Efficient Kalman filter techniques for the assimilation of tide gauge data in three-dimensional modeling of the North Sea and Baltic Sea system

J. V. T. Sørensen and H. Madsen
DHI Water & Environment, Hørsholm, Denmark

H. Madsen
Informatics and Mathematical Modeling, Technical University of Denmark, Kongens Lyngby, Denmark

Received 25 September 2003; revised 27 December 2003; accepted 22 January 2004; published 10 March 2004.

[1] Data assimilation in operational forecasting systems is a discipline undergoing rapid development. Despite the ever increasing computational resources, it requires efficient as well as robust assimilation schemes to support online prediction products. The parameter considered for assimilation here is water levels from tide gauge stations. The assimilation approach is Kalman filter based and examines the combination of the Ensemble Kalman Filter with spatial and dynamic regularization techniques. Further, both a Steady Kalman gain approximation and a dynamically evolving Kalman gain are considered. The estimation skill of the various assimilation schemes is assessed in a 4-week hindcast experiment using a setup of an operational model in the North Sea and Baltic Sea system. The computationally efficient dynamic regularization works very well and is to be encouraged for water level nowcasts. Distance regularization gives much improved results in data sparse areas, while maintaining performance in areas with a denser distribution of tide gauges. *INDEX TERMS:* 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; *KEYWORDS:* tide gauge, data assimilation, North Sea, Baltic Sea

Citation: Sørensen, J. V. T., H. Madsen, and H. Madsen (2004), Efficient Kalman filter techniques for the assimilation of tide gauge data in three-dimensional modeling of the North Sea and Baltic Sea system, *J. Geophys. Res.*, *109*, C03017, doi:10.1029/2003JC002144.

1. Introduction

[2] Marine operational forecasting systems are being increasingly applied for a number of engineering and public service purposes [e.g., *Pinardi and Woods*, 2002; *Erichsen and Rasch*, 2002]. The products are valuable for hindcast, nowcast, and forecast situations, and in all cases, there is a need for higher precision simulation of the physical variables. In order to increase the predictive skill, the numerical models have been continuously improved during the past decades. Better numerical methods have been developed, smaller scales resolved, and improved parameterizations implemented. The developments in the numerical models have been carried over to the operational systems as robustness has been proved. Along with this development, attention has been paid to including an increasing number of physical variables in the models. Hence the portfolio of products has been expanded from the hydrodynamic and thermodynamic parameters to include estimation and prediction of waves, biogeochemical parameters, and sediments. All together, these developments have taken us to the stage we have reached today.

[3] A number of model errors remain despite the clear improvements of the predictive skill that the present operational systems have experienced during their lifetime. However, the measurements that demonstrate this error are often available online and can potentially be used to update the estimation of the ocean state in real time. Methods that pursue this line of thought are referred to as operational data assimilation techniques. Data assimilation is a cross-disciplinary field with a range of uses, for example, the engineering community and meteorological sciences have a long history of successful applications. Data assimilation in ocean models for hindcast studies has also been rather widespread during the past decade.

[4] However, the methodologies are computationally demanding, and hence the use of assimilation approaches has only been applied to a lesser extent in the operational modeling community. Examples are the MERCATOR project [*Bahurel et al.*, 2002] and the MFSTEP project [*Pinardi et al.*, 2002]. In common for these and similar developments is the accessibility of high-performance computational resources and assimilation of a large range of satellite and in situ measurements into three-dimensional regional or global models. For more widespread application, techniques must be applicable on the moderate computational resources available to project engineers and scientists working in applied modeling. For the assimilation of tide

0148-0227/04/2003JC002144

gauge data, operational storm surge forecasting has been one of the targets. Here smaller geographical areas and simpler two-dimensional models have often been considered, which gives some reduction in the required computational resources. Simultaneously, cheap assimilation methods have also been proven successful, hence encouraging their implementation [e.g., *Vested et al.*, 1995; *Gerritsen et al.*, 1995; *Cañizares et al.*, 2001].

[5] A less computationally demanding assimilation approach based on the steady solution of the Riccati equation and subsequent use in a Kalman filter was suggested by *Heemink and Kloosterhuis* [1990]. This approach reduces the computational demands to the same order of magnitude as a standard model execution. *Verlaan and Heemink* [1997] suggested the improved reduced rank square root (RRSQRT) extended Kalman filter, with successful application for storm surge prediction along the Dutch coast. *Bertino et al.* [2002] similarly used a RRSQRT Kalman filter for water level assimilation in the Odra Lagoon. *Madsen and Cañizares* [1999] demonstrated an implementation of the RRSQRT and the Ensemble Kalman Filter (EnKF) [*Evensen*, 1994] in an idealized bay. They showed that the two schemes have similar computational demands and performance. However, computational times are of the order 10^2 times greater than a standard model execution. *Cañizares et al.* [2001] demonstrated a successful application in the North Sea and Baltic Sea system of a Steady Kalman filter using a gain obtained as a time average of the gain produced by the EnKF. Interestingly, spurious correlations caused the results to get worse in data sparse regions, showing the limitation of the Kalman filter approach. On the basis of ideas from *Houtekamer and Mitchell* [1998], *Hamill et al.* [2001] discussed this artifact of the EnKF and suggested that a distance function can be used to control the effect of uncertain ensemble estimates. *Evensen* [2003] argued that such an approach should be avoided because it no longer generates updated ensembles as linear combinations of the forecast ensembles.

[6] The main computational issue in Kalman filter based data assimilation is the propagation of the system error covariance matrix. The EnKF and RRSQRT schemes along with, for example, the SEEK [*Pham et al.*, 1997] and the SEIK filter [*Pham et al.*, 1998], attempt to save computational resources by constructing a low rank approximation of the model error covariance. The Steady filter assumes no time variation, but still requires a solution of the Riccati equation or a more elaborate scheme for the generation of the gain. *Dee* [1991] suggested using a simpler dynamical model propagator for the error propagation. *Fukumori and Melanotte-Rizzoli* [1995] presented a scheme which employed a coarser grid for the error propagation, hence reducing the dimension of the state-space but simultaneously simplifying the dynamics.

[7] The objective of the present study is to investigate the possibility of combining a range of approximate Kalman filter based techniques for the assimilation of tidal gauge data in the North Sea and Baltic Sea system. The techniques are selected in order to provide an optimally efficient scheme for this case, but their nature is discussed in a general regularization perspective. This framework acknowledges the violation of the underlying assumption in the elaborate assimilation schemes and enables the incorporation of prior independent knowledge in the estimation of the ocean state.

[8] Within the regularization framework we describe four approximations to the EnKF. These are temporal smoothing of the Kalman gain, the Steady Kalman gain, a barotropic model error approximation, and a distance dependence of the Kalman gain. The performance of the techniques is presently examined in a hindcast scenario of the North Sea and Baltic Sea system, but the goal is to develop schemes that can be used in an operational forecast setting.

[9] In section 2 the two- and three-dimensional hydrodynamic models employed in this study are presented along with the available tide gauge measurements. Section 3 provides the theoretical basis of the assimilation approaches considered. This encompasses a general discussion of the estimation technique used in Kalman filtering along with a discussion of model and measurement uncertainties. The EnKF is also described in section 3, as are four regularization techniques leading to a Kalman gain smoothing and a Steady Kalman filter as well as a barotropic and a distance regularization. Section 4 presents the design of the numerical experiments. The results are shown and discussed in section 5, while section 6 concludes the paper. The nomenclature suggested by *Ide et al.* [1997] is followed throughout this work, where applicable.

2. Description of Models and Measurements

[10] The Water Forecast is an operational forecasting system covering a large part of the North Sea, the Baltic Sea, and the interconnecting waters [*Erichsen and Rasch*, 2002]. The hydrodynamic model has run operationally as part of the Water Forecast service since June 2001. While the system provides 4-day forecasts of hydrodynamic, water quality, and wave parameters every 12 hours, this study restricts attention to water levels in a hindcast setting.

[11] The hydrodynamic model of the forecast system is the three-dimensional MIKE 3 [*DHI*, 2001], which handles free surface flows. It solves the primitive equations making the hydrostatic and the Boussinesq approximations. The turbulence closure scheme adopted is the k-ϵ model in the vertical and Smagorinsky horizontally. The area covered by the model is shown in Figure 1. Tidally varying water levels are prescribed at the two open boundaries, which are situated in the English Channel and in the northern North Sea between Stavanger in Norway and Aberdeen in Scotland. Wind fields and sea surface pressure are derived from the Vejr2 commercial weather service [*Rogers et al.*, 2001], and force the momentum equations at the sea surface. The vertical resolution is 2 m within the top 80 m. Larger depths are contained in the model bottom layer. The model is nested as displayed in Figure 1, and the horizontal resolution varies from 9 nautical miles to 1 nautical mile in the inner Danish waters and one third nautical mile in a few narrow straits. A two-way nesting technique is employed, securing a dynamic exchange of mass and momentum between grids.

