INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **18** (2002) 1–23 PII: S0266-5611(02)20832-1

Combining geostatistics and Kalman filtering for data assimilation in an estuarine system

Laurent Bertino[1,3], **Geir Evensen**[2] **and Hans Wackernagel**[1]

[1] Centre de Géostatistique, Ecole des Mines de Paris, 35 rue Saint Honoré, F–77305 Fontainebleau, France
[2] Nansen Environmental and Remote Sensing Center, Edvard Griegsvei 3a, N–5059 Bergen, Norway

E-mail: bertino@cg.ensmp.fr

Received 12 January 2001, in final form 9 April 2001
Published 8 January 2002
Online at stacks.iop.org/IP/18/1

Abstract

Data assimilation (DA) has been applied in an estuarine system in order to implement operational analysis in the management of a coastal zone. The dynamical evolution of the estuarine variables and corresponding observations are modelled with a nonlinear state-space model. Two DA methods are used for controlling the evolution of the model state by integrating information from observations. These are the reduced rank square root (RRSQRT) Kalman filter, which is a suboptimal implementation of the extended Kalman filter, and the ensemble Kalman filter which allows for nonlinear evolution of error statistics while still applying a linear equation in the analysis. First, these methods are applied and examined with a simple 1D ecological model. Then the RRSQRT Kalman filter is applied to the 3D hydrodynamics of the Odra lagoon using the model TRIM3D and water elevation measurements from fixed pile stations. Geostatistical modelling ideas are discussed in the application of these algorithms.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Estuarine systems play a key role in our global environment since their richness and diversity strongly influence connecting ecosystems. The problems of decreasing fish resources and of endangered migrating birds are some of the most striking examples. The ecological richness of estuaries is threatened by the huge amount of pollution brought by urban or agricultural activities in the river basin. River estuaries also play an important socio-economical role in international maritime transports and as supports for fishing and touristic activities. Sustainable

[3] Corresponding author.

management of such areas need some help-for-decision tools. Therefore monitoring and forecasting methods have gained interest as computing capabilities have increased enough to allow high-resolution 3D simulations of such a system. This monitoring requires two types of information: (1) a numerical model based on the physical equations that describe the biological, chemical and physical interactions in the system and (2) *in situ* measurements used for model calibration but also in a more intensive way through a data assimilation (DA) scheme that improves the model forecast and evaluates the forecast accuracy.

The numerous DA techniques are often classified into variational methods based on the optimal control theory and sequential methods derived from the theory of statistical analysis. However, most techniques can be derived as suboptimal solution methods for the same general problem definition: 'given a dynamical model and a set of measurements on a spatial and temporal region; what is the most likely state estimate given information about the dynamics from the model and information about the real state from measurements'? This leads to a well posed mathematical problem if one includes information about prior error statistics for both the model and the measurements. Formulating this problem using Bayesian statistics, one can derive traditional methods such as the representer solution of the weak constraint variational problem by assuming Gaussian priors, the strong constraint adjoint method by additionally assuming the model is perfect, and the Kalman filter methods by only propagating information forward in time (see [16]).

The most commonly used method for solving the variational problem, i.e. the adjoint method [23], is very efficient for direct minimization with the model acting as a strong constraint (the model is assumed exact). In this approach, optimization is performed on the initial state by iterated gradient descents. To compute the gradient the adjoint model running backward in time has to be deduced from the direct model, which is a complicated task especially when the system equations are nonlinear. However the adjoint method in the latest and conceptually most complex developments has become extremely popular in atmospheric sciences for successful operational weather forecasts and since the mid-1990s is progressively replacing the optimal interpolation (OI) techniques, as was stated during the Third World Meteorological Organization International Symposium on DA of Observations in Meteorology and Oceanography (Québec City, Canada, 7–11 June 1999). But the OI techniques [10] are still recognized for their numerical efficiency. Even more recently, the adjoint method has also been applied in oceanography [29, 35] and successfully assimilated satellite observations of sea-surface heights in a quasi-geostrophic ocean circulation model [24].

Note, however, that the representer method can be used to solve for the weak constraint variational problem and it its latest formulation uses much the same machinery as the adjoint method and has a similar computational cost [2, 30].

The prototype of the sequential methods, the Kalman filter (KF), was designed in the 1960s for the optimal control of systems governed by linear equations and was first introduced in oceanography at the end of the 1980s [20]. The KF has been mostly applied to quasi-linear tropical ocean situations [6, 17, 39] but it was extended to the nonlinear case [22] with model linearization. This extended Kalman filter (EKF) has been examined with different dynamical models by Evensen [14], Gauthier *et al* [19], and Miller *et al* [28]. A common conclusion from these studies is that the EKF has an apparent closure problem since the error evolution is computed using the model tangent linear operator, and thus, if the model is in an unstable regime there is no nonlinear saturation of error growth.

In high-dimensional problems such as in ocean and coastal models the EKF often has to be simplified by a sub-optimal scheme (SOS hereafter) for reduction of the computational burden. For the common example of a discretized state of size 300×300 the storage of a covariance matrix of approximate size $10^5 \times 10^5$ requires one gigabyte.

Two classes of SOS can be distinguished. Those with simplified model error dynamics, and those using approximations of the error covariance matrix. Among the first group, Dee [11] assumed the forecast error to be divergence free and nearly geostrophically balanced, Fukumori and Malanotte-Rizzoli [18] used a coarser grid for the error covariance propagation. Cohn and Todling [9] used a singular value decomposition of the tangent linear operator in the EKF. A so-called 'kriged Kalman filter' has been proposed by Wikle and Cressie [41] and Mardia *et al* [25]. Their approach, similar to the assimilation of tropical sea levels in [6], is based on a decomposition of both the propagation operator and the state vector on an orthogonal basis of the state-space, thus assuming that the dynamical model is linear and can be decomposed into orthogonal time invariant factors. This approach is highly dependent on the relevance of the decomposition basis and factors *a priori* discarded are not allowed to re-appear during the DA process. Optimization of the decomposition basis has to be done iteratively and *a posteriori* considering all data. Further extensions of these methods to nonlinear and time-varying models have not been proposed yet.

Among the second class of SOS, a steady state KF, i.e. a KF where the gain is computed off-line and is constant, has been applied by Heemink [21] for storm surge forecasting and by Cañizares *et al* [5] for 2D hydrodynamics. Eigenvalue decomposition of the covariance matrix has led to several techniques, e.g., the reduced rank square root (RRSQRT) Kalman filter [37], the singular evolutive extended/interpolated Kalman filters SEEK and SEIK [3, 32] and the Cohn and Todling method based on a Lanczos algorithm [9].

The ensemble Kalman filter (EnKF) was developed by Evensen [15], to resolve previously reported problems with the EKF and linearized error covariance evolution. Starting from a general framework where the error statistics are described by a probability density function (PDF) which evolves according to the Fokker–Planck equation, it was possible to derive a Monte Carlo method for evolution of error statistics. Thus, the EnKF converges to an exact nonlinear error evolution with an infinite sample size. The EKF can be derived from the general EnKF formulation as a special case, and for linear dynamics the KF and the EnKF will produce identical results in the limit of an infinite sample size. The EnKF has been applied with ocean circulation models by Evensen and van Leeuwen [15] and ecological models by Eknes and Evensen [12].

The schemes by Verlaan and Heemink [37] and Pham *et al* [32] were originally derived as simplifications of the EKF. However, by using the nonlinear model operator for evolving the eigen vectors forward in time they can handle highly nonlinear dynamics.

The present paper first reviews the theory of DA in section 2 and focuses on KF in section 3 including the description of the RRSQRT KF and the EnKF, both very popular in oceanographical applications.

Those two DA techniques were extensively compared by Cañizares [4] with a 2D coastal hydrodynamical model for the following criteria: computer time, sensitivity to systematic errors in the model forcing and robustness to incorrect definition of error statistics. He valued them as equally efficient and robust, but his results cannot necessarily be generalized to any field of application. Verlaan and Heemink [38] also applied the two techniques to the Lorenz equations and found that the EnKF was more reliable for highly nonlinear dynamics. In the present paper a comparison of both techniques on an ecological model—for the first time as far as we know—is presented in section 4. The test case is a simple ecological model describing the evolution of nutrients, phytoplankton and zooplankton concentrations in a synthetic water column under light conditions favourable to photosynthesis. The interest of this case is that dynamics are nonlinear and very sensitive to changes in the initial state. Even though the two DA methods are able to recreate accurately a reference annual cycle they both tend to overestimate their own error. The experimental differences between the two methods do not seem to be significant.

Figure 1. Map of the Odra lagoon with the measurement stations Ückermünde, Karnin and Wolgast, the validation stations Odh1 and Odh2. The measurements of the station Koserów are used for boundary conditions. The map also shows the three outlets Peenestrom (1), Świna (2) and Dźwina (3) and the mouth of the Odra river (4).

Independently from this comparison, the following section presents an application of the RRSQRT KF to the estimation of the hydrodynamical state of the Odra lagoon during the extreme event of the Odra flood of summer 1997.

