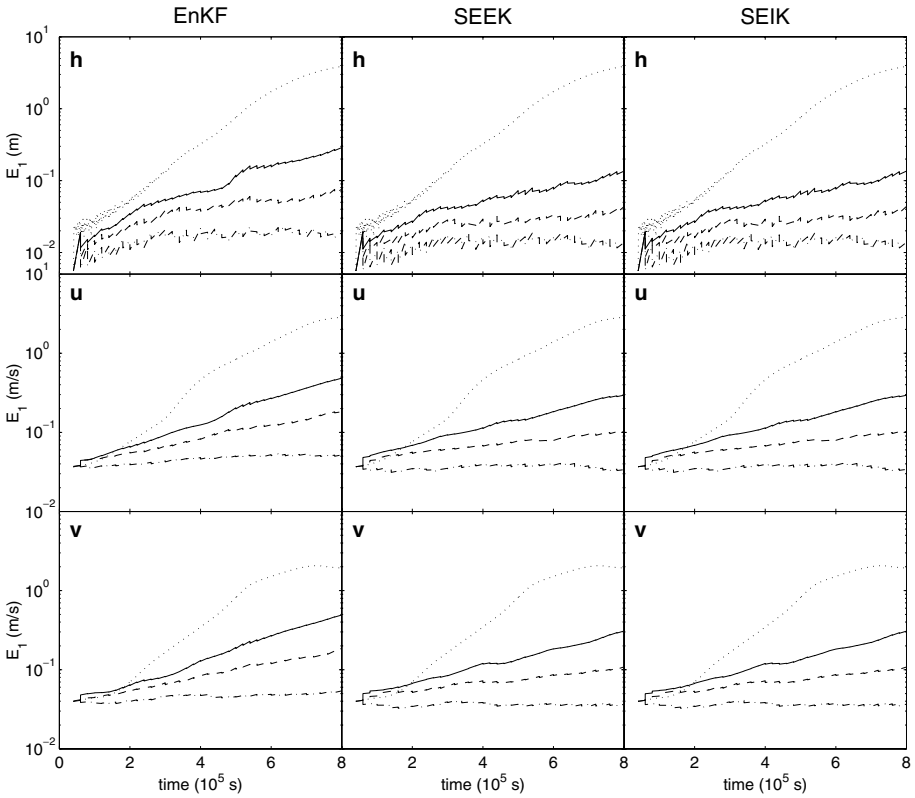

Fig. 5. The rms estimation errors for experiments of type B with excellent initial guess. Shown is the time development of the error of the assimilated state for three ensemble sizes ($N = 30$, solid line; $N = 100$, dashed line; $N = 500$, dash-dotted line) and for a model simulation without assimilation (dotted line). Each column shows the results for one particular filter algorithm. The three rows show the different model fields. As in case A, filter performance and divergence is a function of ensemble size.

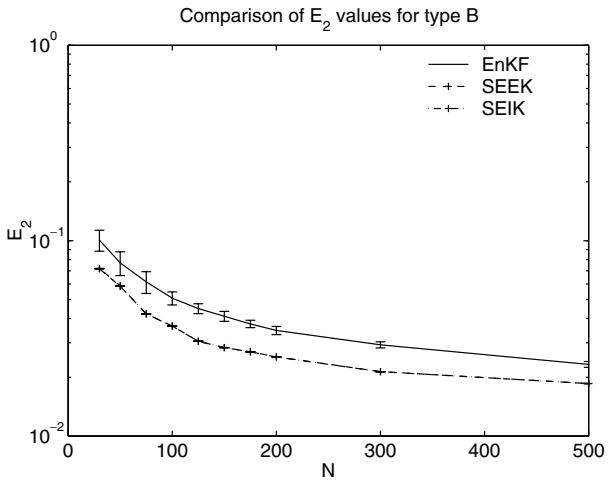

Fig. 6. Normalized time-integrated estimation errors for the three filter algorithms as a function of the ensemble size N for experiments of type B with excellent initial guess. The lines of the SEEK and SEIK filters lie on top of each other. Error bars describe the spread due to the use of random numbers.

comparable estimation errors to the EnKF for about half the number of model evaluations.