[12] The numerical model mentioned above attempts to express the true state of the system in discrete space and time. The model space is spanned by water level, l, velocity, $\mathbf{v}$, temperature, T, and salinity, S, averaged over spatial volumes at discrete times. Let $\mathbf{x}_{M3}(t_{i-1}) \in \mathbf{R}^{n_{M3}}$ be the model estimate of the true state at time t_{i-1}. Hence the one-

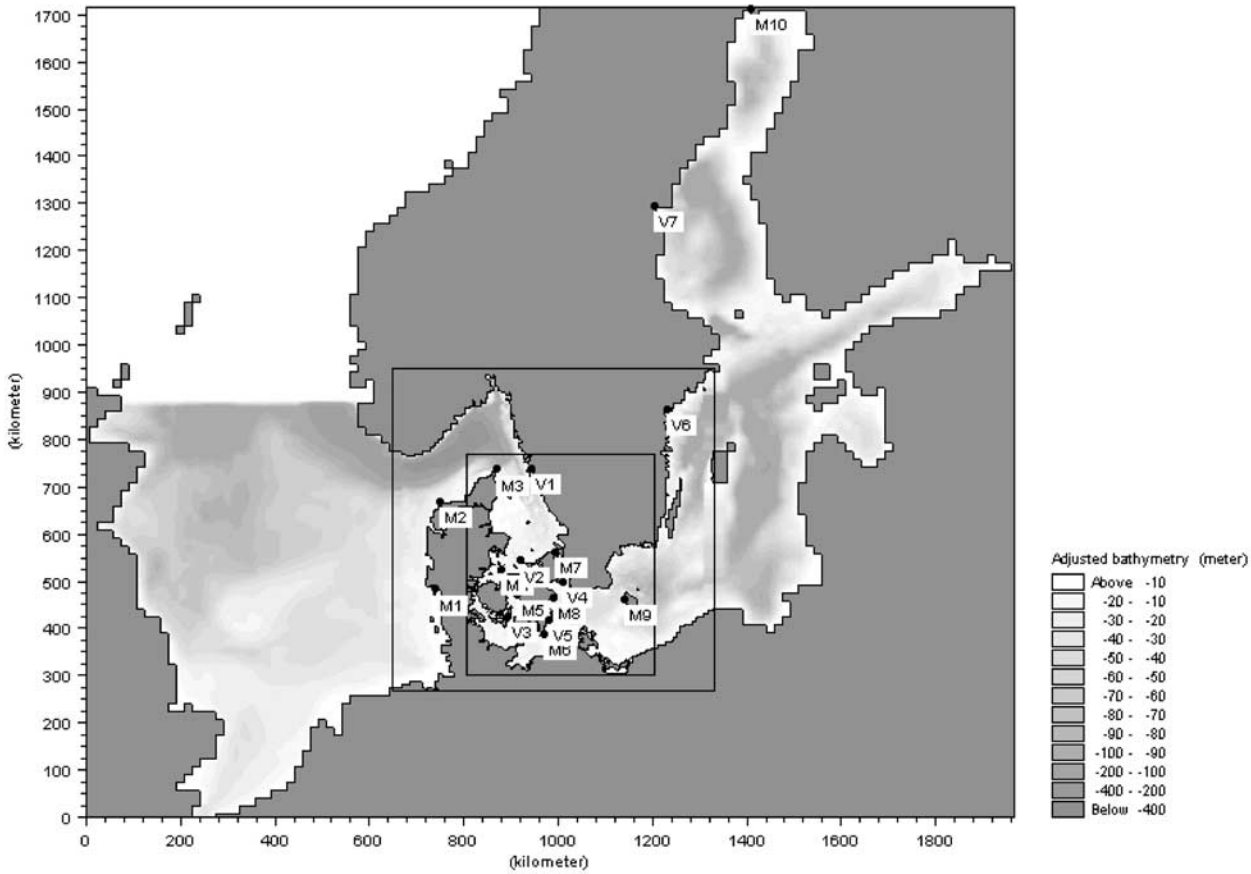

Figure 1. Bathymetry and available tide gauge stations, including 10 measurement stations (M1–M10) and seven validation stations (V1–V7).

time step-ahead model operation, M_{M3}, provides the model estimate at time t_i as

$$\mathbf{x}_{M3}(t_i) = M_{M3}(\mathbf{x}_{M3}(t_{i-1}), \mathbf{u}(t_{i-1})), \tag{1}$$

where $\mathbf{u}(t_{i-1})$ is a vector containing the model forcing.

[13] For the purpose of this study the barotropic two-dimensional model MIKE 21 *DHI* [2002] was set up on the same horizontal grid and with the same forcing. The model has a smaller state-space dimension and simpler dynamics, excluding density variations and collapsing the vertical axis. Only water level, l, and depth-averaged velocities, $\mathbf{V}$, enters the state, $\mathbf{x}_{M21}(t_{i-1}) \in \mathbf{R}^{n_{M21}}$. The model propagator, M_{M21}, provides the model estimate at time t_i as

$$\mathbf{x}_{M21}(t_i) = M_{M21}(\mathbf{x}_{M21}(t_{i-1}), \mathbf{u}(t_{i-1})). \tag{2}$$

Both in terms of state-space dimension and execution times, the barotropic model is significantly cheaper.

[14] For the purpose of this study, 17 tide gauge stations were selected. These are displayed in Figure 1. All stations are situated in Danish or Swedish waters. The 10 tide gauge stations used for assimilation will be referred to as measurement stations and indicated by an "M." The seven stations used for performance assessment will be referred to as validation stations and indicated by a "V." The stations in each of these groups are numbered consecutively, and the corresponding station names can be read from Figures 2 and 3.

[15] A much better data coverage than what is used in this contribution is needed for improved storm surge predictions in the North Sea. However, the Water Forecast model does not have storm surge modeling as a sole objective. The objective also lies in transports as well as ecosystem parameters, and the aim is to apply a unified consistent model for all purposes. Hence we employ a three-dimensional model, and it would be appropriate to also validate results against the velocity. However, very little representative velocity data have been at our disposal in the considered period, and we follow a more traditional storm surge model validation approach in a restricted area, which should be regarded as only a partial validation for the full purpose of the system. In this respect, it is important that there are validation stations (V6 and V7) far from assimilation stations to examine aspects of the consistency of the employed techniques.

3. Assimilation Approach

[16] The schemes used for the assimilation of water level data in the present study can be categorized as sequential estimation techniques. The theory can basically be divided into two parts. One is an estimation of the true state based on the distributions of the model and measured variables,

respectively. The standard approach is to assume no bias and use the best estimator in a minimal variance sense. This estimator is presented in section 3.1. The other part is a specification and a subsequent propagation of the stochastic model state in between measurement times. The observational error also needs to be quantified. The specification of an error model for the numerical model as well as for the observations is built on a number of assumptions. A discussion of these and a description of the error models employed in the present study are presented in section 3.2. In a dynamical model the model error is continuously altered by the model dynamics, and hence the error description needs to be propagated in time. A Markov Chain Monte Carlo approach is followed here, leading to the Ensemble Kalman Filter (EnKF) described in section 3.3.

[17] The ensemble approach is an efficient way of making the work load of the model error propagation tractable, by reducing the degrees of freedom in the description dramatically. However, the resulting scheme is still too expensive for many operational systems, which are typically pushed close to the limit in terms of computational resources in order to resolve as many processes as possible. Further, the EnKF scheme may introduce spurious correlations in data-sparse regions due to an inaccurate model error description and the stochastic nature of the scheme. Hence, despite risking introducing nondynamical modes in the system, various forms of regularization of the gain is proposed. Section 3.4.1 presents a simple temporal smoothing of the Kalman gain matrix, section 3.4.2 describes a Steady Kalman gain approach, and section 3.4.3 presents a dynamic regularization based on the assumption that the errors are barotropic. Finally, section 3.4.4 describes a distance regularization technique.

3.1. BLUE Estimator

[18] In this section the estimation of the state of the system is under consideration. This is often referred to as the analysis step. The state is essentially a multivariate continuous four-dimensional field. Observations are noisy samples from this field and are typically integral measures over some spatial and temporal scale. Similarly, model variables represent per definition spatial averages of the true state. The spatial and temporal representation of the three-dimensional model is taken as a common reference frame, and the state is described in terms of its projection onto it, $\mathbf{x}^t(t_i) \in \mathbf{R}^n$. Here t_i denotes time indexed by i, and $\mathbf{x}^t(t_i)$ is further restricted to include the model variables, water level, l_i and velocity, $\mathbf{v}_i$, hence excluding temperature and salinity. This approximation is due to later time savings in the EnKF error propagation and is made in order to facilitate the barotropic regularization in section 3.4.3. Next, let the prediction by a numerical model, $\mathbf{x}^f(t_i)$, describe the first moment of the stochastic state and assume its error covariance matrix, $\mathbf{P}^f(t_i)$, to be known.