The Odra lagoon straddles the Polish–German border. Its surface covers about 600 km^2, its average depth is 5 m, thus its volume is about 3×10^9 km^3. Fresh water enters the lagoon from several rivers, with the Odra being the main input. Intermittently salt water intrudes in small amounts through the three narrow channels from the Baltic sea (see figure 1). Such intrusions are caused by water level differences between the Baltic sea and the lagoon and have a large impact on the whole ecosystem. The system is observed by five fixed pile stations. The water level measurements from three of these five points are assimilated while those from the two remaining stations are kept aside for validation.

The 3D hydrodynamical model used is TRIM3D, it has been adapted to the Odra lagoon [34] and already efficiently reproduces water levels. The question addressed here is: 'can DA improve the results of the model'? The long term goal of this study is to build with DA a tool for state forecasting giving also an estimate of the forecast error.

The RRSQRT KF has already been applied to two and 3D coastal hydrodynamics [5, 37], but the Odra lagoon system required a different implementation than coastal and marine systems since the open boundaries are very narrow. Previous DA of water levels in the Odra lagoon has been presented by Wolf *et al* [42] using the RRSQRT KF and presenting a simple wave model for evaluating the spatial covariance structure of the model error. In the present paper, geostatistical ideas are discussed in the application of DA to an estuarine system. The RRSQRT

KF is efficient for assimilation of water levels, it significantly reduces the model errors within computation times that are short enough to enable operational forecasting. From there, a method for improving the observation network is given and application to other variables can be investigated. Furthermore, knowing the hydrodynamical state with high accuracy will be helpful for predictions of nutrient transport and ecological monitoring of the lagoon.

2. The data assimilation problem: continuous and discretized

2.1. The state-space model

Let $\boldsymbol{Z}(x, t)$ be a k-dimensional random function of space and time. In this paper, as usual in applied environmental issues, we will discretize the spatial dependences, that is, we will consider a multivariate random function $\boldsymbol{Z}(t)$ where all l locations x for all k variables are different components of the $k \times l$ state vector $\boldsymbol{Z}(t)$, without loss of generality.

The state $\boldsymbol{Z}(t)$ is supposed to follow the equations of the physical model

$$\frac{\partial \boldsymbol{Z}}{\partial t} = \boldsymbol{f}(\boldsymbol{Z}), \tag{1}$$

where $\boldsymbol{f}$ is often a nonlinear function of $\boldsymbol{Z}$ and of the spatial derivatives of $\boldsymbol{Z}$. Equation (1) is called the evolution equation of the system.

At every measurement step t_n the system state $\boldsymbol{Z}(t_n)$ is observed through a measurement vector $\boldsymbol{z}_n$ of m observations. The observed variables are not necessarily the state variables of $\boldsymbol{Z}$ but they are linked through the observation equation

$$\boldsymbol{z}_n = L\boldsymbol{Z}(t_n), \tag{2}$$

L being a $m \times (k \times l)$ matrix.

The system of both equations (1) and (2) is called a state-space model. This system is ill-conditioned and generally has no solution since every measurement contradicts the model predictions, but the introduction of random errors in both equations allows for finding an optimal solution that satisfies best both constraints.

When discretized in time, the above equations (1) and (2) become for timestep n:

$$\boldsymbol{Z}_{n+1} = \boldsymbol{f}_n(\boldsymbol{Z}_n) \tag{3}$$

$$\boldsymbol{z}_n = L\boldsymbol{Z}_n, \tag{4}$$

where $\boldsymbol{Z}_n$ remains a vector of length $k \times l$ and $\boldsymbol{z}_n$ a vector of length m.

2.2. The support effect

Within the discretized framework, it should be noted that measurements and model state values have different spatio-temporal supports for most real case studies: a measurement is an average of the variable of interest on a spatio-temporal volume that is a technical characteristic of the sensing device. The elements of the model state, on the other hand, are averages on a whole grid cell during a model time step that are dependent on the system modelling and represent much larger volumes.

Therefore, measurements usually have much higher variability than model output and they are usually smoothed by averaging in order to compare model output with observations that have similar variability. From a rigorous point of view, the smoothing step before comparison is an abusive operation and solves the support question only in appearance. In the case of assimilation of turbulent water currents this operation can even lead to measurement biases, e.g. taking a vortex that has a radius short enough to be included in a model grid cell. If the

sampling is not performed in the eye of the vortex, measurements will indicate a constant flow in the direction tangent to the vortex, while the average flow on the grid cell may be zero. In such a pathological example, time averaging cannot sweep out the measurement bias and spatial averaging requires measurements with a resolution finer than the model description, in which case we would barely need any modelling.

Statistical change-of-support models are reviewed in Chilès and Delfiner [8] but will not be discussed in our applications since we limit the study to the case of very sparse measurement stations (<10) which is insufficient for such models.

Accounting for the support effect for both applications will be discussed in the respective description of each DA systems.

2.3. General solution to the nonlinear problem

The evolution equation (1) can be perturbed by a random model noise $\mathrm{d}\boldsymbol{w}_t$, classically considered a multivariate centred reduced white noise, and it becomes an Itô stochastic differential equation,

$$\mathrm{d}\boldsymbol{Z}(t) = \boldsymbol{f}(\boldsymbol{Z}, t)\,\mathrm{d}t + \boldsymbol{g}(\boldsymbol{Z}, t)\,\mathrm{d}\boldsymbol{w}_t, \tag{5}$$

where $\boldsymbol{g}(\boldsymbol{Z}, t)$ is the function spreading the noise on the whole model grid. When defined, the PDF, $\phi_t(\boldsymbol{Z})$, of $\boldsymbol{Z}(t)$ evolves in time according to the Fokker–Planck equation,

$$\frac{\partial \phi_t}{\partial t} = -\sum_i \frac{\partial (f_i(\boldsymbol{Z}, t)\phi_t)}{\partial Z_i} + \sum_{i,j} \frac{1}{2} G_{ij} \frac{\partial^2 \phi_t}{\partial Z_i \partial Z_j}, \tag{6}$$

with $G = \overline{\boldsymbol{g}\boldsymbol{g}^T}$. This equation is a deterministic advection–diffusion equation and can be solved as a partial differential equation in low state dimensions, e.g. less than four. For higher-dimensional problems the numerical discretization becomes practically impossible, and the solution process is further complicated by constraints of positivity and unit sum of probabilities. The measurements are integrated with the model system by Bayes theorem which states that the posterior PDF can be expressed as,

$$\phi_t(\boldsymbol{Z}|z_1, \ldots, z_n) = A\phi_t(\boldsymbol{Z}|z_1, \ldots, z_{n-1})\phi_t(z_n|\boldsymbol{Z}) \tag{7}$$

where A is a normalization constant. The first PDF on the right-hand side is the PDF resulting from the model integration from the previous data time t_{n-1} to t_n, and the second PDF is the density for the observations. Thus, the general sequential DA problem would involve forward integration of the Fokker–Planck equation (6) and update of the PDF using Bayes theorem (7) every time measurements are available. A more elaborate discussion is given by Evensen and van Leeuwen [16].

Miller *et al* [27] found a general numerical solution to the double-well problem, which is a nonlinear monovariate problem. But for high-dimensional problems such as ocean models no practical solution method has been found yet.

In the case of a linear dynamical model and zero-mean Gaussian errors, it can be shown that the KF is the general sequential solution of the weak constraint problem under the classical Markovian hypotheses [31]. However, in the general case with nonGaussian prior error statistics and nonlinear evolution of error statistics all existent assimilation techniques provide solutions which do not coincide with the maximum likelihood solution and therefore must be characterized as sub-optimal.

3. Sequential DA methods

In the sequential approach, the measurements z_n are used to correct the state vector. The unknown true state at time n will be denoted $\boldsymbol{Z}_n^{\mathrm{t}}$, while the associated model forecast before correction will be $\boldsymbol{Z}H_n^{\mathrm{f}}$ and the analysed state will be $\boldsymbol{Z}_n^{\mathrm{a}}$. Assuming Gaussian PDF in equation (7), the analysis step equation (9) is linear. This analysed state $\boldsymbol{Z}^{\mathrm{a}}$ then recursively feeds the next propagation step,

$$\boldsymbol{Z}_n^{\mathrm{f}} = \boldsymbol{f}_n(\boldsymbol{Z}_{n-1}^{\mathrm{a}}), \tag{8}$$

$$\boldsymbol{Z}_n^{\mathrm{a}} = \boldsymbol{Z}_n^{\mathrm{f}} + K_n(z_n - L\boldsymbol{Z}_n^{\mathrm{f}}), \tag{9}$$

where the *a priori* nonlinear function $\boldsymbol{f}_n$ denotes the model equations integrated from the time $n-1$ to n, $\boldsymbol{f}_n$ is varying in time as the phenomena under consideration can be subject to time varying forcing like light conditions or a river discharge.

Techniques of OI [10] are based on empirical evaluation of the matrix K_n by spatial statistics. Other methods evaluating K_n from empirical criteria are referred as 'simple methods', a review is given by Robinson *et al* [33]. 'Simple methods' are easy to implement but they provide no information about the uncertainty in the state estimate or the time evolving covariance functions which describe the influence of observations on the model state.

Hereafter we will consider the KF in which a measure of the state uncertainty evolves jointly with the state estimate.