It is possible to interchange the methods of Monte Carlo sampling and second-order exact sampling between the EnKF and SEIK algorithms, as discussed in Section 2. Dependent on the ensemble size, this yields in the experiments of both types for the EnKF a 5–10% better filter performance, while the performance of the SEIK filter degrades by about the same amount. After interchanging the initializations, the SEIK filter still performs better than the EnKF. This is caused by the introduction of noise into the ensemble by the observation ensemble required in the analysis scheme of the EnKF, as discussed below.

7. Quality of error statistics

To gain insight into the reasons for the different filter performances of the three algorithms, we extend the analysis of the model state estimate after assimilation. We further examine the sampling quality of the represented state covariance matrices in the following. At first, some additional analysis quantities are defined. Based on these quantities, we then discuss how the different variants of forecasting and different choices of ensembles can lead to estimates of the covariance matrix, and hence the error subspace, of strongly different quality.

7.1. Definition of analysis quantities

To define analysis quantities to assess the sampling quality, let us reconsider the filter algorithms. The SEEK filter evolves the state estimate with the non-linear dynamic model and the eigenmodes of the low-rank approximated state covariance matrix with the linearized dynamic model or a finite difference approximation of it. The EnKF and SEIK algorithms both evolve an ensemble of model states with the non-linear dynamic model. The capability of the forecast phase to provide a realistic representation of the error subspace is reflected by the sampling quality of the state covariance matrix $\mathbf{P}$.

To discuss the analysis step, we consider the covariance matrix to consist of submatrices as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{hh} & \mathbf{P}_{hu} & \mathbf{P}_{hv} \\ \mathbf{P}_{uh} & \mathbf{P}_{uu} & \mathbf{P}_{uv} \\ \mathbf{P}_{vh} & \mathbf{P}_{vu} & \mathbf{P}_{vv} \end{pmatrix}. \tag{58}$$

Here the submatrices $\{\mathbf{P}_{ij} = \mathbf{P}_{ji}^{\mathrm{T}}\}$ are $m \times m$ matrices ($m = n/3$) with $\mathbf{P}_{hh}$, $\mathbf{P}_{uu}$, and $\mathbf{P}_{vv}$ respectively containing the covariances of the height field and the two velocity components. The off-diagonal submatrices $\{\mathbf{P}_{ij}, i \neq j\}$ contain the cross-covariances between different state fields. The measurement operator projects a state vector onto its height field part, and thus $\mathbf{H} = (\mathbf{I}_{m\times m}, \mathbf{0}_{m\times 2m})$ where $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ is the matrix containing only zeros. In the experiments, all observations were assumed to be uncorrelated with variances of constant value σ_h^2. Thus, the observation error covariance matrix is $\mathbf{R} = \sigma_h^2 \mathbf{I}_{m\times m}$.

With this specification, the analysis equation for the state in the SEEK (eq. 20) and SEIK (eq. 52) filters simplifies to

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \sigma_h^{-2} \begin{pmatrix} \mathbf{P}_{hh}^{\mathrm{a}} \\ \mathbf{P}_{uh}^{\mathrm{a}} \\ \mathbf{P}_{vh}^{\mathrm{a}} \end{pmatrix} \mathbf{d} \tag{59}$$

with observation-state residual, also called innovation, $\mathbf{d} = \mathbf{y}^{\mathrm{o}} - \mathbf{h}^{\mathrm{f}}$ where $\mathbf{h}^{\mathrm{f}}$ is the estimated forecast height field. For the EnKF, the analysis equation for the ensemble states (eq. 31) is also valid for the ensemble mean (Evensen, 1994). Here it simplifies to

$$\mathbf{x}^{\mathrm{a}} = \overline{\mathbf{x}^{\mathrm{f}}} + \begin{pmatrix} \mathbf{P}_{hh}^{\mathrm{f}} \\ \mathbf{P}_{uh}^{\mathrm{f}} \\ \mathbf{P}_{vh}^{\mathrm{f}} \end{pmatrix} \left[\mathbf{P}_{hh}^{\mathrm{f}} + \sigma_h^2 \mathbf{I}\right]^{-1} \mathbf{d} =: \overline{\mathbf{x}^{\mathrm{f}}} + \mathbf{A}\mathbf{d}. \tag{60}$$

According to eqs. (59) and (60) only the covariances $\mathbf{P}_{hh}$ in the height field and the cross-covariances $\mathbf{P}_{uh}$ and $\mathbf{P}_{vh}$ between the height field and the velocity components participate in the analysis update of the state estimate. The other submatrices are updated as well during the analysis update of the covariance matrix and all parts of $\mathbf{P}$ determine the forecast quality.