[19] Now, let the observation at time t_i, $\mathbf{y}_i^o \in \mathbf{R}^p$ be related to $\mathbf{x}^t(t_i)$ through the linear measurementequation,

$$\mathbf{y}_i^o = \mathbf{H}_i \mathbf{x}^t(t_i) + \boldsymbol{\epsilon}_i. \tag{3}$$

The matrix $\mathbf{H}_i \in \mathbf{R}^{p \times n}$ is a linear operator projecting the state representation onto the measurement space, and the measurement noise term $\boldsymbol{\epsilon}_i$ is assumed to be an i.i.d. random process. Assume the first and the second moments of this noise to be known, respectively, $\mathbf{0}$ and $\mathbf{R}_i$.

[20] When information from the true system becomes available in the form of measurements, an improved state estimate can be obtained. One procedure for doing this is to assume a linear combination of the unbiased model prediction and the observation that gives the minimum variance estimate, $\mathbf{x}^a(t_i)$. This approach is called the best linear unbiased estimate (BLUE). A derivation is given by *Jazwinski* [1970], yielding the following estimator:

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}_i\left(\mathbf{y}_i^o - \mathbf{H}_i \mathbf{x}^f(t_i)\right). \tag{4}$$

The Kalman gain matrix, $\mathbf{K}_i \in \mathbf{R}^{n \times p}$, is given by

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^T\left(\mathbf{H}_i \mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}. \tag{5}$$

The error covariance, $\mathbf{P}^a(t_i)$, of $\mathbf{x}^a(t_i)$ will always be less than or equal to $\mathbf{P}^f(t_i)$ and can be calculated as

$$\mathbf{P}^a(t_i) = \mathbf{P}^f(t_i) - \mathbf{K}_i \mathbf{H}_i \mathbf{P}^f(t_i). \tag{6}$$

The set of equations (4) and (5) supplies the variance minimizing analysis among the class of linear equations, and equation (6) supplies the a posteriori error covariance, if $\mathbf{R}_i$ and $\mathbf{P}^f(t_i)$ indeed were the real a priori error covariances.

3.2. Error Descriptions

[21] The optimality of the BLUE estimator for the analysis relies on a correct specification of the model and measurement error covariances. Hence any misspecification of these will make the scheme suboptimal. This section takes a closer look at model and observation errors and how they are quantified in the scheme.

3.2.1. Measurement Error

[22] The error, $\boldsymbol{\epsilon}_i$, in the measurement equation (3) includes both an instrumental error and a representation error and is thus properly referred to as a measurement constraint error, as suggested by *Fukumori et al.* [1999]. Depending on the observation considered, either instrumental or representation error can be the dominating source. The instrumental error refers to the actual error in measurement of the physical property under consideration. Often such statistics can be assessed rather precisely. However, the instruments may be badly calibrated, and electronic or mechanical malfunctioning may induce systematic errors. Hence an elaborate quality check on the data must be performed.

[23] The other contribution to the measurement constraint error is due to the fact that the state estimation is done in the model space. *Fukumori et al.* [1999] provide a nice discussion of this, arguing that the resulting measurement representation error contributes to the measurement constraint error. A measurement typically represents a physical property averaged over a different spatial and temporal scale than the model representation. As an example, the measured water level in a corner of a harbor needs not be representative of the water level averaged over an area of 1×1 km^2. If we had retained the continuous reality as the space in which we estimate the state, then the spatial discretization of the model should have been described as a model error.

However, the adoption of the discrete model space as the state-space moves the error to the measurement equation, now expressing that the measurement only approximately represents the spatial average adopted by the model.

[24] Similarly, observed signals that are caused by processes not included in the numerical model can be described as a representation error. The only difference is that we now consider the dynamical subspace spanned by the model rather than the spatial subspace. It must be stressed that the adoption of the model space as the state-space is a choice. We could have chosen another projection of reality, but important in either case is that the error process description is formulated according to this choice.

[25] When using the state definition discussed above, a white noise error does not provide a good description of the expected measurement error, and hence the entire premise of the Kalman filter is violated. The representation error must be expected to be colored and it should be described as such. This can be done by augmenting the state-space by its colored components of a suitable measurement error model. This is, however, a difficult task, and the rather crude approximation of merely increasing the white measurement error standard deviation is taken here, as suggested by *Fukumori et al.* [1999].

[26] The measurement error is usually given some predefined value, by considering instrumentation errors and representation errors as discussed above. The measurement error can easily be time varying if justified by such considerations. In the present implementation, the error at a given tide gauge station is assumed independent of all other stations. For the instrumental errors, this is true, but for representation error, this might be violated.

3.2.2. Model Error

[27] Let $\mathbf{x}_M^t(t_i)$ define the true state represented in the space of the model at time t_i. A system equation can then be formulated as

$$\mathbf{x}_M^t(t_i) = M_M\left(\mathbf{x}_M^t(t_{i-1}), \mathbf{u}(t_{i-1})\right) + \boldsymbol{\xi}_i. \tag{7}$$

Thus the model error, $\boldsymbol{\xi}_i$, describes the error imposed by the model operator, M_M, at time t_i. This error must be described along with the error covariance of the state at an initial point in time in order to provide a stochastic description of the system.

[28] The description of model error is a complex task. The exclusion of processes at the very level of the definition of the mathematical model and the spatial discretizations used in the state description are model errors, but are described in terms of the representation error as discussed in section 3.2.1. Errors in the mathematical formulation of processes we wish to describe (including feedback from undescribed processes!), and the numerical methods used to solve the equations as well as numerical truncation errors and parameter specifications, all impose errors in the model simulation. Finally, incorrect forcing terms are potentially major sources of model error. The model error has a complex spatially varying structure and is dynamically altered throughout its propagation in time. It is thus presently intractable to describe accurately. However, an approximate second-order description of its statistical properties is not out of reach.

[29] When looking at the sources of model error in a well-calibrated hydrodynamic model of a coastal area, it is a good first approximation to assume that the main error source at each time step comes from the forcing terms. The system is quite strongly driven by its forcing, and these are known to be inaccurate. Atmospheric forcing is provided by meteorological forecast or hindcast models, and open boundary water levels are typically described by a model of harmonic constituents. In the present implementation it is assumed that all other model errors are neglectable and hence a model error description can be provided if the error sources in the forcing terms can be propagated throughout the system. The errors in the forcing terms are assumed to be colored processes described by an autoregressive model with a spatially co-varying error driving it, i.e.,

$$\boldsymbol{\xi}_i = M_{AR(1)}(\boldsymbol{\xi}_{i-1}, \boldsymbol{\eta}_i) = \mathbf{A}\,\xi_{i-1} + \boldsymbol{\eta}_i, \tag{8}$$

where $\mathbf{A} = diag(\boldsymbol{\alpha})$. For the sake of simplicity, the noise process $\boldsymbol{\eta}_i$ is assumed Gaussian with zero mean and error covariance matrix, $\mathbf{Q}_i^{\eta} \in \mathbf{R}^{r\times r}$. Hence $\mathbf{x}^t(t_i)$ is augmented with the open boundary water level and wind velocity error description, and an extended operator, $M = (M_M, M_{AR(1)})^T$, is introduced. This leads to a system equation with additive noise, which will be used in the remainder of this work,

$$\begin{aligned}\begin{bmatrix}\mathbf{x}_M^t(t_i)\\ \boldsymbol{\xi}_i\end{bmatrix} &= \mathbf{x}^t(t_i) = M(\mathbf{x}^t(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\eta}_i)\\ &= \left[M_M\left(\mathbf{x}_M^t(t_{i-1}), \mathbf{u}(t_{i-1})\right) + \mathbf{x}\mathbf{i}_{i-1}\right]\\ &\quad [\mathbf{A}\ \mathbf{x}\mathbf{i}_{i-1}]\end{aligned} \tag{9}$$

The error covariance of $\begin{bmatrix}0\\ \boldsymbol{\eta}_i\end{bmatrix}$ is $\mathbf{Q}_i = \begin{bmatrix}0 & 0\\ 0 & \mathbf{Q}_i^{\eta}\end{bmatrix}$.