3.1. Principles of the Kalman filter: the stochastic model

Kalman filtering applied to oceanography is described in detail by Bennett [1]. Discretization is required in our applications for resolving the advection–diffusion equations. The basic KF is designed for linear models F_n such that

$$\boldsymbol{Z}_{n+1} = F_n\boldsymbol{Z}_n + \boldsymbol{q}_n, \tag{10}$$

where $\boldsymbol{q}_n$ is the model error, a random noise corresponding to the $\boldsymbol{g}(\boldsymbol{Z}, t)\,\mathrm{d}\boldsymbol{w}_t$ term in equation (5).

The observation equation (2) also has a random additive error $\boldsymbol{r}_n$, called the measurement error in

$$z_n = L\boldsymbol{Z}_n + \boldsymbol{r}_n. \tag{11}$$

The classical Markovian hypotheses are that both error processes $\boldsymbol{q}_n$ and $\boldsymbol{r}_n$ are temporally independent centred Gaussian white noise:

- The errors are Gaussian and unbiased,

$$E(\boldsymbol{q}_n) = E(\boldsymbol{r}_n) = 0.$$

- The model and measurement errors are uncorrelated in time:

$$\forall p \neq n, \mathrm{corr}(\boldsymbol{r}_p, \boldsymbol{r}_n) = \mathrm{corr}(\boldsymbol{q}_p, \boldsymbol{q}_n) = 0.$$

- They are also independent from each other,

$$\forall (n, p), \mathrm{corr}(\boldsymbol{r}_n, \boldsymbol{q}_p) = 0.$$

The random vector $\boldsymbol{q}_n$ is spatially defined on the model grid and $\boldsymbol{r}_n$ on the measurement locations. Their probability law is determined by their variance–covariance matrices respectively Σ_m and Σ_o that can be made time-dependent without any influence on the ongoing calculations. These are symmetric matrices of rank respectively $(k \times l)$ and m.

The measurement error reflects inaccuracy in the measurement value and position. Therefore, its covariance matrix is rather simple to evaluate, it is a diagonal matrix if we assume these inaccuracies to be independent from one station to another, and the variance of these errors can be computed both *a priori* from the technical characteristics of the sensors and *a posteriori* from the short range variability in the measurement timeseries. The model error is more delicate to determine since it reflects many sources of uncertainty among which the discretization, the model forcing, inaccurate parameter values and neglected dynamical processes. Furthermore the knowledge of the entire 'true state' of the system is generally too fragmentary for any *a posteriori* evaluation of these error statistics. Thus filling in the covariance matrix Σ_m requires severe assumptions. As a function of space, $\boldsymbol{q}_n$ is generally assumed stationary of order 2 which means that, in addition to the zero mean assumption above, its covariance function

$$C(\boldsymbol{x}, \boldsymbol{h}) = \mathrm{cov}(\boldsymbol{q}_n(\boldsymbol{x} + \boldsymbol{h}), \boldsymbol{q}_n(\boldsymbol{x}))$$

has to be translation independent

$$\forall(\boldsymbol{x}, \boldsymbol{h}), C(\boldsymbol{x}, \boldsymbol{h}) = C(\boldsymbol{h})$$

here written in spatial continuous notations, so that it can be fitted with covariance functions (see [8] or [40]). The DA results are highly dependent on the stochastic model since it defines the trust we place in each information. Unfortunately the knowledge we have of both measurement and especially model noise is often very poor and simplifying assumptions are needed in most case studies.

3.2. Equations of the linear KF

According to the stochastic model above, all random variables are Gaussian and therefore the maximum likelihood estimator of $\boldsymbol{Z}_n^{\mathrm{t}}$ coincides with its expectation, conditionally to the data. The calculations of the estimators and of their associated error covariance matrices can be found in [26]. The time step equation of the linear KF can be expressed as

$$\boldsymbol{Z}_n^{\mathrm{f}} = E(\boldsymbol{Z}_n^{\mathrm{t}} | \boldsymbol{Z}_0, \boldsymbol{z}_i, i = 1, \ldots, n-1) \tag{12}$$

$$= E(\boldsymbol{Z}_n^{\mathrm{t}} | \boldsymbol{Z}_{n-1}^{\mathrm{a}}) \tag{13}$$

$$= F_n \boldsymbol{Z}_{n-1}^{\mathrm{a}}, \tag{14}$$

the estimation error $\epsilon_n^{\mathrm{f}} = \boldsymbol{Z}_n^{\mathrm{t}} - \boldsymbol{Z}_n^{\mathrm{f}}$ has zero expectation and the evolution of its covariance matrix is obtained by taking the moments of the Fokker–Planck equation (6),

$$C_n^{\mathrm{f}} = F_n C_{n_1}^{\mathrm{a}} F_n^T + \Sigma_m. \tag{15}$$

For the measurement step equation, the Bayes theorem (7) under the KF assumptions—Gaussian distribution and Markovian observation error processes—shows that the maximum likelihood estimator coincides with the least squares estimator

$$\boldsymbol{Z}_n^{\mathrm{a}} = E(\boldsymbol{Z}_n^{\mathrm{t}} | \boldsymbol{Z}_0, \boldsymbol{z}_i, i = 1, \ldots, n) \tag{16}$$

$$= E(\boldsymbol{Z}_n^{\mathrm{t}} | \boldsymbol{z}_n) \tag{17}$$

minimizing the variance of the estimation error $\epsilon^{\mathrm{a}} = \boldsymbol{Z}_n^{\mathrm{t}} - \boldsymbol{Z}_n^{\mathrm{a}}$. This results in the same linear analysis equation as in all 'simple' sequential DA methods (9) cited in the beginning of the section but here the computation of the gain matrix K_n is more elaborate. The gain matrix K_n and updated covariance C_n^{a} are

$$K_n = C_n^{\mathrm{f}} L^T (L C_n^{\mathrm{f}} L^T + \Sigma_o)^{-1} \tag{18}$$

$$C_n^{\mathrm{a}} = C_n^{\mathrm{f}} - K_n L C_n^{\mathrm{f}}. \tag{19}$$

Then from a geostatistical point of view the analysis step is an interpolation of the model-to-data misfit $\boldsymbol{z} - L\boldsymbol{Z}^{\mathrm{f}}$ to the whole model grid by simple cokriging.

Depending on which variable is assimilated, the measurement step can respect or violate the physical properties of the system. For example the assimilation of water levels introduces masses of water where the model elevation was lower than the measured elevation and other masses of water vanish in the opposite case. This operation may contradict the mass balance principle. However in an open boundary system the exchanges through the boundary conditions are only inaccurately known and this uncertainty can be taken for responsible for the correction to be done in the mass balance.

Further, the dangers of assimilating data with respect to mass balance have been stated by Cañizares [4], when he builds a KF with model errors only in the momentum equation. This conservative KF may provide a solution worse than an ill-calibrated model alone, whereas assimilating data without respect to mass balance can yield better estimates. Cañizares then emphasizes that DA should be conservative only if the initial solution are accurate enough, or else the KF will displace the initial error, but may not be able to get rid of it.

3.3. Extension of the KF to the nonlinear case

Ocean and coastal dynamics are often nonlinear and the KF method has to be modified to the EKF, in which the system propagation equation is nonlinear:

$$\boldsymbol{Z}_n = \boldsymbol{f}_n(\boldsymbol{Z}_n, \boldsymbol{q}_n) \tag{20}$$

and the covariance matrix is propagated by successive linearizations of the model $\boldsymbol{f}_n$, when the Taylor expansion of the model is generally stopped at the first-order. Developments to higher orders can be necessary in some cases, e.g. see [14] for a limitation of the EKF due to neglecting higher order moments, but this is rarely done due to the increasing complexity of the equations and the numerical load associated with storage and evolution of the higher order statistics.

Here two alternative approaches are used, one based on a decomposition onto a reduced space spanned by eigenvectors of the error covariance matrix, the other based on an ensemble representation of error statistics and error evolution using Monte Carlo simulations.

3.3.1. Reduced rank square root (RRSQRT) Kalman filter. The RRSQRT was developed by Verlaan and Heemink [37]. The method relies on a simplification where the covariance matrix C_n is represented by only the q largest eigenvalues in the eigen decomposition

$$C_n = P_n D_n P_n^T. \tag{21}$$

The eigen decomposition is performed at every assimilation step, and only the square root S_n of $C_n = S_n S_n^T$ with reduced rank p is computed through the KF process, in both analysed and forecast stages S_n^{a} and S_n^{f}. The ith column of S_n is interpreted as a perturbation vector $\Delta \boldsymbol{Z}_{i,n}$ and the ensemble of p perturbations is an orthogonal basis of the p-dimensional subspace of the state-space where the state error vector is expected to take its values.