To compare the three filter algorithms despite their different analysis equations, we define update matrices $\mathbf{B}$. For the SEEK and SEIK filters we define the elements $\{\mathbf{B}_{(i,j)}, 1 \leq i \leq n, 1 \leq j \leq m\}$ by

$$\mathbf{B}_{(i,j)}^{\mathrm{a}} := \sigma_h^{-2} \mathbf{P}_{(i,j)}^{\mathrm{a}} \mathbf{d}_{(j)}. \tag{61}$$

For the EnKF the definition is analogously

$$\mathbf{B}_{(i,j)}^{\mathrm{f}} := \mathbf{A}_{(i,j)} \mathbf{d}_{(j)}. \tag{62}$$

Here, $\mathbf{d}_{(j)}$ denotes the element j of the vector $\mathbf{d}$. The update matrices $\mathbf{B}^{\mathrm{a}}$, $\mathbf{B}^{\mathrm{f}}$ correspond to the matrix–vector products in eqs. (59) and (60) without performing the summation. For the SEEK and SEIK filters, this amounts to a scaling of the covariances by the inverse variance σ_h^{-2} and the elements of the residual vector. Thus, the update matrices take into account not only the different sampling qualities of the state covariance matrix but also the different residuals $\mathbf{d}$. Accordingly, an estimate of the analysis quality for the single state fields will be provided by the sampling quality of the submatrices, $\mathbf{B}_{hh}$, $\mathbf{B}_{uh}$ and $\mathbf{B}_{vh}$.

To quantify the sampling quality, we compare the computed update matrices with those of an assimilation with the EnKF with an ensemble size $N = 5000$, referred to as the 'ideal' update matrix $\mathbf{B}^{\mathrm{ideal}}$. For the comparison we compute correlation coefficients $\rho_{\mathbf{B}}$ between the sampled and ideal update submatrices and regression coefficients $\beta_{\mathbf{B}}$ from the ideal to the sampled update submatrices. Over time, the state estimates of the different filter algorithms deviate increasingly. For this reason, we focus on the very first analysis step.

7.2. Influence of ensemble size in type A

In Table 2 experiments of type A are examined for assimilation with an ensemble size $N = 30$. The correlation and regression coefficients, $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$, are displayed for the height field $\mathbf{h}$ and the zonal velocity component $\mathbf{u}$. The coefficients for the meridional velocity component $\mathbf{v}$ are similar to those for $\mathbf{u}$ and thus are not shown. In addition, the relative estimation error

$$E_3(f) = \frac{E_1^{\mathrm{assim}}(f, t_1)}{E_1^{\mathrm{free}}(f, t_1)} \tag{63}$$

Table 2. Examination of the first analysis for experiments of type A with $N = 30$. Shown are relative estimation errors E_3 and the correlation $\rho_{\mathbf{B}}$ and regression $\beta_{\mathbf{B}}$ coefficients between the ideal and sampled update submatrices for the height field $\mathbf{h}$ and the zonal velocity $\mathbf{u}$. In addition, the correlation ρ_σ and regression β_σ coefficients of the variance part for the height field are shown

	Field	E_3	$\rho_{\mathbf{B}}$	$\beta_{\mathbf{B}}$	ρ_σ	β_σ
EnKF		0.168	0.305	0.091	0.961	0.071
SEEK	**h**	0.089	0.325	0.107	0.959	0.086
SEIK		0.135	0.320	0.107	0.959	0.084
EnKF		0.309	0.126	0.015		
SEEK	**u**	0.179	0.188	0.035		
SEIK		0.273	0.130	0.017		

after the first analysis is shown for the fields $f \in (\mathbf{h}, \mathbf{u})$. For comparison, the values of E_3 for the ideal experiment are much smaller with $E_3(\mathbf{h}) = 0.005$ and $E_3(\mathbf{u}) = 0.04$. Thus, the filter performance will increase strongly with growing ensemble size and the improvement will be larger for the height field than for the velocity components. The order of the values of E_3 for the three filters is the same as that of the time-integrated E_2 values, which are shown in Fig. 4 for $N = 30$. The SEEK filter has the smallest value of E_3, followed by the SEIK filter and then the EnKF. The ratio of the time-integrated E_2 for the EnKF to that of the SEIK filter is 1.59. It is larger than the corresponding ratio of E_3 values of 1.24 after the first analysis update. This is caused by the use of an observation ensemble in the analysis of the EnKF, which destabilizes the assimilation process, as examined in more detail below.