[30] *Dee* [1995] devised a technique for estimating error covariance parameters, but it requires the number of observations at a single time step to exceed the number of tunable parameters by 2 or 3 orders of magnitude. Hence, in real application of assimilation schemes, the determination of the error covariance $\mathbf{Q}_i^{\eta}$ as well as $\mathbf{R}_i$ in equation (3) is based on experience and theoretical considerations. Thus a moderate sensitivity to these parameters is vital for their successful application.

3.3. Ensemble Kalman Filter

[31] The Kalman filter based data assimilation schemes used today are all based on the BLUE estimator. They differ mainly in the way they propagate the stochastic state representation. The foundation of the Ensemble Kalman Filter (EnKF) is to approximate the propagation of the full pdf using a Markov Chain Monte Carlo technique [*Evensen*, 1994]. While the deterministic model in equation (1) or equation (2) propagates the state assuming the model and forcing to be perfect, the EnKF takes the stochastic nature of the model prediction and the nonlinearities explicitly into account.

[32] An ensemble of q state realizations is defined at an initial point in time. In the approach presented here, the same initial state defines all ensembles with zero variance at the beginning of a spin-up period. During this period the forcing errors are propagated throughout the system to provide the initial mean state estimate and model error covariance matrix.

[33] Each ensemble member is propagated according to

$$\mathbf{x}_j^f(t_i) = M\left(\mathbf{x}_j^a(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\eta}_{j,i}\right), \qquad j = 1, \ldots, q. \tag{10}$$

The model error, $\boldsymbol{\eta}_{j,i}$, is randomly drawn from a predefined distribution with zero mean and covariance, $\mathbf{Q}_i^{\eta} \in \mathbf{R}^{r \times r}$. With each ensemble propagated by equation (10), the mean state estimate and model error covariance estimate are provided by the following equations:

$$\hat{\mathbf{x}}^f(t_i) = \frac{1}{q}\sum_{j=1}^{q} \mathbf{x}_j^f(t_i) \tag{11}$$

$$\mathbf{P}^f(t_i) = \mathbf{S}_i^f \left(\mathbf{S}_i^f\right)^T, \quad \mathbf{s}_{j,i}^f = \frac{1}{\sqrt{q-1}}\left(\mathbf{x}_j^f(t_i) - \hat{\mathbf{x}}^f(t_i)\right). \tag{12}$$

The vector, $\mathbf{s}_{j,i}^f \in \mathbf{R}^n$, is the jth column of $\mathbf{S}^f{}_i \in \mathbf{R}^{n \times q}$. The update can be performed by equations (4) and (5), when given the proper interpretation in an ensemble setting. For computational efficiency, an algebraically equivalent set of equations are used.

[34] In order to maintain correct statistical properties of the updated ensemble, each ensemble member must be updated rather than the ensemble state estimate. For the same reason, an ensemble of measurements must be generated and used for each ensemble member update accordingly rather than the measurement itself [*Burgers et al.*, 1998]. Hence

$$\mathbf{y}_{j,i}^o = \mathbf{y}_i^o + \boldsymbol{\epsilon}_{j,i}, \quad j = 1, \ldots, q. \tag{13}$$

Randomly generated realizations, $\boldsymbol{\epsilon}_{j,i}$, of $\boldsymbol{\epsilon}_i$ are added for each member. The update scheme presented here specifically uses the uncorrelated measurement structure to assimilate simultaneous measurements sequentially. The updating algorithm for every ensemble member, j, reads [*Chui and Chen*, 1991],

$$\mathbf{x}_{j,m}^a(t_i) = \mathbf{x}_{j,m-1}^a(t_i) + \mathbf{k}_{i,m}\left(y_{j,i,m}^o - \mathbf{h}_{i,m}\mathbf{x}_{j,m-1}^a(t_i)\right), \quad m = 1, \ldots, p, \tag{14}$$

and $\mathbf{x}_{j,0}^a(t_i) = \mathbf{x}_j^f(t_i)$. In equation (14), $y_{j,i,m}^o$ is the mth element in $\mathbf{y}_{j,i}^o$ and $\mathbf{h}_{i,m}$ is the mth row of $\mathbf{H}_i$. Treating one measurement at a time, the Kalman gain is a vector, $\mathbf{k}_{i,m}$, given by

$$\mathbf{k}_{i,m} = \frac{\mathbf{S}_{i,m-1}^a \mathbf{c}_{i,m}}{\mathbf{c}_{i,m}^T \mathbf{c}_{i,m} + \sigma_{i,m}^2}, \qquad \mathbf{c}_{i,m} = \left(\mathbf{S}_{i,m-1}^a\right)^T \mathbf{h}_{i,m}^T. \tag{15}$$

The mth diagonal element in $\mathbf{R}_i$ is denoted $\sigma_{i,m}^2$ The matrix $\mathbf{S}_{i,m}^a$ in equation (15) is calculated as

$$\mathbf{S}_{i,m}^a = \left[\mathbf{s}_{1,i,m}^a \ldots \mathbf{s}_{q,i,m}^a\right], \quad \mathbf{s}_{j,i,m}^a = \frac{1}{\sqrt{q-1}}\left(\mathbf{x}_{j,m}^a(t_i) - \hat{\mathbf{x}}_m^a(t_i)\right), \tag{16}$$

for $m = 1, \ldots, p$ and $\mathbf{S}_{i,0}^a = \mathbf{S}_i^f$. Now, equations (14), (15), and (16) provide the update equations of all ensemble members, one measurement at a time. The time consumption of the EnKF is of the order of q times a standard model execution.

3.4. Regularization

[35] In the EnKF, most of the computational effort is used for providing an estimate of the Kalman gain matrix, $\mathbf{K}_i$. This matrix contains $n \times p$ elements which are calculated based on q ensembles. This leads to uncertain estimates which in particular can have an unwanted effect in data-sparse regions with large model variability. Such areas are susceptible to erroneous updates from spurious correlation estimates [*Hamill et al.*, 2001]. However, even for $q = \infty$, the gain estimate will only have a limited accuracy because of the simplistic nature of the models used to describe measurement and model error. Propagating an approximate error source gives an approximate error covariance matrix.

[36] Regularization methods allow the expression of a prior knowledge about the elements in $\mathbf{K}_i$ and their interdependence to be taken into account [*Hastie et al.*, 2001]. The techniques can usually be cast in a Bayesian framework; for example, if prior information about the model error covariance, $\mathbf{P}^{prior}$, is available for $\mathbf{P}^f$, then the posterior estimate, $\mathbf{P}^{posterior}$, is

$$\left(\mathbf{P}^{posterior}\right)^{-1} = \left(\mathbf{P}^{prior}\right)^{-1} + \left(\mathbf{P}^f\right)^{-1}. \tag{17}$$

Such an approach is not tractable in the high-dimensional state space under consideration. However, this line of thought can still provide a useful angle on Kalman filtering. Is there knowledge about the model error covariance that clearly conflicts with the estimates provided by, for example, the EnKF? Regularization methods deliberately make biased estimates in order to lower the variance of the estimated elements. Because of the approximate error models and structural model errors, the estimates of the $\mathbf{K}_i$ elements will typically be biased anyhow, so it makes sense to express this in order to lower the total prediction error of the elements, which is a sum of squared bias and variance.

3.4.1. Smoothing of Kalman Gain

[37] Not to be confused with Kalman smoothing for the state vector estimate, a temporal smoothing factor, s, is introduced. It is used to regularize the EnKF derived gain matrix in equation (5) and implemented in equation (15). With the instantaneous Kalman gain still being denoted $\mathbf{K}_i$, a smoothed Kalman gain, $\mathbf{K}_i^S$, which replaces $\mathbf{K}_i$, is obtained as

$$\mathbf{K}_i^S = (1-s)\mathbf{K}_{i-1}^S + s\mathbf{K}_i, s \in [0;1]. \tag{18}$$

This approach reduces the stochastic variability of the gain estimate at the cost of leaving out high-frequency signals in the gain as well as introducing a phase error. In general, the use of a smoothing factor gives a good performance even for insufficient ensemble sizes [*Sørensen et al.*, 2004a]. Thus it allows a smaller q to be chosen for the same performance, implicitly saving computational time. This proves the need for regularization techniques for efficient filtering. It can be regarded as an intermediate method in between the Ensemble Kalman Filter and the Steady Kalman filter described subsequently.

3.4.2. Steady Approximation

[38] The Steady Kalman filter can be regarded as an ad-hoc regularization method. Instead of calculating the Kalman gain at every measurement time, it can be assumed that the state and measurement error covariances are the same at every update, which yields a constant Kalman gain. This gain is calculated as a long-term average of Kalman gains estimated by the EnKF. Since the gain actually is varying, this introduces a bias in the gain, but the time averaging that creates the steady gain smoothes the gain and lowers the variance. This variance reduction possibly lowers the prediction error of the gain elements if the time-varying bias indeed is not too big. Using a snapshot of the gain from the EnKF would similarly be expected to make the estimate worse, since it still would result in an increased bias without lowering the variance.