The propagation of the perturbation by the physical model in the ith direction, is performed through a first-order approximation

$$\begin{aligned} \frac{\partial \boldsymbol{f}}{\partial Z_i}(\boldsymbol{Z}_{n-1}^{\mathrm{a}}) &= \lim_{\epsilon \to 0} \frac{\boldsymbol{f}(\boldsymbol{Z}_{n-1}^{\mathrm{a}} + \epsilon \Delta \boldsymbol{Z}_{i,n-1}^{\mathrm{a}}) - \boldsymbol{f}(\boldsymbol{Z}_{n-1}^{\mathrm{a}})}{\epsilon} \\ \boldsymbol{Z}_n^{\mathrm{f}} + \Delta \boldsymbol{Z}_{i,n}^{\mathrm{f}} &\approx \frac{\boldsymbol{f}(\boldsymbol{Z}_{n-1}^{\mathrm{a}} + \epsilon \Delta \boldsymbol{Z}_{i,n-1}^{\mathrm{a}}) - \boldsymbol{f}(\boldsymbol{Z}_{n-1}^{\mathrm{a}})}{\epsilon}. \end{aligned} \tag{22}$$

If the linearization parameter ϵ tends to zero, under usual differentiability assumptions, the above fraction tends to the derivative of $\boldsymbol{f}$ in the $\Delta \boldsymbol{Z}^{\mathrm{a}}_{i,n-1}$ direction, and one can conclude that the ϵ should be an infinitesimal real parameter. However due to locally discontinuous or strongly nonlinear dynamics, setting small values for ϵ can lead to divergent error growth as in our applications below. These instabilities usually disappear when increasing the value of ϵ, which then requires some tuning. It may also be the case that the instability is due to the integration of orthogonal perturbations that are physically meaningless in the nonlinear model equations. These perturbations are empirical orthogonal functions (EOFs) defined by statistical criteria but may not be constrained to physical properties of the variable studied. Therefore as these EOFs are introduced in the physical model through the equation above, their unrealistic features could lead to model divergence. But in this last hypothesis it is surprising that the model divergence disappears when ϵ increases in our examples below. In both hypotheses the instability remains unpredictable and this is a drawback of the RRSQRT KF and of many other EKF schemes using such first-order approximations.

About computational performance, the number p of eigenvalues to be retained is often inferior to a hundred while the state-space has dimension 10^5 in most hydrodynamical applications. As the most time consuming steps are the forward propagation in time of the reduced square root covariance matrix, and the matrix algebra operations, this reduction of the state-space represents a speeding of a factor up to one thousand compared to the full rank EKF. More details are given in [4] with variable number of retained eigenmodes. Verlaan [36] also describes methods for further speeding up of the RRSQRT KF.

3.3.2. Ensemble Kalman filter (EnKF). In the EnKF, introduced by Evensen [15], the model state and the error statistics are represented by a Monte Carlo method on an ensemble of simulated states for solving the Fokker–Planck equation (6). The analysis approximates the Bayes equation (7) by assuming predicted statistics to be Gaussian. The prediction expectation and uncertainty are approximated by the first- and second-order statistical moments of the ensemble population. Therefore the covariance matrix operations (15) and (19) are not necessary any more but the state propagation (8) and update (9) are repeated as many times as there are members in the ensemble. As a consequence, no model linearization is necessary as in the EKF, thus avoiding the tuning of a linearization parameter as in the RRSQRT KF procedure above. It should also be noticed that all EnKF members are realistic state vectors and should not lead to disruptions in the physical model.

The key parameter of the EnKF is the size of the ensemble. The ensemble members are not orthogonal vectors but only randomly scattered equiprobable states. Thus, we need a large ensemble, with size much larger than the number of dimensions we need to span, in order to compensate for information redundancy. Usual ensemble sizes are around 100–200 samples for practical applications so that the method is rather expensive compared to its competitors in terms of computation time. Here, as the computation of the Kalman gain in equation (18) is negligible compared to a model integration, the filter requirements are approximately proportional to the ensemble size. Therefore a significant gain can also be achieved by using the EnKF instead of the EKF.

3.4. Extension to the nonGaussian case

An extension of the EnKF method to nonGaussian PDF is straightforward by anamorphosis. When the variables of interest $\boldsymbol{Z}$ are nonlinear functions of a Gaussian variable $\boldsymbol{Y}$

$$\boldsymbol{Z} = \boldsymbol{h}(\boldsymbol{Y}),$$

the EnKF is performed on the Gaussian transform $\boldsymbol{Y} = \boldsymbol{h}^{-1}(\boldsymbol{Z})$ and first- and second-order statistics are computed on the back-transformed ensemble. For example, some positive variables such as concentrations are wrongly described by the Gaussian model and can be treated by assimilation of their log transform, the log normal model being also useful in applications where measurement uncertainty depends on the measurement value and when positivity of some variables has to be ensured in the dynamical model.

Other methodologies which represent error statistics by the covariance matrix rather than using Monte Carlo simulations are more difficult to extend to nonGaussian models. Indeed, if we refer to the probabilistic interpretation of the assimilation in equation (16), a back transform of the assimilated Gaussian $\boldsymbol{Y}^{\mathrm{a}}$ by the nonlinear function $\boldsymbol{h}$ is a biased estimator of $\boldsymbol{Z}^{\mathrm{a}}$ since the expectation does not commute with nonlinear functions. Bias correction is possible (see [8] for a discussion) but makes the system much more complex.

4. A simple 1D ecological model

4.1. The model

A simple zero-dimensional ecological model was used by Evans and Parslow [13] to describe annual cycles of plankton and especially spring blooms. Three variables—nutrient (N), phytoplankton (P) and herbivores (H)—interact through differential equations, the units for all ecosystem variables are mmol of nitrogen per m^3 (mmol N m^{-3}). Herbivores are grazing phytoplankton and their mortality is supposed constant. Nutrients are taken up by phytoplankton in the photosynthesis process, and phytoplankton growth responds to light. The light conditions and mixed layer depth vary throughout the year but present the same pattern from one year to another. All parameters have been calibrated so that the annual cycle is appropriate for a water column offshore Newfoundland in the Atlantic Ocean. The model step divides the year in 400 time intervals.

This model was extended by Eknes and Evensen [12] to contain a vertical dimension and diffusion terms along this direction have been added to the differential equations. Their model characteristics have been reproduced here: 20 vertical cells of height 10 m form a 200 m deep water column. The boundary conditions specify zero flux of the three variables at the water surface and a constant nutrient flux of 10 mmol N m^{-3} at the bottom, simulating an idealized nutrient input from the sediment phase. Phytoplankton and herbivores concentrations at the bottom are set to zero.

A reference year was simulated by starting the model with initial conditions set to 10 mmol N m^{-3} for the nutrients and 0.1 mmol N m^{-3} for both the phytoplankton and the herbivores concentrations and spinning up the model for 5 years until the annual cycle repeats itself accurately. Using this reference year as a true state, Eknes and Evensen performed DA and reported the capability of the EnKF to control the evolution of the system under different observation conditions.

A common feature of ecological models is that the chemical and biological growth rates are modelled by Michaelis hyperbolae. In the non-saturated case these rates are directly proportional to the substrate concentrations as long as this substrate concentration is much lower than the saturation concentration. This model leads to exponential-like population growth and the dynamics can be strongly nonlinear.

The ecological model is also very sensitive to initial conditions. Random errors of order 0.01 mmol N m^{-3} in the initial concentrations of the three variables can produce large deviations from the reference solution up to 1 mmol N m^{-3} for all three variables which represents 100% in relative variations. The main deviation occurs during the spring bloom as shown in figure 2 for

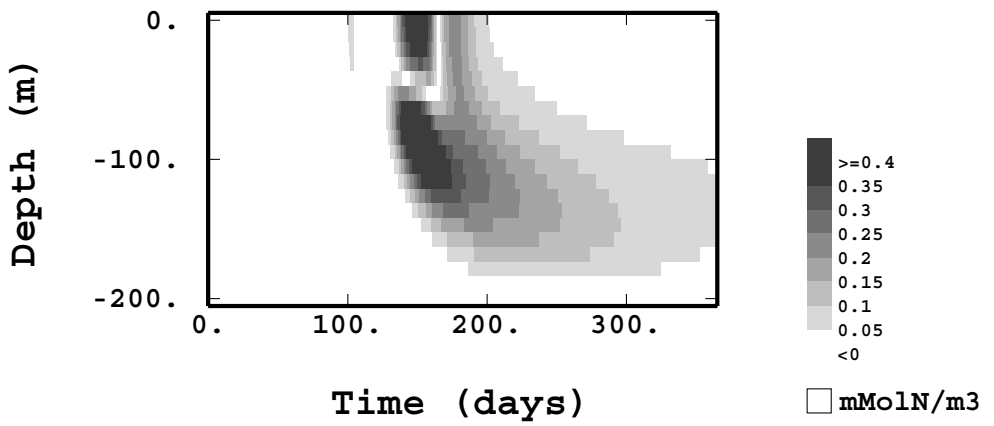

Figure 2. Absolute differences between nutrient concentrations in the reference case and those from a single model run with initial state errors as in table 1. The maximum error occurs during the spring bloom. Note that the scales are 10 times larger than in figure 3. Similar graphs have also been obtained for the variables phytoplankton and herbivores. Average errors are reported in table 4.

Table 1. Error statistics for DA using the 1D ecological model, variances expressed in (mmol N m^{-3})^2.