The correlation and regression coefficients, $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$, reflect the different filter performances of the first analysis update. Overall, it is visible that there is a significant correlation between the sampled and the ideal submatrices. The small regression coefficients show in addition that the amplitudes are strongly underestimated. The amount of underestimation decreases when observations with larger error are assimilated (not shown). The underestimation is even more pronounced when we consider only the correlation and regression coefficients for the variance part, i.e. the diagonal, of the height field update submatrix. These coefficients are also shown in Table 2, denoted ρ_σ and β_σ. The correlation coefficients ρ_σ are for $N = 30$ already very near to unity, but the regression coefficients β_σ show a very strong underestimation of the variance. In the experiments, the structure of the update submatrix $\mathbf{B}_{hh}$ corresponding to a single grid point, as well as the covariance submatrix $\mathbf{P}_{hh}$, consists of noise of rather low amplitude and a significantly larger peak with a radius of about two grid points around the location of the specified grid point. Accordingly, the variance will dominate the analysis while most of the noise will average out when computing the product $\mathbf{P}_{hh}\mathbf{d}$. The smaller values of $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$ for $\mathbf{h}$ for the EnKF point to the fact that here the analysis is less accurate than for the SEEK and SEIK filters. This is confirmed by the value of E_3, which is larger for the EnKF than for the two other filters. For the difference between the SEEK and SEIK filters this is less obvious. The smaller regression coefficients $\beta_{\mathbf{B}}$ and β_σ for the EnKF include also the 'inbreeding effect', which results from the update of the state ensemble using a gain computed from the same ensemble (Houtekamer and Mitchell, 1998, 1999; van Leeuwen, 1999).

The sampling quality of $\mathbf{B}$ is generally worse for the velocity components than for the height field. This is due to the fact that only $\mathbf{h}$ is observed. Hence, $\mathbf{u}$ and $\mathbf{v}$ are updated via the covariance submatrices $\mathbf{P}_{uh}$ and $\mathbf{P}_{vh}$, which have a structure showing multiple extrema and are more difficult to sample than the variance-dominated $\mathbf{P}_{hh}$ (not shown). For $N = 30$ the values of $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$ are nearest to unity in the case of the SEEK filter. This is consistent with the filter's small value of E_3. In experiments of type A, the SEEK filter is able to sample the submatrices $\mathbf{P}_{uh}$ and $\mathbf{P}_{vh}$ for small ensembles significantly better than the SEIK and EnKF algorithms.

Table 3. Examination of the first analysis for experiments of type A with $N = 200$. The same quantities as in Table 2 are shown

	Field	E_3	$\rho_{\mathbf{B}}$	$\beta_{\mathbf{B}}$	ρ_σ	β_σ
EnKF		0.015	0.756	0.570	0.996	0.477
SEEK	**h**	0.035	0.554	0.277	0.988	0.227
SEIK		0.012	0.756	0.598	0.995	0.503
EnKF		0.103	0.502	0.315		
SEEK	**u**	0.191	0.324	0.121		
SEIK		0.081	0.496	0.332		

For $N = 200$ the sampling quality of the update matrices is examined in Table 3. The estimation errors E_3 after the first analysis are much smaller in comparison to $N = 30$. This decrease is smaller for the velocity components than for the height field due to the worse sampling of cross-correlations between $\mathbf{h}$ and the velocity components, $\mathbf{u}$ and $\mathbf{v}$. The increased regression coefficients $\beta_{\mathbf{B}}$ show that the underestimation of the correlations has diminished. In addition, according to the increased correlation coefficients $\rho_{\mathbf{B}}$ and ρ_σ, covariances as well as variances are sampled much more realistically. The values of the coefficients for the SEIK and EnKF algorithms are now more similar but the SEIK filter still shows the better sampling quality. The estimation error measure E_3 for $N = 200$ is larger for the SEEK filter than for the SEIK and EnKF algorithms. This is consistent with the values of $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$, which are smaller for the SEEK filter than for the two other filters. This inferior sampling quality of the SEEK filter is caused by the direct forecast of the eigenmodes of the state covariance matrix $\mathbf{P}$. The modes with larger index represent gravity waves which do not provide useful information to the error subspace, causing the filter performance to stagnate. For the estimated velocity components, the experiments show that this can even lead to a small decrease in filter performance for increasing N.