[39] The Steady Kalman filter uses equation (9) for the model propagation with $\boldsymbol{\eta}_i = 0$,

$$\mathbf{x}^f(t_i) = M(\mathbf{x}^a(t_{i-1}), \mathbf{u}(t_{i-1}), \mathbf{0}). \tag{19}$$

Subsequently, equation (14) is used for the analysis, where $\mathbf{k}_i$ is calculated as a time average from an execution of the EnKF, $p = 1$ and $y^o_{j,i,m} = y^o_{i,m}$. The time consumption of the Steady Kalman filter is of the order of a standard model execution.

[40] The Kalman gain is calculated as a long-term average of the gain from an EnKF, where the error has been assumed to lie in the open boundary water level and in the wind velocity. The noise is thus included as a quadratic term in the momentum equations. On average, this leads to an overestimation of the values in the Steady Kalman gain, but in periods with strong winds it is underestimated. This is the approximation made by the steady assumption.

3.4.3. Barotropic Approximation

[41] The method described in this section belongs to the group of methods that apply simplified dynamics for calculating the model error covariance and hence the Kalman gain. The idea is that since the water level response to variations in tides and wind-forcing is mainly barotropic, its error covariance due to errors in open boundary conditions and wind velocity can be well modeled by a depth-averaged barotropic model. The forecast step is composed of an ensemble forecast step using the 2D model and a single forecast of the 3D model according to equation (19).

[42] The first component of the analysis step consists of the EnKF analysis for the 2D model. The other component is to update the full 3D forecast based on the ensemble statistics from the 2D model. The augmented $AR(1)$ error model part of the state-space has the same size and interpretation in both model spaces, and hence it can be carried directly over from 2D to 3D. Depth-averaged velocity, $\mathbf{V}$, and water level, l, also have similar interpretation, and hence in the corresponding subspace the 2D Kalman gain can be applied directly in 3D.

[43] However, temperature, T, salinity, S, and the three-dimensional velocity, $\mathbf{v}$, are not included in the 2D state-space, and thus additional assumptions must be imposed. The error covariances between T and S and water levels in the measurement points are all assumed to be zero. This means that the thermodynamic variables are unaffected by the analysis. The velocity, on the other hand, needs to be updated. When $\mathbf{V}$ is updated, then for consistency, $\mathbf{v}$ must be updated as well, since $\mathbf{V}$ is the depth average of $\mathbf{v}$. A vertical structure, $s(z_k)$, must be chosen under the constraint that its depth average is not zero. Let V_x^a be the updated depth-averaged x velocity component in the 3D model. Let $v_x^f(z_k)$ and $v_x^a(z_k)$ be the forecast and analysis of the x velocity in the 3D model at depth z_k. Now the updated full velocity field can be found by solving the following equations for $\mathbf{v}_x^a = (v_x^a(z_1), \ldots, v_x^a(z_{Kmax}))$ and the zonal structure scaling parameter θ:

$$V_x^a = \left(\mathbf{v}_x^a\right)^T \mathbf{dz} \tag{20}$$

$$\mathbf{v}_x^a = \mathbf{v}_x^f + \mathbf{s}\theta, \tag{21}$$

where $\mathbf{v}_x^f = (v_x^f(z_1), \ldots, v_x^f(z_{K_{max}}))$, $\mathbf{s} = (s(z_1), \ldots, s(z_{K_{max}}))$, $\mathbf{dz} = (dz(z_1), \ldots, dz(z_{K_{max}}))$ is a vector of layer depth, and K_{max} is the number of layers.

[44] The solution to equations (20) and (21) is

$$\theta = \frac{V_x^a - \left(\mathbf{v}_x^f\right)^T \mathbf{dz}}{\mathbf{s}^T \mathbf{dz}} \tag{22}$$

$$\mathbf{v}_x^a = \mathbf{v}_x^f + \mathbf{s}\frac{V_x^a - \left(\mathbf{v}_x^f\right)^T \mathbf{dz}}{\mathbf{s}^T \mathbf{dz}}. \tag{23}$$

A similar set of equations can be solved for the other velocity component. The vertical velocity is updated by the mass conservation equation. In the present study, $\mathbf{s}$ was chosen to be $\mathbf{s} = (1, \ldots, 1)$. This corresponds to simply moving the entire forecast velocity profile to match the updated depth-averaged velocity.

[45] In the light of regularization, the scheme assumes all elements in $\mathbf{K}_i$ that are used for updating T and S to be zero and elements for updating the velocity components to be related through equation (23). Again, this certainly may introduce a bias, if the assumption of no correlation to the observed water levels or the vertical interdependence of the velocity errors break down, but a much lower variance has been obtained. Most importantly, a huge time reduction has been won in the model error propagation.

[46] *Sørensen et al.* [2002] compared the barotropic approximation applied to the Steady Kalman filter to the standard EnKF as well as the Steady Kalman filter with no barotropic approximation in an idealized bay setup. All methods showed similar performance, but the barotropic approximation has the lowest computational requirements both in terms of time and memory demand.

3.4.4. Distance Regularization

[47] The use of distance regularization comes down to a trade-off between accepting inaccurate elements in the Kalman gain and introducing spurious or nondynamical modes in the analysis. The model error covariance is modeled dynamically by assuming errors in the forcing terms. There are many good reasons for doing so, but it may lead to correlations in the error of the state that conflicts with our prior knowledge, and hence a regularization can be performed taking this into account.

[48] The distance regularization is an ad hoc procedure for expressing that we do not believe any tide gauge

observation should be used for updating state variables that are positioned far away. This is implemented by constructing a vector, with coefficients between 0 and 1, which are a Gaussian function ϕ of their geographical distance, d_m to observation, m, according to

$$\phi(d_m) = \exp\left(-\frac{d_m^2}{2D^2}\right). \tag{24}$$

The parameter, D, specifies the spatial decorrelation scale. This regularization can be used in either the EnKF or the Steady Kalman filter ($p = 1$) presented above, by modifying the analysis equation (14) according to

$$\mathbf{x}^a_{j,m}(t_i) = \mathbf{x}^a_{j,m-1}(t_i) + \hat{\mathbf{k}}_{i,m}\left(y^o_{j,i,m} - \mathbf{h}_{i,m}\mathbf{x}^a_{j,m-1}(t_i)\right), \quad m = 1, \ldots, p$$
$$\hat{\mathbf{k}}_{i,m} = \begin{pmatrix} \hat{k}_{i,m}(1) \\ \vdots \\ \hat{k}_{i,m}(n) \end{pmatrix} = \begin{pmatrix} k_{i,m}(1)\phi(d_m) \\ \vdots \\ k_{i,m}(n)\phi(d_m) \end{pmatrix}. \tag{25}$$

4. Description of Experiments

[49] The main objective of this study is to demonstrate the hindcast performance of the time-efficient barotropic approximation for a Steady Kalman filter in the Water Forecast model. Further, the impact of applying a time-varying EnKF with barotropic and distance regularization is investigated. No comparison is made to an Ensemble Kalman Filter in the full three-dimensional setting, because it is not operationally feasible. Further, *Sørensen et al.* [2002] demonstrated that the dynamic regularization has similar performance to the full three-dimensional implementation of the time-varying EnKF in a simple test case.

[50] All experiments span the period: 0000 local time (LT) January 1 to 0000 LT January 29, 2002. The initial state is obtained from the database of the operational system. The steady gain employed in the study is based on the period 0000 LT January 2 to 0000 LT January 6, 2002. Figure 8 (in section 5) shows that this period includes a single storm surge event and average winter conditions the rest of the time.

[51] The results will be compared to a reference run, which is obtained from a hindcast execution of the Water Forecast system. All assimilation runs make use of the barotropic approximation and hence do not have higher demands to the computational hardware than the Water Forecast itself, and have operational execution times less than 2.5 times that of the reference run. The model runs can be summarized as follows:

[52] • Reference Run is a standard 3D water forecast model execution with no use of data assimilation.

[53] • Steady is a 3D model execution with the Steady Kalman filter. The gain is obtained from the 2D model using the EnKF with temporal smoothing.

[54] • Steady Dist is a 3D Model execution like Steady, but with distance regularization used in the 3D environment.

[55] • EnKF is a 3D model execution using a time-varying gain obtained from the 2D model employing temporal smoothing. Distance regularization is enabled in the 3D execution.