Variable	*N*	*P*	*H*
Initial	6.25e−4	1.0e−4	1.0e−4
Model	1.0e−05	1.0e−05	1.0e−05
Top boundary	0	0	0
Bottom boundary	1.0e−05	1.0e−05	1.0e−05
Measurements	6.25e−4	1.0e−4	1.0e−4

the case of nutrient concentration variable. The phytoplankton and herbivores concentrations also severely diverge from the reference case during the spring bloom. Can the RRSQRT KF and the EnKF correct the wrong initial state?

4.2. Data assimilation

In the present case study, the data from the simple ecological model are simulated during a one year reference model run, samples are taken on eight cells on the vertical dimension, they are regularly scheduled throughout the year every ten model steps—i.e. approximately every 10 days—and a Gaussian white noise of constant variance equal to the measurement error variance, see table 1, is added to the samples to simulate measurement variability. Both methods are assimilating the same data and are started from the wrong initial state detailed above.

DA procedures are then compared with respect to their ability to recreate the reference run and to estimate their own error. From this sampling procedure we can notice that both spatial and temporal supports of measurement and model output are the same which is more convenient than in real applications for which they differ.

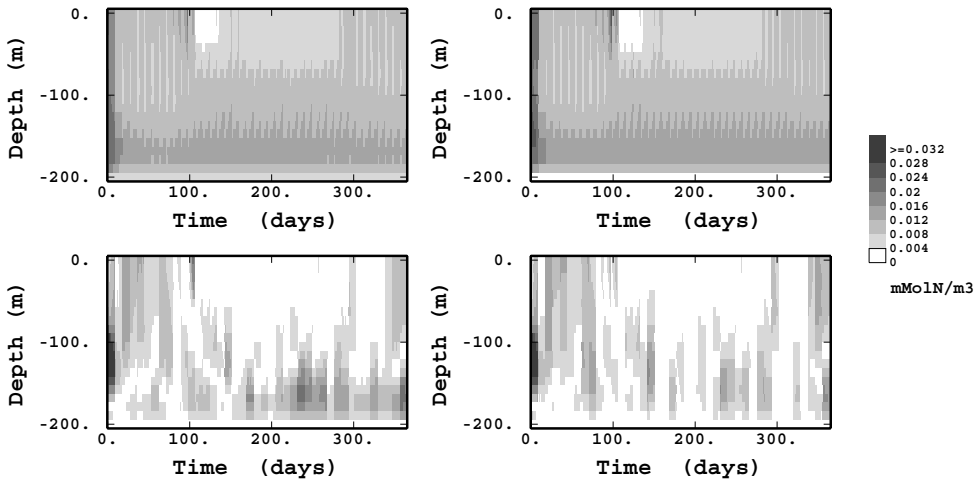

Figure 3. EnKF (left) versus RRSQRT KF (right): SD predicted by the Kalman filter (top) and observed RMS errors computed from the reference solution (bottom) for the nutrients. Both methods give rather similar results, the same observation can also be made for the phytoplankton and herbivores concentrations.

4.3. Comparison of the EnKF and the RRSQRT KF

4.3.1. Error statistics. In the present simulated case, all error statistics have been kept identical as in [12], they realistically reproduce the features of ecological problems. The measurement error is about 10% of the mean value, and the model is also assumed erroneous, although in the present test case the reference solution has been generated by this same model. The model error variance is apparently lower than that of the measurements, see table 1, but since data are assimilated every ten model steps, both model and measurement can be considered as equally accurate.

Both methods compared below (RRSQRT and EnKF) use the same statistics. The initial covariance matrix C_0^{a} is that of a smooth Gaussian auto covariance function of range 50 m—one fourth of the total column depth—and no cross-covariance has been set between the three variables, thus assuming that they are initially independent from each other.

The model error—denoted $\boldsymbol{q}_n$ in the previous section—is made from two parts. The first part is an error on the system state—i.e. on all grid cells of the water column for the three variables—with the same spatial covariance as the initial error. The second part is an error on the system forcing—i.e. the variables concentrations imposed at top and bottom of the water column. The model is assumed exact in the top boundary condition but erroneous at the lower boundary. This error stands for a lack of knowledge of the bottom solid to liquid interactions, which are more important in ecological modelling than interactions at the surface boundary.

Measurement error is in both cases a nugget effect—which stands for 'spatial white noise' in geostatistics—since it seems unrealistic that measurement cells measuring different variables or located at different locations have correlated errors. The variance of this measurement error has also been used for generating the measurements by adding noise to the reference state. Three experiments using different seeds for the random number generator and different realizations of the random measurement errors have been carried out, the results in table 2 show that the influence of these random numbers is not significant. The variances of the errors are given in table 1.

Table 2. RMS errors in mmol N m^{-3} for DA runs done with three different realizations of measurements and three different random seeds (left: EnKF, right: RRSQRT).

	EnKF				RRSQRT			
Trial	*N*	*P*	*H*	Total	*N*	*P*	*H*	Total
n. 1	0.007 35	0.007 30	0.004 44	0.011 27	0.006 53	0.003 98	0.002 29	0.007 98
n. 2	0.007 38	0.006 26	0.003 89	0.010 43	0.006 70	0.003 29	0.002 05	0.007 74
n. 3	0.007 04	0.007 27	0.004 47	0.011 06	0.006 74	0.004 18	0.002 54	0.008 33

In most multivariate DA experiments using the RRSQRT, it is necessary to normalize the error covariance before the eigen decomposition in order to avoid cutting off the modes associated with the variables that have lower absolute values or that are expressed in larger units. As all three variables have similar magnitudes here, this normalization has not been done.

4.3.2. Stability. The RRSQRT KF has instability problems arising from the model linearization in equation (22). The linearization ϵ is supposed to be small for a better precision of the tangent linear model but for values too small sudden bursts in the estimation variance are observed and lead to filter divergence. These instabilities might be caused by a strong local nonlinearity and numerical instability. They are usually countered by tuning the linearization parameter ϵ. In the present case, the RRSQRT KF running with values of ϵ inferior to 1 had divergent solutions with sudden error growths, on the other hand, all runs using values of ϵ superior or equal to 1 remained stable and reproduced equal results on short tuning runs. However the bursts of estimation variance remain unpredictable and a higher value of $\epsilon = 7$ has been retained for more security so that the problem did not show up in this application.

4.3.3. Discussion. The state dimension is 60 in this case study (three variables times 20 vertical nodes), for the RRSQRT Kalman filter the number of eigenmodes has been arbitrarily set to 36 so that the reduction is not too dramatic. In spite of this reduction, the RRSQRT will propagate more than 36 times the model: for the covariance propagation step in equation (15), the RRSQRT runs 60 additional vectors accounting for model error in all vertical nodes for the three variables and also three more vectors accounting for the bottom boundary condition errors. Together with the propagation of the estimated state vector (8) this sums up to 100 forward integrations needed at every timestep.

In order to make the EnKF similarly efficient, 1000 ensemble members have been run, which is much more than in usual EnKF applications.

From the point of view of computation times the RRSQRT propagates 10 times fewer state vectors but contrarily to the EnKF the RRSQRT KF performs eigen decompositions of a 99×99 matrix at each model step for rank reduction, which is also time consuming. The RRSQRT is then approximately five times faster than the EnKF. This has been discussed in detail in Cañizares [4] for the case of a 2D coastal hydrodynamical model.

Figure 3 shows on the top graphs the estimation standard deviation (SD) given by the EnKF (on the left) and the RRSQRT (on the right), and on the bottom graphs the associated true root mean square (RMS) errors between the estimation and the 'true' reference solution. The single nutrient variable is shown but the two other variables, phytoplankton and herbivores, can be commented on in a similar way.

At first sight, both methods seem rather equivalent, the estimated error SD overestimates the actual RMS error on the whole water column during the whole year run. If we look at the

Table 3. SD in mmol N m^{-3} for DA runs done with three different realizations of measurements and three different random seeds (left: EnKF, right: RRSQRT).

	EnKF				RRSQRT			
Trial	*N*	*P*	*H*	Total	*N*	*P*	*H*	Total
n. 1	0.010 12	0.008 88	0.005 88	0.014 69	0.010 54	0.010 67	0.006 90	0.016 51
n. 2	0.010 15	0.008 65	0.005 76	0.014 53	0.010 54	0.010 84	0.006 95	0.016 64
n. 3	0.010 07	0.008 91	0.005 92	0.014 69	0.010 54	0.010 85	0.006 97	0.016 65

Table 4. RMS errors in mmol N m^{-3} of the model run from an erroneous initial state without DA.

N	*P*	*H*	Total
0.116 31	0.349 32	0.109 33	0.384 06

average SD and RMS values in tables 2 and 3, the EnKF and the RRSQRT both overestimate their actual RMS error for all N, P and H variables. This indicates that the error modelling is efficient—which is relatively easy in synthetic cases since the mechanisms that produced the errors are known—and also validates the choices of the number of eigenmodes in the RRSQRT and of the ensemble size in the EnKF. If these numbers were too small, then both DA schemes would model an error subspace of dimension too small and would not describe accurately enough the error. This would result in an underestimation of the RMS error as can be seen in [4].

By comparison with the 'wrong' model run without DA in figure 2, the errors due to a modification of the initial state are significantly reduced by DA. This can also be stated for other variables and using three different realizations of the measurement set with different random error, see tables 2 and 4.