7.3. Sampling differences between the EnKF and the SEIK filter

The different sampling qualities of the EnKF and the SEIK filter are due to the distinct variants to generate the state ensembles in both algorithms. Interchanging the initialization methods between the algorithms results, at the first analysis step, in an exchange of the values of E_3, $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$. Neglecting model errors, both filters are equivalent at the first analysis step with respect to the update of the state estimate when using the same ensemble because the predicted error subspaces are identical. Such an equivalence does not exist for the update of $\mathbf{P}$ due to the implicit update of this matrix in the EnKF algorithm. In the EnKF, the

update of $\mathbf{P}$ is given implicitly by

$$\tilde{\mathbf{P}}^{\mathrm{a}} = (\mathbf{I} - \mathbf{KH})\tilde{\mathbf{P}}^{\mathrm{f}}(\mathbf{I} - \mathbf{KH})^{\mathrm{T}} + \mathbf{K}\tilde{\mathbf{R}}\mathbf{K}^{\mathrm{T}} + \mathcal{O}\left[\langle \delta\mathbf{x}^{\mathrm{f}}(\delta\mathbf{y}^{\mathrm{o}})^{\mathrm{T}}\rangle\right]. \quad (64)$$

Here, $\tilde{\mathbf{R}}$ is the observation error covariance matrix as sampled by the ensemble of observation vectors. $\tilde{\mathbf{P}}^{\mathrm{f}}$ and $\tilde{\mathbf{P}}^{\mathrm{a}}$ are the covariance matrices of the forecast and analysis state ensembles, respectively. The last term $\mathcal{O}[\langle \delta\mathbf{x}^{\mathrm{f}}(\delta\mathbf{y}^{\mathrm{o}})^{\mathrm{T}}\rangle]$ denotes the spurious covariances between the state and observation ensembles. In the SEEK and SEIK filters, this last term is zero and $\tilde{\mathbf{R}}$ is replaced by the prescribed matrix $\mathbf{R}$. Further, $\tilde{\mathbf{P}}$ denotes the rank-r approximated state covariance matrix. Hence, eq. (64) reduces to the KF update equation (eq. 12) for a covariance matrix $\tilde{\mathbf{P}}$ in the case of the SEEK and SEIK filters. For the EnKF, the sampled matrix $\tilde{\mathbf{R}}$ and the correlations between the state and observation ensembles insert noise into the analysis ensemble, which represents the state covariance matrix. Whitaker and Hamill (2002) discussed this effect in a simple one-dimensional system. In order to quantify the introduction of noise, the two definitions (61) and (62) of $\mathbf{B}$ can be examined. Without sampling errors, both definitions are equally valid. Thus, for the SEEK and SEIK filters the update matrices computed from either equation are identical. For the EnKF, the resulting update matrices are different.

In Table 4 the coefficients $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$ for update matrices computed with eqs. (61) or (62) are compared for the EnKF algorithm with $N = 30$ for type A. The values of $\rho_{\mathbf{B}}$ computed from the forecast covariances according to eq. (62) are about 1.5 times larger compared to those computed with eq. (61) from the analysis covariances. Despite this, the regression coefficients $\beta_{\mathbf{B}}$ remain almost unchanged. Also, the coefficients ρ_σ and β_σ show an analogous but much smaller ratio. The introduction of noise to the ensemble states at each analysis step leads to more unstable forecasts in the EnKF in comparison to the SEIK filter. Over the course of the assimilation process, the estimation error E_1 deviates increasingly for the two filters. This leads to the values of E_2 shown in Fig. 4 in which the performance difference between the EnKF and SEIK algorithms is larger than just for the first analysis.