[56] When adopting the barotropic regularization in all assimilation runs, it is implied that the central model forecast (equation (19)) is employed for the 3D model, while the ensemble forecast (equations (10) and (11)) is employed for the calculation of 2D steady and time-varying Kalman gains.

[57] In any assimilation approach it is important first to correct the measurement datum such as to approximately represent model datum in order to allow proper intercomparison between observed and modeled quantities. Model datum is determined by the open boundary levels and a long-term average dynamical balance. In order to assess the model datum, the water levels at all measurement stations were extracted from a 1-year simulation spanning all of 2002. The time average was calculated for each station, and the corresponding measurement was adjusted to match this average. Note that the model error may have a seasonal dependence, and hence the datum-corrected measurements may still contain an off-set in January, where the study is performed. The measurements were adjusted to the model datum for both the 2D and the 3D model.

[58] A number of parameters need to be specified in the filtering schemes. The assimilation system is too complex for statistically based parameter estimation, and hence first guesses based on experience and theoretical considerations are used. For tide gauges, the measurement representation error is in general dominating over the instrumentation error. The water level readings can be expected to measure the truth projected onto the model space with an accuracy around 0.05 m. Hence the tidal gauge measurement errors are assumed to have mutually uncorrelated, unbiased Gaussian distributions with a standard deviation of 0.05 m. However, sometimes less trust is put in the measurement in order to constrain the model less. This is the case in M1, Esbjerg, M9, Rønne and M10, Kalix, where the standard deviations were assumed to be 0.15, 0.08, and 0.15, respectively.

[59] The model wind error was assumed to have a temporal correlation scale of 5.7 hours and a spatial correlation scale of 300 km. The standard deviation of the white noise in equation (8) was assumed to be 0.3 m/s leading to a standard deviation of 3.0 m/s for the wind. The model error in open boundary water levels was assumed to have a temporal correlation scale of 1.7 hours and a spatial correlation scale of 95 km. The standard deviation of the white noise in the boundary error was assumed to be 0.05 m. This leads to a standard deviation of 0.27 m for the water level.

[60] An ensemble size of 100 and a smoothing factor of 0.05 was used for the EnKF runs of the study. Measurements were available every 30 min. These were linearly interpolated in time, and the model was updated from the interpolated time series every 10 min. The distance parameter, D, in the distance regularization was set to 250 km.

[61] In the work of *Sørensen et al.* [2004a] a study of the root mean square error (RMSE) sensitivity to a range of filter parameters of the EnKF was undertaken in an ideal test case. The study investigated both misspecified filter parameter values and misspecified error structure. The general conclusion was that the filter performance is robust with respect to moderate parameter variability and also significantly improves results even with different structural error sources. On the basis of this analysis as well as our

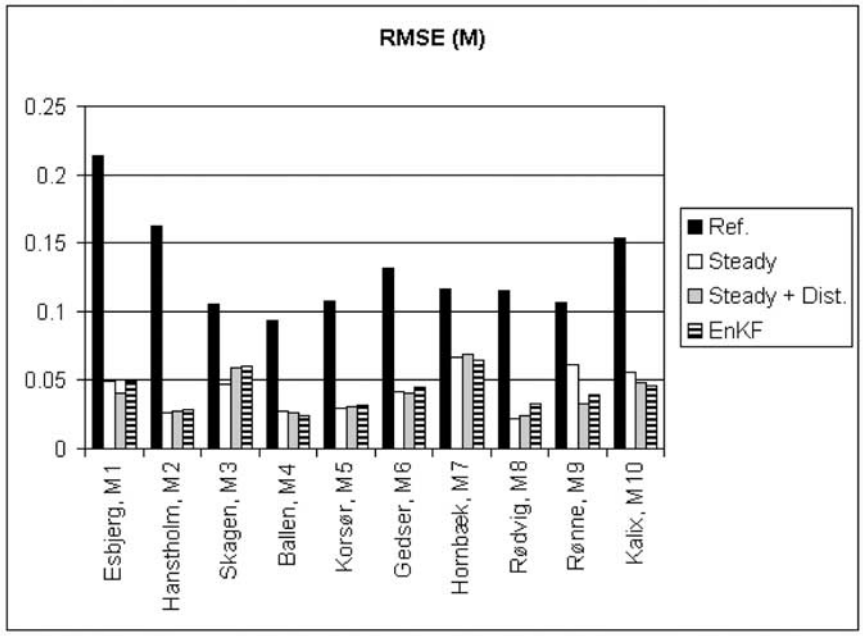

Figure 2. RMSE performance of measurement stations.

experience in applying the filter, it is fair to assume that the sensitivity in the Water Forecast model is moderate at worst. As a measure of the filter performance, the RMSE of water levels, l, calculated over the 28-day simulation period for each measurement and validation station is used,

$$RMSE = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(l_{obs}(t_i) - l_{predicted}(t_i)\right)^2}. \qquad (26)$$

5. Results and Discussion

[62] The performance of the reference run and the three assimilation runs are summarized in Figures 2 and 3 for measurement and validation stations, respectively. For the reference run, RMSE is in the range 0.10 m to 0.15 m for most stations with M1-Esbjerg peaking above 0.20 m.

[63] The RMSE can be decomposed into a standard deviation and a bias component. Such an analysis shows that the datum correction method equating 1-year averages discussed in section 4 leaves a variability at monthly time-scales with biases in the range -0.12 to 0.07 for the reference run. This might be due to long-term variability in meteorological error (the boundaries cannot explain such long-term variability) or long-term error components in the model (biases in annual cycle of the density modeling, etc.). However, at present, the bias is accepted as the working conditions, adhering to the bias correction properties of filters using a colored noise implementation [*Sørensen et al.*, 2004b].

[64] Figure 2 also shows that all assimilation runs significantly reduce the RMSE in measurement points. The remaining error is in fair agreement with the standard deviation of 0.05 m assumed in most stations. Figure 3 shows similar good performance in stations close to measurement points. However, in the Baltic Sea (V6 and V7) far from measurements, the Steady Kalman filter without distance regularization significantly degrades the results.

[65] Figures 4–7 show examples of the water level part of the Steady Kalman gain for the measurement stations, M3-Skagen and M9-Rønne, with and without the distance regularization imposed. In the M3-Skagen station the gain is clearly affected by the error assumed in the tidal signal, while the M9-Rønne station is dominated by the wind-driven dynamics. In this latter case the gain without distance regularization shows large corrections in the entire southern part of the Baltic. This can in part be explained by the

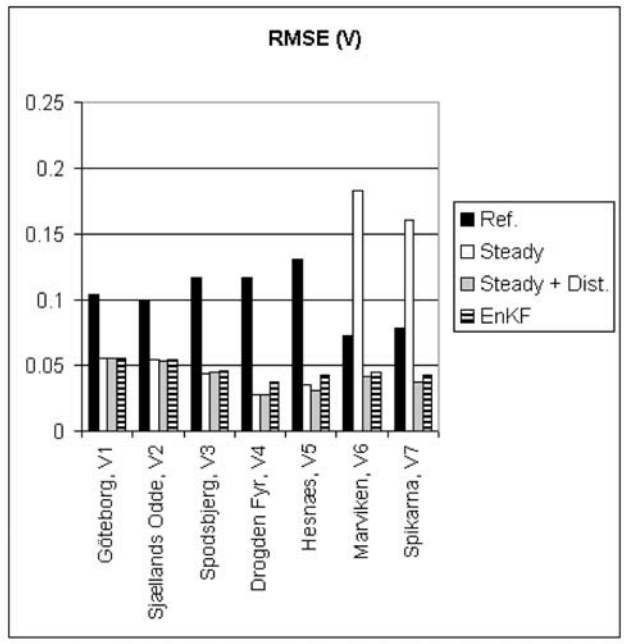

Figure 3. RMSE performance of validation stations.