In figure 3 vertical lines appear in the SD plots on the top. These are the traces of the two steps of the KF: first the propagation step increases the forecast error variance with model error and then the measurement step reduces the analysed error variance. This can be physically understood since the measurements correct the inaccurate forecast state, the estimated uncertainty is reduced.

We can also observe that the errors in the nutrient concentrations are lower in the upper half of the water column from days 100 to 300 (from the beginning of April to the end of October). This part of the water column and time period are those of the maximum biological activities, excited by the solar irradiation. During the rest of the year and on the bottom of the water column, the errors are more or less similar to the model error as specified in table 1. This shows that the model dynamics play a significant role in the evolution of the error in the Kalman filter.

A more careful look to the bottom graphs of figure 3 indicates that the RMS errors are lower in the RRSQRT assimilation than in the EnKF. This observation is confirmed in table 2 for all variables and is surprising since there is no theoretical reason why the EnKF running with a large ensemble like this should perform less well than the RRSQRT. This result is also not consistent with those from Cañizares [4] where both methods were equivalent.

An increase of the number of ensemble members should put the EnKF on the same level of performance as the RRSQRT, but at the expense of computational efficiency. Surprisingly we noticed that decreasing the EnKF ensemble size to 100 or even 36 members only weakly deteriorated the results, thus indicating that the size of the ensemble is large enough and that it is not the most sensitive parameter in this case. This observation may only be valid for low-dimensional systems as in the present synthetic test case.

A closer look at the average SD values in table 3 also shows that the RRSQRT produces higher error SD than the EnKF even if the RMS errors are smaller. There is also no theoretical explanation for this difference and the amplitude of this difference is a little too low to be significant.

These two small differences between EnKF and RRSQRT may also depend on the system and cannot be generalized to other DA situations. As a conclusion to the present comparison, the two methods are equal for performing DA on the present 1D ecological model.

5. Case of the Odra lagoon hydrodynamics

5.1. The model

The TRIM3D model is a 3D numerical model for hydrodynamics [7]. It solves the Navier–Stokes equations for free-surface flows under the hydrostatic hypothesis. A semi-implicit finite difference scheme is carried out on an Arakawa grid with a model timestep of 5 min; see [34] and [42] for numerical simulations of the Odra lagoon using TRIM3D.

In the following case study, the rectangular grid has a 250 m horizontal resolution and spreads on 357 $\times$ 259 square nodes among which 16 053 are within the Odra lagoon. Three vertical layers—two with depth 3 m and the third one down to the bottom—are considered which finally results in 24 455 active nodes.

The size of the state and the computational burden could be reduced by increasing the grid cell size or the model timestep, but the model would then lose the ability to describe small scale processes that are significant in the Odra lagoon hydrodynamics. In this application a size reduction occurs in the course of the KF propagation and on statistical criteria so that the neglected processes are those inducing the least possible variations, whatever the scale length.

5.2. Initial and boundary conditions

As no global information about the full state is available, the initial state of the model is such that all hydrodynamical variables—water levels, horizontal and vertical velocities—are set to zero. This means that the Odra lagoon is initially a still water body and the system forcing alone set it into motion. The model is supposed to recover by itself from this lack of knowledge. There are three boundary conditions forcing the lagoon hydrodynamics:

- The Odra river discharge is measured at a station upstream. Discharge values are given every 4 hours and linearly interpolated at every model step. This is sufficient since the discharge values are slowly varying.
- The wind stress is a very sensitive parameter for shallow water systems. Hourly averaged measures are supplied in the pile station Odh1 and since the wind field is almost homogeneous in this region, the single measure is used for the whole Odra lagoon.
- The water elevation at the Baltic sea interface is also a very sensitive forcing of the Odra lagoon hydrodynamics since they control the flux directions in the three channels leading to the Baltic. Unfortunately the water levels are only measured in one station Koserów on the Baltic coast, see the map in figure 1, and this single measure is assigned to the three interfaces. These measurements are also hourly averages.

5.3. Data assimilation

5.3.1. Error statistics. Implementing a Kalman filter requires the knowledge of both observation and model error covariance matrices Σ_o and Σ_m. However we do not have much

knowledge of the error incurred in the measurements and in the complex data processing. There is also no objective knowledge of the model error, as always in real cases, because the true state and dynamics of the system are unknown.

Observation errors can be considered as spatial white noise because they depict a lack of accuracy in the measurement protocol and are seldom correlated from one device to another. The variance of this white noise has been set empirically to the unstructured part of the variability of the measurements time series—which is called the nugget effect part of the variogram in geostatistical terminology—i.e. the SD is around 3 mm of water level, which seems physically satisfying for an hourly average of water levels.

Rosenthal *et al* [34] used TRIM3D in the Odra lagoon and demonstrated that the dynamics are accurately described by the model, therefore the main error in the modelling is assumed to come from the determination of boundary conditions.

The water levels at the outlets to the Baltic sea are given by a single station although these outlets are distant from 70 km, this should be the main source of model error. Another—minor—source of error is the unknown wind field which is supposed homogeneous all over the domain. This second assumption is more consistent with the observations of the Odra lagoon but can remain a source of error since for shallow water hydrodynamics the influence of the wind direction is very strong.

The model errors are modelled as an uncorrelated noise in the wind field (of amplitude 0.5 m s^{-1}) and a correlated noise in the Baltic interface water levels. Since the data available are spatially poor—only five stations on the whole domain—but temporally very rich, an assumption was made by Wolf *et al* [42] that the error at the boundary was propagated according to a wave equation. Statistical analysis of the measurement timeseries confirmed this assumption and gave credibility to the computation of the wave speed. Then the temporal covariance structures computed could be converted to spatial structures, inaccessible by data analysis. The eventual spatial covariance of the water level error at the boundary appeared very flat, which is consistent with the physics of the system since the Odra lagoon water levels show very little variations over its whole area, and has an amplitude of 1.5 cm.

5.3.2. Observations. The assimilated observations are water levels measured as the magnetic signature of a 20 cm floating ball kept in a tube. Hourly measurements are sampled in three pile stations near the coast. Two stations located in the 'Kleines Haff' are kept for validation, see figure 1, their measurements are quasi-continuously transmitted by satellite to the GKSS. Data calibration and 'cleaning' are rather complex and we will not attempt to describe this protocol in the present paper.

Considering the temporal support, hourly averages of the water level measurements were satisfactory since they presented the same statistical characteristics as the model time series with a model time step of 5 minutes. Then for the spatial support the quasi-punctual measurements can reasonably be taken as representative for a whole grid cell of size 250×250 m^2 because water levels present very smooth spatial variations where data is available—even during the flood period of August 1997. In the case of variables having more spatial variations like nutrients, currents or salt, the support effect may have a greater impact on DA and should be introduced in the measurement error model.

5.3.3. Results. Here the RRSQRT has been applied independently from the comparison done in section 4. The number of retained eigenmodes is set to 50 by considering the shape of the eigenvalues diagram. As in the previous case study the linearization parameter ϵ has been tuned on different dynamical situations: during the flood period and during a more quiet

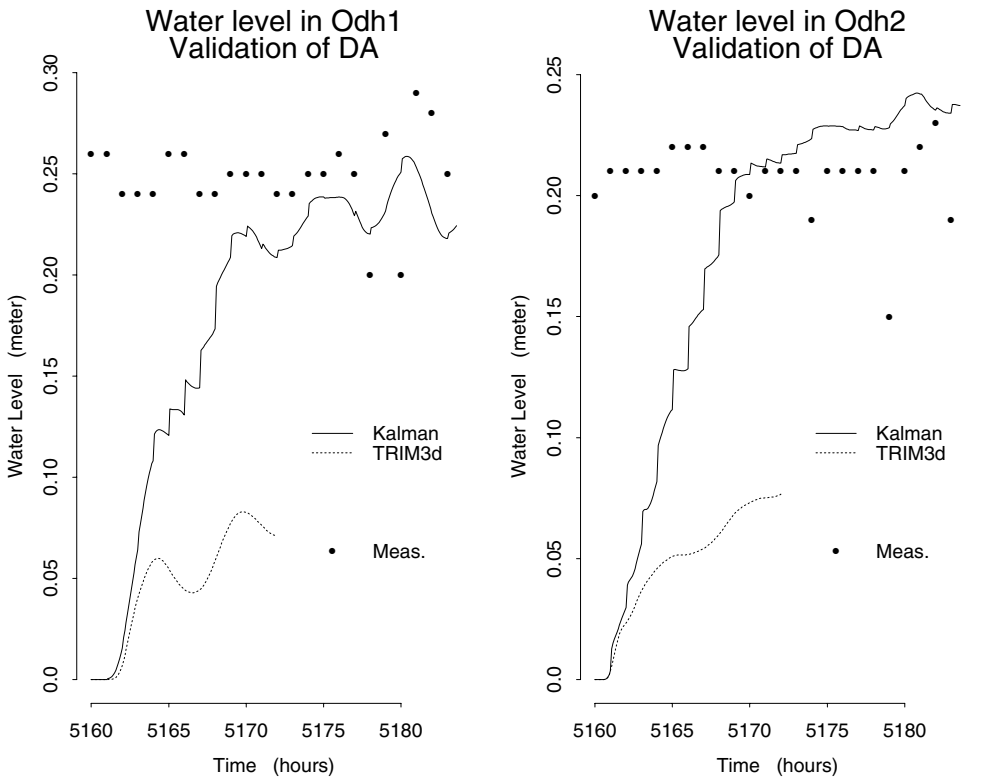

Figure 4. Water levels in the stations Odh1 (left) and Odh2 (right) on the 4th August 1997, dots are the measurements and the full curve is the result of DA of the measurements of water levels from the stations Ückermünde, Karnin and Wolgast and the dotted curve is the result of the TRIM3D model without assimilation driven by the boundary conditions. DA is efficient for correcting the effect of an arbitrary initial condition.

period, and it has eventually been set to 1. The RRSQRT filter was unstable for values of ϵ inferior to 0.1 and stable for the values 1 and 10.