Table 4. Comparison of the sampling quality of the update submatrices for the EnKF with $N = 30$ for experiments of type A. Shown are correlation ($\rho_{\mathbf{B}}$) and regression ($\beta_{\mathbf{B}}$) coefficients for sampled update submatrices computed from the forecast covariance matrix ($\mathbf{B}^{\mathrm{f}}$, eq. 62) and from the analysis covariance matrix ($\mathbf{B}^{\mathrm{a}}$, eq. 61). In addition, the correlation and regression coefficients (ρ_σ, β_σ) for the variance part of the height field update submatrix are shown

$\mathbf{B}$	Field	$\rho_{\mathbf{B}}$	$\beta_{\mathbf{B}}$	ρ_σ	β_σ
$\mathbf{B}^{\mathrm{f}}$, eq. (62)	**h**	0.305	0.091	0.961	0.071
$\mathbf{B}^{\mathrm{a}}$, eq. (61)	**h**	0.207	0.093	0.937	0.072
$\mathbf{B}^{\mathrm{f}}$, eq. (62)	**u**	0.126	0.015		
$\mathbf{B}^{\mathrm{a}}$, eq. (61)	**u**	0.082	0.014		

Table 5. Examination of the first analysis for experiments of type B with $N = 30$. The same quantities as in Table 2 are shown

	Field	E_3	$\rho_{\mathbf{B}}$	$\beta_{\mathbf{B}}$	ρ_σ	β_σ
EnKF		0.446	0.408	0.206	0.973	0.150
SEEK	**h**	0.431	0.425	0.171	0.944	0.119
SEIK		0.431	0.425	0.171	0.944	0.119
EnKF		1.045	0.175	0.090		
SEEK	**u**	1.135	0.366	0.213		
SEIK		1.137	0.367	0.213		

Table 6. Examination of the first analysis for experiments of type B with $N = 200$. The same quantities as in Table 2 are shown

	Field	E_3	$\rho_{\mathbf{B}}$	$\beta_{\mathbf{B}}$	ρ_σ	β_σ
EnKF		0.273	0.802	0.703	0.996	0.630
SEEK	**h**	0.269	0.847	0.651	0.991	0.533
SEIK		0.269	0.847	0.650	0.991	0.532
EnKF		0.981	0.519	0.559		
SEEK	**u**	0.872	0.766	0.729		
SEIK		0.875	0.766	0.728		

7.4. *Experiments with idealized set-up (type B)*

The sampling quality of the update matrices for experiments of type B for ensembles of size $N = 30$ and $N = 200$ are shown in Tables 5 and 6, respectively. For the SEEK and SEIK filters, the values of E_3, $\rho_{\mathbf{B}}$ and $\beta_{\mathbf{B}}$ are identical for **h** and almost identical for **u** and **v** for both ensemble sizes. Thus, the SEEK filter shows no problem with the forecast of the modes in this type of experiment. This can be related to the different structure of the covariance matrix, which leads to mode forecasts that provide realistic directions of the error subspace even for high eigenvalue indices. For **h**, the EnKF shows a slightly larger estimation error E_3 than the SEIK filter. This corresponds to the smaller values of $\rho_{\mathbf{B}}$, which show that the update matrices are less realistically sampled for the EnKF compared to the SEIK filter. However, the EnKF underestimates the amplitude of the covariances to a smaller degree than the SEIK filter does. The variance part of the update matrices is represented better by the EnKF than by the SEIK filter as is visible from the values of both ρ_σ and β_σ. The smaller regression coefficients in the case of the SEIK filter result from the low-rank approximation of the matrix $\mathbf{P}$, which systematically underestimates the overall variance. Due to the structure of $\mathbf{P}$ in the experiments of type B, as discussed in Section 5, the discarded variance is non-negligible here even for $N = 200$.

Compared to the experiments of type A, the estimates of the velocity components are much worse here. For $N = 30$, the values of E_3 even increase, showing that the sampled cross-covariances are not realistic. For $N = 200$, a small decrease of the estimation error is visible, which is stronger for the SEIK filter compared to the EnKF. Because the ideal values are $E_3(\mathbf{h}) = 0.2$ and $E_3(\mathbf{u}) = 0.75$, there will be no strong decrease in E_3 any longer for larger ensembles. However, over the whole assimilation period the performance of all three filters is better than at the first analysis step. While the non-assimilated state diverges from the true state, the data assimilation keeps the estimation error almost constant. This leads to the small values of the time-integrated estimation error E_2 displayed in Fig. 6.