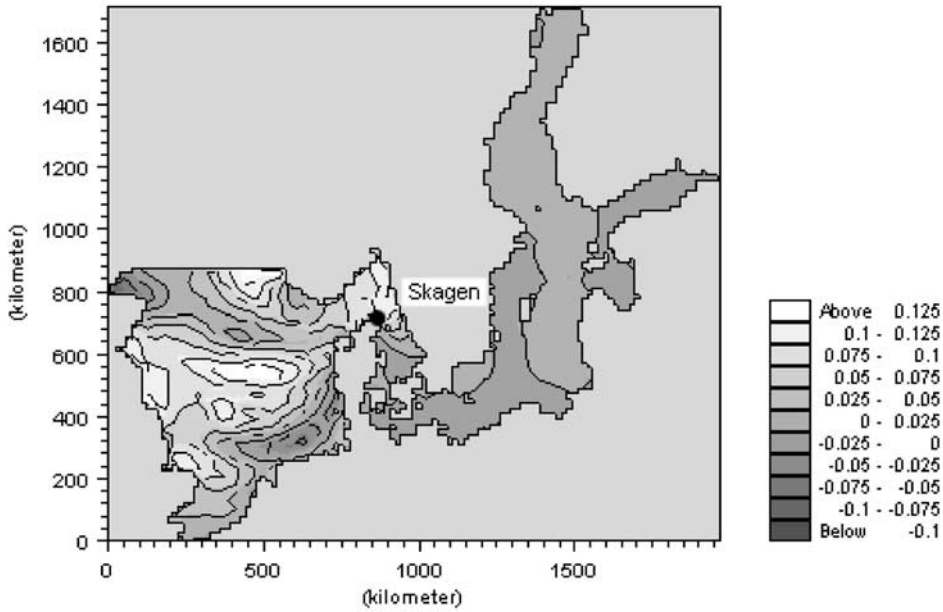

Figure 4. Water level part of the Steady Kalman gain for M3-Skagen.

assumed wind error model, which has spatial correlation scale of 300 km. The distance regularization of the M3-Skagen station effectively filters the gain structure in the North Sea, which contrasts our prior understanding of the system and our modeling capability.

[66] The model water level has a large variance far from observations in the Baltic, and hence even a small covariance will provide an impact on the update in this area. The stochastic variability of the gain is filtered out in the Steady approach and thus does not contribute significantly to the gain structure. The distance regularized gain structure dampens the effect of distant correlations by imposing the assumption that such error correlation does not exist despite its prediction by the filter. As is evident in Figure 3, this significantly improves the results in the data-sparse Baltic. The distance and barotropic regularized Steady Kalman filter adds significant state estimation skill in all measurement and validation points at a very low computational cost both for the generation of the gains and for execution, enabling use in an operational setting.

[67] A time series plot of measured water level in the validation station, V1-Göteborg, is displayed in Figure 8

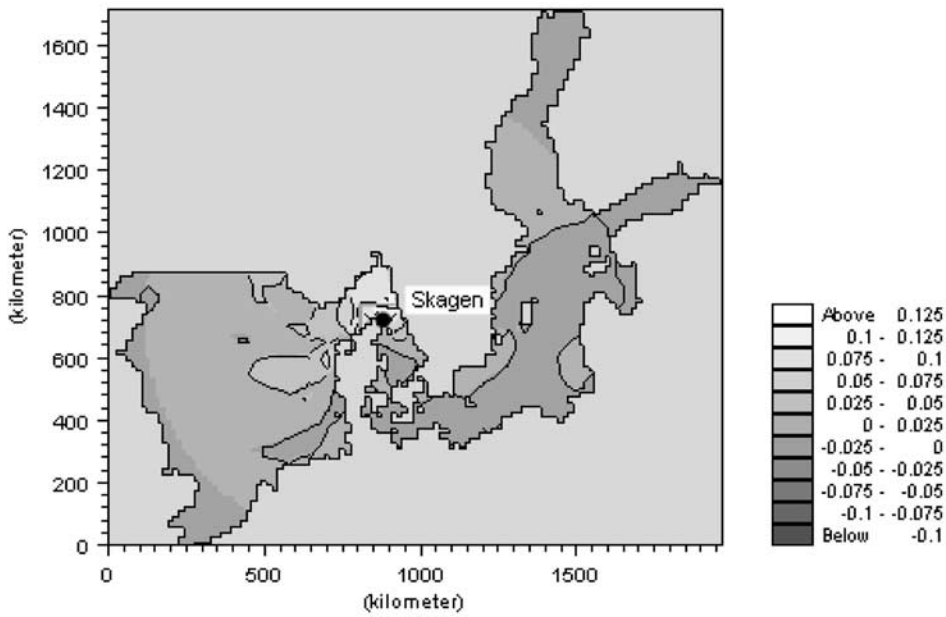

Figure 5. Water level part of the distance regularized Steady Kalman gain for M3-Skagen.

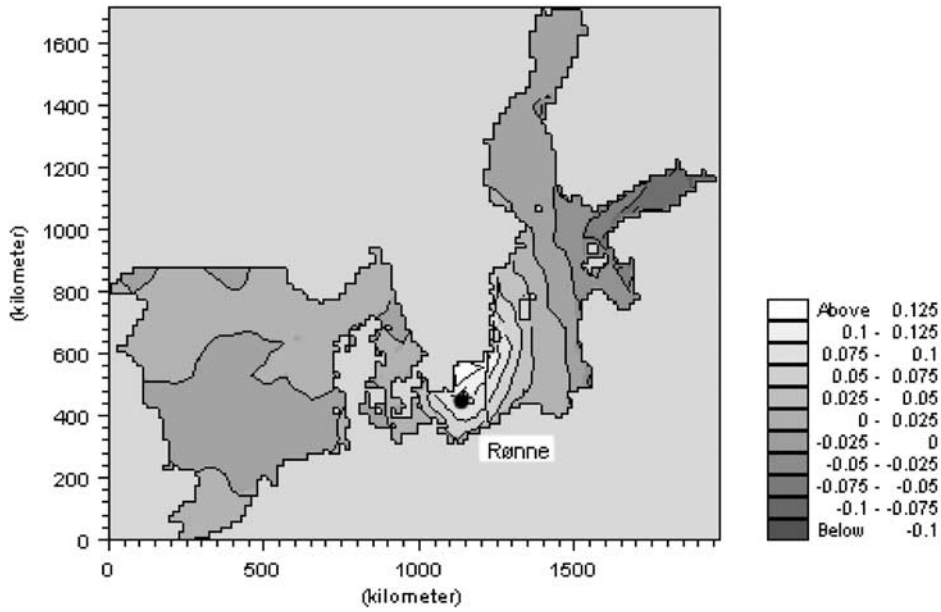

Figure 6. Water level part of the Steady Kalman gain for M9-Rønne.

along with estimates by the reference run and the distance and barotropic regularized Steady Kalman filter. This plot shows the good performance previously expressed by the statistical RMSE measure in a visually interpretative form.

[68] The use of a time-varying gain from the EnKF with the barotropic approximation and distance regularization was compared with the successful Steady approach. Figures 2 and 3 show the results as the last horizontally striped bar. The performance is similar to that of the Steady distance regularized scheme. However, its implementation is more demanding on computational time, although the 2D EnKF execution with 100 ensemble members has a similar speed as a single 3D model execution, and hence still can be applied in operational settings.

[69] Figures 9 and 10 show the variance of the EnKF derived gain for the water level portion of the gain for the M9-Rønne station with and without smoothing in the ensemble run. The variability of the nonsmoothed Kalman gain shows its maximum values far from the station itself indicating spurious correlations. This also explains the model problems that has been encountered when applying

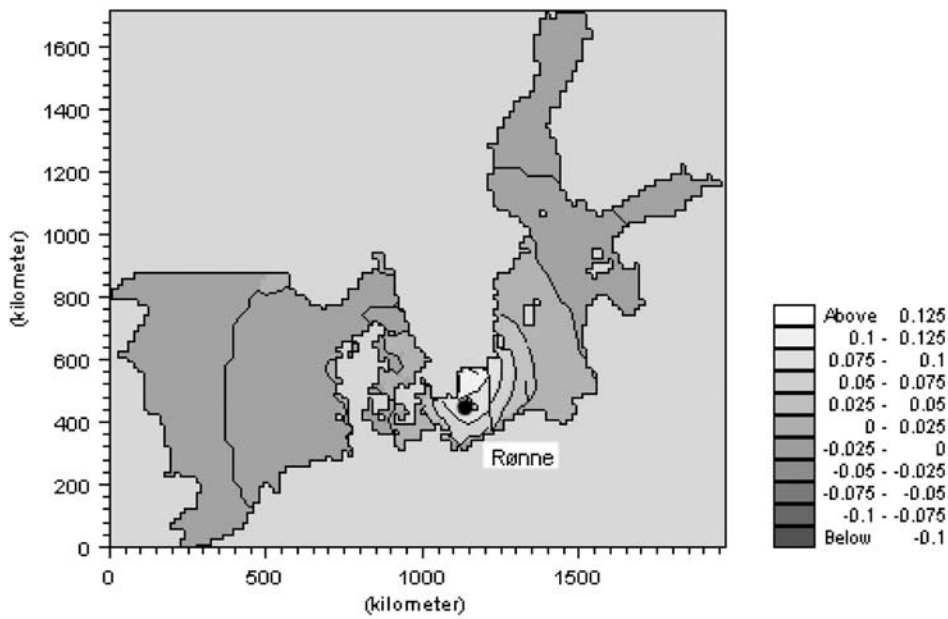

Figure 7. Water level part of the distance regularized Steady Kalman gain for M9-Rønne.