The RRSQRT KF simulates the period from the 4th of August 1997 to the 14th of August 1997. The computation took 5 days on a Sun Sparc machine and further reduction of the computation times can be achieved by tuning the number of eigenmodes. Therefore the RRSQRT can provide forecasts of the hydrodynamical state of the Odra lagoon starting from an existing DA run and propagating forward in time the state and RRSQRT covariance matrix as an expectation of the future state and an estimate of the forecast error. Forecast of the boundary condition values also have to be supplied for this purpose.

The impact of an erroneous initial condition can also be reduced by the use of DA. The time series in figure 4 show the water levels in the two measurement stations kept aside for validation at the beginning of the DA run. The model is initialized by uniform zero water levels but the measurements indicate water levels of around 20 cm. As the dynamical model driven by the boundary conditions needs more than 5 days to initialize, one day is enough for the KF estimate to reach the measurements in both validation stations. This can be understood as the KF is also a spatial interpolation method.

In the course of a 40 days run, the model TRIM3D slightly underestimates the water levels in Odh1 from 3 cm and produces RMS errors of amplitude 4 cm, see table 5. This shows that

Table 5. Mean errors, RMS errors and total errors expressed in m from different estimators of water levels in the validation station Odh1: the crude model TRIM3D without DA, the average of the assimilated water levels measured in three stations and the RRSQRT KF assimilated estimations. Statistics are computed on 40 days, i.e. 961 samples, from the 4th August to the 13th of September 1997.

	Bias	RMS	Total
TRIM3d	+0.035	0.041	0.054
Samp. mean	+0.018	0.026	0.032
RRSQRT KF	+0.005	0.027	0.027

Table 6. Mean errors, RMS errors and total errors expressed in *m* from different estimators of water levels in the validation station Odh2: the crude model TRIM3D without DA, the average of the assimilated water levels measured in three stations and the RRSQRT KF assimilated estimations. Statistics are computed on 40 days, i.e. 961 samples, from the 4th August to the 13th of September 1997. It should be noted that an extract without the 4th of August of present DA run was also used for calibration of the measurements in Odh2.

	Bias	RMS	Total
TRIM3d	+0.031	0.042	0.052
Samp. mean	+0.018	0.045	0.047
RRSQRT KF	+0.002	0.031	0.031

the model is already accurate enough for a correct hydrodynamical description of the lagoon, however the use of DA can subsequently reduce the model bias by 86% and the RMS error by 34%, thus enabling higher accuracy applications such as water constituent transport balance and ecological modelling. In station Odh2, the water level calibration could not be performed by physical means and the offset of the water level timeseries was estimated using the DA analysis timeseries from the 2nd day to the 40th day to skip the initialization. Therefore DA naturally corrects the model bias in Odh2 and the remaining 2 mm bias reflects the effects of DA initialization, see table 6. By comparison with another coastal hydrodynamics application working with comparable measurement frequency, spatial coverage and model grid cell size, see Cañizares *et al* [5], both model and assimilated estimates of water levels are more accurate in the present case, which is probably due to the unusual geometry of the Odra lagoon: an almost closed basin with only three narrow open boundaries. In other estuarine systems with wider open boundaries, the DA errors may be larger, even in similar model and observations conditions.

To assess the DA method, the arithmetic mean of the assimilated measurements has also been tested as an estimator. This sample mean is a better estimator than the crude TRIM3D model in the case of the station Odh1 since this station is located in the middle of the assimilation stations, but the estimation is poorer in the case of the station Odh2 that is outside of this region. Moreover, the arithmetic average of the stations measurements is also a biased estimator for the two validation stations, see tables 5 and 6.

It should be noted that the above comparison of validation measurements with model values or DA values implicitly assumes that the values are comparable from the point of view of their spatio-temporal support. As stated in the beginning of the section, this is possible in the case of water levels in the Odra lagoon since spatial variations are very smooth and quasi-punctual measurements can be considered representative for a larger volume.

If we look now for spatial features of the assimilated water levels on the assimilated water levels map in figure 5, the RRSQRT Kalman filter has realistic smooth state estimates—around 10 cm water level differences from the East coast to the West coast of the basin—corrections

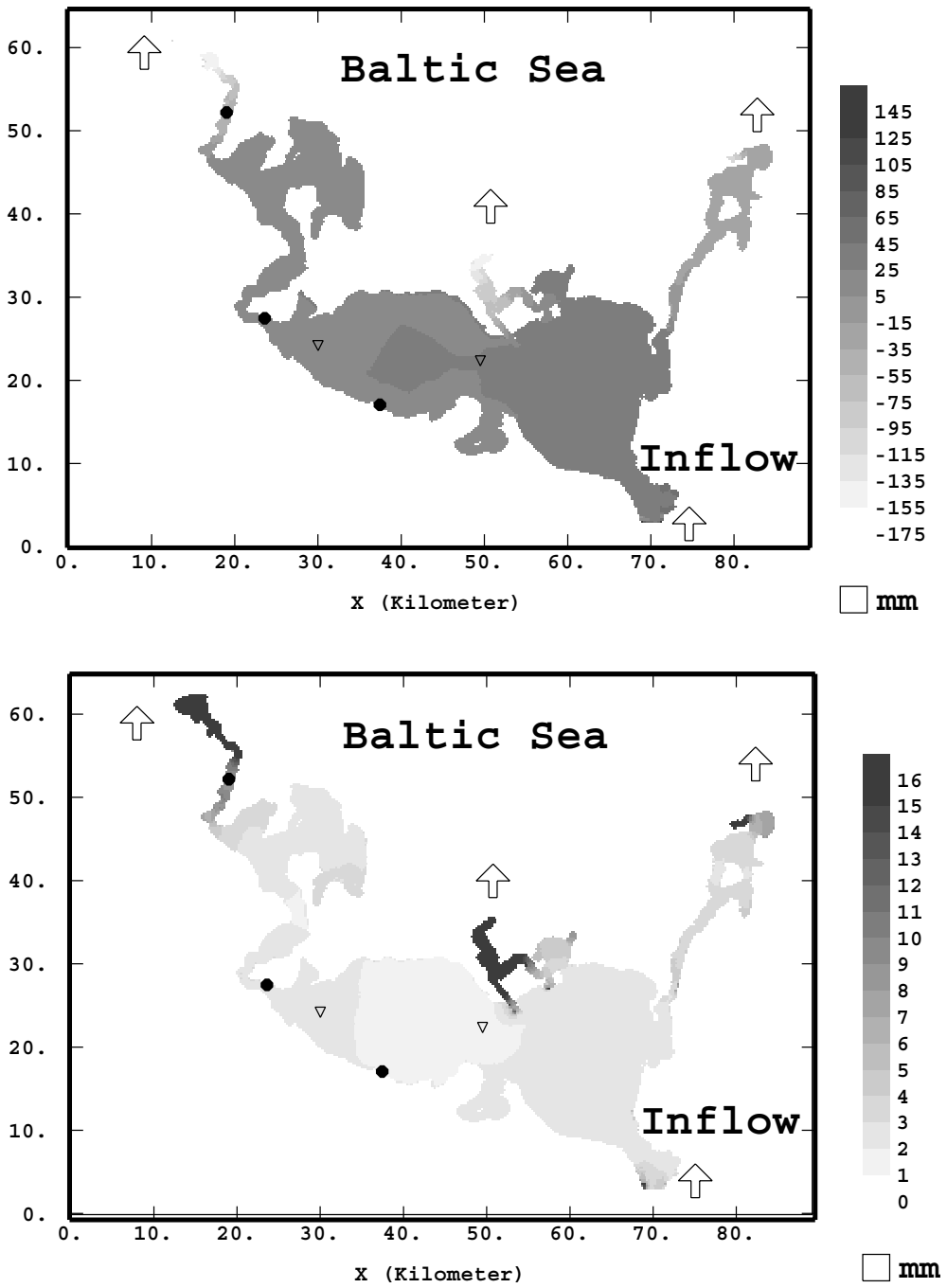

Figure 5. DA by RRSQRT KF: maps of the water level (top) assimilated and SD (bottom) of the assimilation after 10 days and 240 DA steps. The Odra inflow comes at the bottom right of the graph and the three channels to the Baltic sea are on the top left middle and right of the domain. Measurement stations are indicated by a ●, validation stations by a ▽.

of the measurement step are performed on the whole basin and not only in the neighbourhood of the measurement stations.