8. Summary and conclusions

Three different ESKF algorithms, the EnKF, SEEK and SEIK algorithms, have been compared. With regard to their theoretical formulation, the algorithms are reviewed in the context of statistical estimation. The algorithms are presented for the first time in unified notation. The theoretical benefits of these algorithms over the EKF have been discussed. The comparison of the three filter algorithms, focusing on their capabilities in data assimilation with large-scale non-linear models, has shown that the EnKF and the SEIK filter are comparable as ensemble methods but use different ensemble initializations. In addition, the analysis scheme in the EnKF is computationally more costly and introduces noise into the ensemble, which is caused by the requirement of an ensemble of observation vectors. A main difference between the SEIK filter and the EnKF lies in the efficiency of the representation of the covariance matrix $\mathbf{P}$. In general, the EnKF will require a larger ensemble size N for the same performance as the SEIK filter, as is illustrated in detail utilizing identical twin experiments. The SEEK filter is initialized similarly to the SEIK algorithm, and also their analysis steps are analogous. However, due to the direct forecast of the covariance modes the predicted error subspace within the SEEK algorithm can be strongly distinct from that predicted by the SEIK filter.

All three filters are relatively easy to implement. The EnKF has the plainest structure, but also the SEIK filter, using the most advanced mathematics of the filters studied here, can be implemented within a few hundred lines of code.

In addition to the theoretical study, the behaviour of the three filter algorithms was assessed on the basis of identical twin experiments. The experiments utilized a shallow-water model with non-linear evolution and assimilated synthetic observations of the surface elevation. Two types of experiment were performed, which differed in the initialization of the state estimate and covariance matrix. For identical initial conditions, the filter algorithms showed quite different abilities to reduce the estimation error. In addition, the dependence on the ensemble size differed. Under some circumstances, the SEEK filter shows a behaviour distinct from the two other algorithms caused by the direct evolution of modes of the state covariance matrix. For the experiments of type A, in which the covariance matrix is dominated by a small number of large-scale modes, the performance of the SEEK filter differed strongly from that of the EnKF and SEIK algorithms. For experiments of type B, in which the covariance matrix is variance dominated, the SEEK and SEIK filters perform almost identically. The superior performance of the SEEK filter for the smallest ensemble sizes in experiments of type A appears to be a consequence of the highly red EOF spectrum used, but shows that a mode-based filter algorithm can, under some circumstances, yield a better filter performance than the ensemble-based filters. The SEEK filter has the computationally cheapest analysis scheme of the three filters examined here. It is well suited to filter rather coarse structures in which non-linearity is not pronounced. The EnKF and SEIK algorithms show similar convergence with increasing ensemble size. The SEIK filter exhibits superior performance compared to the EnKF due to the initialization by minimum second-order exact sampling. This leads to a better ensemble representation of the covariance matrix, in particular for small ensembles. Additionally, the SEIK filter does not suffer from noise introduced into the state ensemble by an observation ensemble as required by the original EnKF.

Statistical analyses of the quality of the sampled state covariance matrices have shown how the represented covariance matrices differ for the examined algorithms. The structure of the variances is in all cases quite well represented, but their amplitude is underestimated. Dependent on the structure of the covariance matrix, the low-rank initialization used in the SEEK and SEIK filters tends to underestimate the variances even more than the Monte Carlo initialization used in the EnKF. The full covariance submatrices for the single state fields are less well sampled by all three filters compared to the variances. The representation of the covariances for the height field is significantly better due to the variance-dominated structure than the cross-correlations between the height field and the velocity components. Here a better sampling quality can be achieved, at least for the SEIK filter and the EnKF, by increasing the ensemble size.

Overall, the numerical experiments have confirmed the theoretical findings. In particular, it is apparent that the SEIK filter is an ensemble algorithm comparable to the EnKF with the benefit of a very efficient scheme for analysis and resampling. Furthermore, initialization methods such as the second-order exact sampling are recommendable due to the better representation of the state covariance matrix, in particular for small ensembles. Like the EnKF and SEIK algorithms, the SEEK filter is able to provide good state estimates. The SEEK filter is, however, sensitive to the mode vectors it needs to evolve.

The experiments performed here are of course highly idealized. For example, an inclusion of the model error would be desirable. However, for the EnKF and SEIK algorithms, it can be expected that this will not lead to significant changes in the

relative filter performances, because both algorithms can treat the model error in the same way.