The error SD map in figure 5 shows that the error expected by the RRSQRT KF is reasonable (mostly below one centimeter SD) and remains stable. It also shows that the system noise has maximum variance—around 1.6 cm—near the coastal boundaries and this can be mainly explained by the error of amplitude 1.5 cm that is introduced in the sea boundary water levels every model step. However the KF propagates the error covariance and it is not clear whether the estimated error at the center of the basin is due to the error in the sea boundary water levels or the error in the wind field.

This indicates a method for improving the model observing network: by successively switching off each of the two sources the error, SD estimation can quantify how much one dominates the other at any location of the domain. Then knowing where the estimation should be most accurate in the application, this method indicates which further efforts in the development of the observation network should be done.

6. Conclusions

Coupling a KF to a numerical model of an estuary can make it account for uncertainties in the system, measurements and boundary conditions. The filter gives a least squares estimate of the true state of the system, but also provides an analysis of the estimation error.

When applied to high-dimensional systems driven by nonlinear equations such as the hydrodynamics of the Odra lagoon, KF is highly demanding in CPU time as well as disk space and leads to numerical problems.

In the present paper, two suboptimal DA schemes—the RRSQRT KF and EnKF—were efficiently applied to a 1D ecological model that is very sensitive to uncertainties in the initial state. The comparison of their results did not show significant differences between both schemes and proved that the accuracy of the DA results is dependent on the accuracy of the model dynamics. The RRSQRT KF was also implemented in a 3D hydrodynamical model of the Odra lagoon. Although the lagoon looks almost closed, the main source of model error has been identified as the poor knowledge of the three water level boundary conditions in the Baltic sea. The RRSQRT KF corrected this error and the initialization error with high accuracy.

The DA schemes presented here have overcome two major problems of estuarine applications. The first one is the spatial extent of the system that leads to high-dimensional state models. The second one is the nonlinearity of the dynamical models. The two suboptimal KF schemes provide empirically efficient solutions for estuarine systems. Their success in this application is a sign that their operational use is a realistic perspective, and that the methodology can be applied to other physical, chemical or biological application fields.

The inputs of geostatistics in DA methods are in spatial modelling of model and measurement errors but also in accounting for support effects while comparing *in situ* measurements with model forecasts.

Acknowledgments

This work was carried out within the EC funded MAST III project PIONEER (description: http://pioneer.geogr.ku.dk) and has been partly supported by a grant of computer time from the Norwegian Supercomputing Committee (TRU).

References

[1] Bennett A F 1992 *Inverse Methods in Physical Oceanography* (Cambridge: Cambridge University Press)

[2] Bennett A F, Chua B S and Leslie L M 1996 Generalized inversion of a global numerical weather prediction model *Meteorol. Atmos. Phys.* **60** 165–78

[3] Brasseur P, Ballabrera-Poy J and Verron J 1999 Assimilation of altimetric data in the mid-latitude oceans using the singular evolutive extended Kalman filter with an eddy-resolving, primitive equation model *J. Mar. Syst.* **22** 269–94

[4] Cañizares R 1999 On the application of data assimilation in regional coastal models *PhD Thesis* TU Delft, Delft, The Netherlands (Rotterdam: Balkema)

[5] Cañizares R, Madsen D, Jensen H and Vested H J 2000 Developments in operational shelf sea modelling in danish waters *Estuarine Coastal Shelf Sci. JONSMOD'98* special issue

[6] Cane M, Kaplan A, Miller R N, Tang B, Hackert E and Busalacchi A 1996 Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter *J. Geophys. Res.* **101** 22599–617

[7] Casulli V and Cattani E 1994 Stability, accuracy and efficiency of a semi-implicit method for three-dimensional shallow water flow *Comput. Math. Appl.* **27** 99–112

[8] Chilès J P and Delfiner P 1999 *Geostatistics: Modeling Spatial Uncertainty Wiley Series in Probability and Statistics* (New York: Wiley)

[9] Cohn S and Todling R 1996 Appropriate data assimilation schemes for stable and unstable dynamics *J. Meteorol. Soc. Japan* **74** 63–75

[10] Daley R 1991 *Atmospheric Data Analysis* (Cambridge: Cambridge University Press)

[11] Dee D P 1991 Simplification of the Kalman filter for meteorological data assimilation *Q. J. R. Meteorol. Soc.* **117** 365–84

[12] Eknes M and Evensen G 1999 An ensemble Kalman filter with a 1D marine ecosystem model *J. Mar. Syst.* submitted

[13] Evans G and Parslow J 1985 A model for annual plankton cycles *Biol. Oceanogr.* **3** 327–47

[14] Evensen G 1992 Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model *J. Geophys. Res.* **97** 17905–24

[15] Evensen G 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics *J. Geophys. Res.* **99** 10143–62

[16] Evensen G and van Leeuwen P J 2000 An ensemble Kalman smoother for nonlinear dynamics *Mon. Weather Rev.* **128** 1852–67

[17] Fukumori I 1995 Assimilation of Topex sea level measurements with a reduced-gravity, shallow water model of the tropical pacific ocean *J. Geophys. Res.* **100** 25027–39

[18] Fukumori I and Malanotte-Rizzoli P 1995 An appropriate Kalman filter for ocean data assimilation: an example with an idealized Gulf stream model *J. Geophys. Res.* **100** 6777–93

[19] Gauthier P, Courtier P and Moll P 1993 Assimilation of simulated wind lidar data with a Kalman filter *Mon. Weather Rev.* **121** 1803–20

[20] Ghil M and Malanotte-Rizzoli P 1991 Data assimilation in meteorology and oceanography *Adv. Geophys.* **33** 141–266

[21] Heemink A and Kloosterhuis H 1990 Data assimilation for nonlinear tidal models *Int. J. Numer. Methods Fluids* **11** 1097–112

[22] Jazwinski A A 1970 Stochastic and filtering theory *Mathematics in Sciences and Engineering* vol 64 (New York: Academic)

[23] le Dimet F X and Talagrand O 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects *Tellus* A **38** 97–110

[24] Luong B, Blum J and Verron J 1998 A variational method for the resolution of a data assimilation problem in oceanography *Inverse Problems* **14** 979–97

[25] Mardia K, Goodall C, Redfern E and Alonso F 1998 The kriged Kalman filter *Test* **7** 217–85

[26] Maybeck P 1979 Stochastic models, estimation, and control *Mathematics in Science and Engineering* vol 141–1 (New York: Academic)

[27] Miller R N, Carter E F Jr and Blue S T 1999 Data assimilation into nonlinear stochastic models *Tellus* A **51** 167–94

[28] Miller R N, Ghil M and Gauthiez F 1994 Advanced data assimilation in strongly nonlinear dynamical systems *J. Atmos. Sci.* **51** 1037–56

[29] Nechaev V and Yaremchuk M 1994 Applications of the adjoint technique to processing of a standard section data set: world ocean circulation experiment section S4 along 67 deg S in the Pacific ocean *J. Geophys. Res.* **100** 875–79

[30] Ngodock H E, Chua B S and Bennett A F 2000 Generalized inverse of a reduced gravity primitive equation model and tropical atmosphere-ocean data *Mon. Weather Rev.* **128** 1757–77
[31] Øksendal B 1992 *Stochastic Differential Equations, an Introduction with Applications* (Berlin: Springer)
[32] Pham D, Verron J and Roubaud M 1998 A singular evolutive extended Kalman filter for data assimilation in oceanography *J. Mar. Syst.* **16** 323–40
[33] Robinson A R, Lermusiaux P F J and Sloan N Q III 1998 Data assimilation *The Global Coastal Ocean. Processes and Methods. The Sea* vol 10, ed A R K H Brink (New York: Wiley) pp 541–93
[34] Rosenthal W, Wolf T, Witte G, Buchholz W and Rybaczok P 1998 Measured and modelled water transport in the Odra estuary for the flood period july/august 1997 *German J. Hydrogr.* **50** 215–30
[35] Thacker W and Long R 1988 Fitting dynamics to data *J. Geophys. Res.* **93** 1227–40
[36] Verlaan M 1998 Efficient Kalman filtering algorithms for hydrodynamic models *PhD Thesis* TU Delft, Delft, The Netherlands
[37] Verlaan M and Heemink A W 1997 Tidal flow forecasting using reduced rank square root filters *Stoch. Hydrol. Hydraulics* **11** 349–68
[38] Verlaan M and Heemink A W 1999 Nonlinearity in data assimilation applications: a practical method for analysis *Mon. Weather Rev.* submitted
[39] Verron J, Gourdeau L, Pham D, Murtugudde R and Busalacchi A 1998 An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific: method and validation *J. Geophys. Res.* submitted
[40] Wackernagel H 1998 *Multivariate Geostatistics* 2nd edn (Berlin: Springer)
[41] Wikle C K and Cressie N 1999 A dimension-reduced approach to space-time Kalman filtering *Biometrika* **86** 815–29
[42] Wolf T, Sénégas J, Bertino L and Wackernagel H 2001 Application of data assimilation to three-dimensional hydrodynamics: the case of the Odra lagoon *GeoENV III: Geostatistics for Environmental Applications* ed Monestiez, Allard and Froidevaux (Amsterdam: Kluwer) at press