9. Acknowledgments

We wish to thank Ricardo Todling for a careful review and many helpful remarks. Also Gennardy Kivman and Stephan Frickenhaus are thanked for fruitful discussions and comments on the manuscript. We are grateful to Sergey Danilov for his help regarding the shallow-water model. Thanks also go to two anonymous reviewers for their thoughtful comments which helped to improve the consistency of the paper.

References

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.

Bennett, A. 1992. *Inverse Methods in Physical Oceanography*. Cambridge University Press, New York.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.

Brusdal, K., Brankart, J. M., Halberstadt, G., Evensen, G., Brasseur, P. and co-authors. 2003. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.* **40–41**, 253–289.

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

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteorol. Soc. Japan* **75**(1B), 257–288.

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

Evensen, G. 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10 143–10 162.

Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.

Evensen, G. and van Leeuwen, P. J. 1996. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon. Wea. Rev.* **124**, 85–96.

Fukumori, I. 2002. A partitioned Kalman filter and smoother. *Mon. Wea. Rev.* **130**, 1370–1383.

Golub, G. H. and van Loan, C. F. 1989. *Matrix Computations*. John Hopkins University Press, Baltimore, MD.

Hamill, T. M. and Whitaker, J. S. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–1790.

Heemink, A. W., Verlaan, M. and Segers, A. J. 2001. Variance reduced ensemble Kalman filtering. *Mon. Wea. Rev.* **(129)**, 1718–1728.

Hoteit, I., Pham, D-T. and Blum, J. 2002. A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *J. Mar. Syst.* **36**, 101–127.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.

Houtekamer, P. L. and Mitchell, H. L. 1999. Reply. *Mon. Wea. Rev.* **127**, 1378–1379.

Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137.

Ide, K., Courtier, P., Ghil, M. and Lorenc, A. C. 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Japan* **75**(1B), 181–189.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Kalman, R. E. and Bucy, R. S. 1961. New results in linear filtering and prediction theory. *Trans. ASME, J. Basic Eng.* **83**, 95–108.

Keppenne, C. L. and Rienecker, M. M. 2002. Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean circulation model. *Mon. Wea. Rev.* **130**, 2951–2965.

Lermusiaux, P. F. J. and Robinson, A. R. 1999. Data assimilation via error subspace statistical estimation. Part 1: Theory and schemes. *Mon. Wea. Rev.* **127**, 1385–1407.

Mitchell, H. L., Houtekamer, P. L. and Pellerin, G. 2002. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.* **130**, 2791–2808.

Natvik, L.-J. and Evensen, G. 2003. Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 1. data assimilation experiments. *J. Mar. Syst.* **40–41**, 127–153.

Ott, E., Hunt, B., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.

Pham, D. T. 2001. Stochastic methods for sequential data assimilation in strongly non-linear systems. *Mon. Wea. Rev.* **129**, 1194–1207.

Pham, D. T., Verron, J. and Roubaud, M. C. 1998a. A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.* **16**, 323–340.

Pham, D. T., Verron, J. and Gourdeau, L. 1998b. Singular evolutive Kalman filters for data assimilation in oceanography. *C. R. Acad. Sci., Ser. II* **326**(4), 255–260

Sadourny, R. 1975. The dynamics of finite-difference models of the shallow-water equations. *J. Atmos. Sci.* **120**, 680–689.

Tippett, M. K. Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

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

van Leeuwen, P. J. 2001. An ensemble smoother with error estimates. *Mon. Wea. Rev.* **129**, 709–728.

Verlaan, M. and Heemink, A. W. 1995. Reduced rank square root filters for large scale data assimilation problems. In: *Proceedings of the 2nd International Symposium on Assimilation in Meteorology and Oceanography*. World Meteorological Organization, March 1995, 247–252.

Verron, J., Gourdeau, L., Pham, D. T., Murtugudde, R. and Busalacchi, A. J. 1999. An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the tropical Pacific ocean: method and validation. *J. Geophys. Res.* **104**(C3), 5441–5458.

Voorrips, A. C., Heemink, A. W. and Komen, G. J. 1999. Wave data assimilation with the Kalman filter. *J. Mar. Syst.* **19**, 267–291.

Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1927.