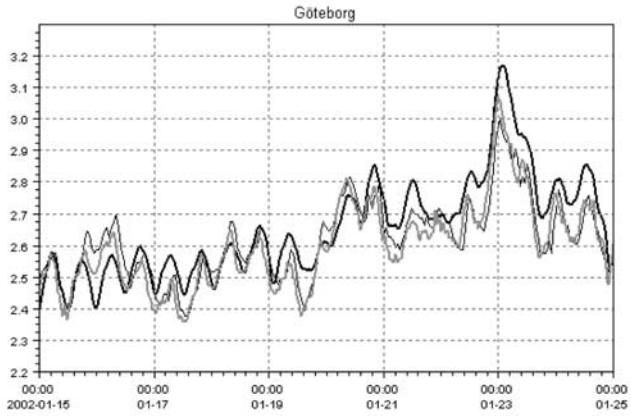

Figure 8. Time series of water level in V1-Göteborg. The thin black line is the measured level. The thick black line shows the reference solution. The thick shaded line shows the solution with the barotropic and distance regularized Steady Kalman filter.

the standard EnKF without distance regularization. In this case, the analysis may impose a state estimation which is not a likely outcome in the real system despite the fact that the EnKF always produces its analyzed ensemble members as linear combinations of the forecast members (an example is pulling water out of a shallow region until a water point is dried out). A spurious correlation can last over a dynamically significant length of time due to the colored noise implementation.

[70] The variability near M9-Rønne is quite similar with and without smoothing in the EnKF gain calculation, and in both cases the variability is small compared to the actual size of the steady gain in Figure 6. This small Kalman gain variability in regions where the update is also largest explains the similar performance of the Steady filter and the EnKF.

6. Conclusions

[71] The water level estimation problem has been discussed and the well-known Ensemble Kalman Filter technique has been presented for solving the problem. In this sequential setting the estimation of the water levels requires an estimation of the elements of the Kalman gain matrix as an intermediate step, which is important for understanding the behavior of the scheme. The estimate of the gain elements possess both a bias and a variance, because of inaccurate measurement and model error descriptions and the stochastic variability in the EnKF. This uncertainty is discussed from the viewpoint of regularization techniques, and a Kalman gain smoothing, a Steady Kalman filter, a barotropic approximation, and a distance regularization are discussed in this light.

[72] These techniques are combined and tested for the assimilation of water levels in the Water Forecast operational system. The Steady and the barotropic approximations show the best performance at the lowest cost. The use of distance regularization has been demonstrated to be important for data sparse regions, while maintaining performance in areas with denser data coverage. The difference in the RMSE of the various filter algorithms is moderate in the Inner Danish Waters, and it must be kept in mind that the sensitivity to parameter values is likely on the same scale.

[73] The distance and barotropic regularized Steady Kalman filter has a good estimation skill in all areas of the model. Further, its low computational cost enables easy operational implementation.

[74] Future developments will investigate the use of regularization techniques for controlling the bias-variance

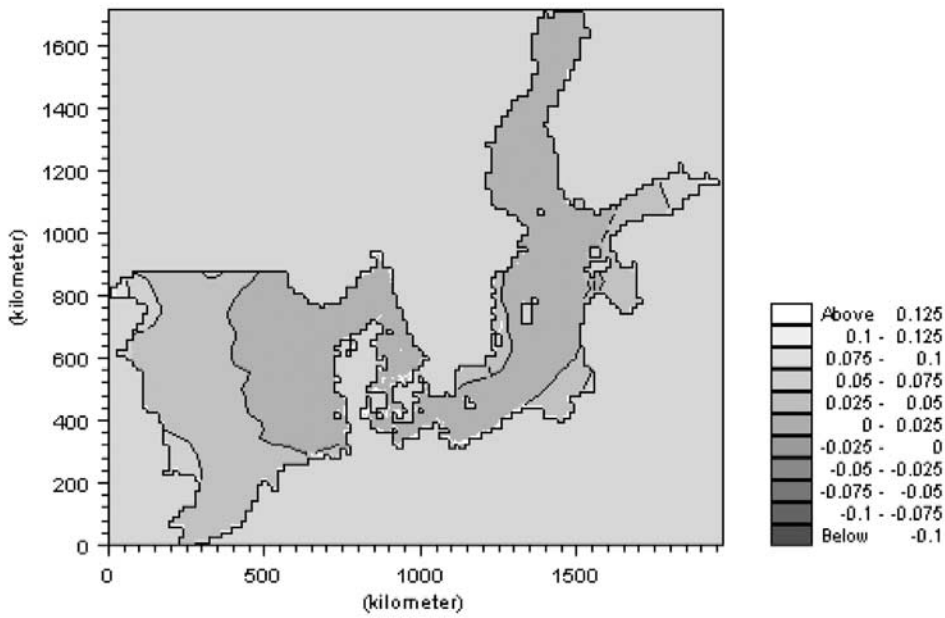

Figure 9. Standard deviation of M9-Rønne water level part of the Kalman gain derived using smoothing.

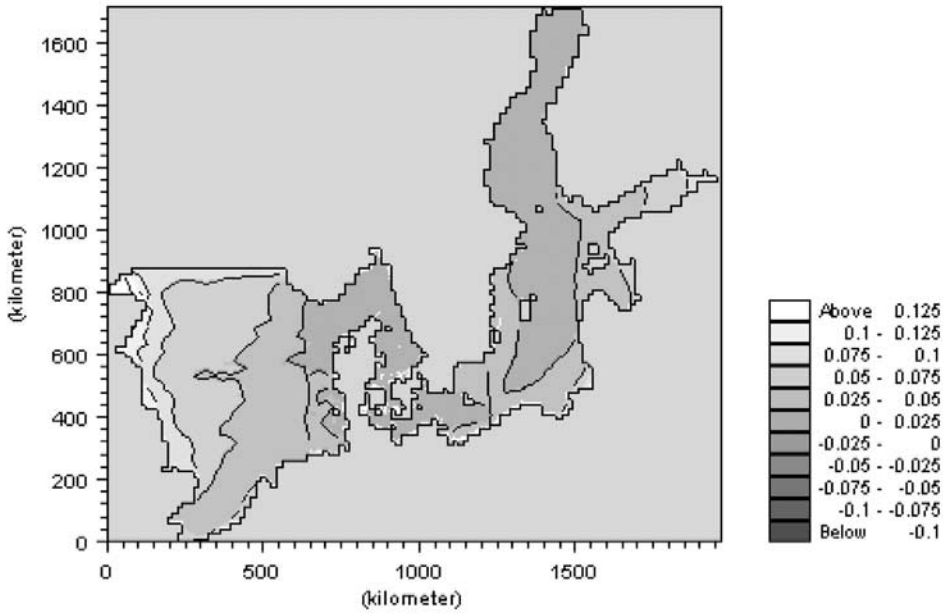

Figure 10. Standard deviation of M9-Rønne water level part of the Kalman gain derived without using smoothing.

trade-off together with attempting to improve model error description. Also, more work needs to be done on estimating the bias of the measurements when applied in a model datum frame work. Most important in an operational setting is the forecast skill. This will be addressed in a future study.

[75] **Acknowledgments.** This research was carried out jointly at DHI Water and Environment and the Technical University of Denmark under the Industrial Ph.D. Programme (EF835). Contribution of tide gauge data from the Danish Meteorological Institute, the Royal Danish Administration of Navigation and Hydrography, and the Swedish Meteorological and Hydrological Institute is acknowledged.

References

Bahurel, P., P. De Mey, C. Le Provost, and P.-Y. Le Traon (2002), A GODAE prototype system with applications-Example of the MERCATOR system, paper presented at International Symposium "En route to GODAE," Cent. Natl. d'Etudes Spatiales, Toulouse, France.

Bertino, L., G. Evensen, and H. Wackernagel (2002), Combining geostatistics and Kalman filtering for data assimilation in an estuarine system, *Inverse Prob.*, *18*, 1–23.

Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), Analysis scheme in the Ensemble Kalman Filter, *Mon. Weather Rev.*, *126*, 1719–1724.

Cañizares, R., H. Madsen, H. R. Jensen, and H. J. Vested (2001), Developments in operational shelf seamodelling in Danish waters, *Estuarine Coastal Shelf Sci.*, *53*, 595–605.

Chui, C. K., and G. Chen (1991), *Kalman Filter With Real-Time Applications*, 195 pp., Springer-Verlag, New York.

Dee, D. P. (1991), Simplification of the Kalman filter for meteorological data assimilation, *Q. J. R. Meteorol. Soc.*, *117*, 365–384.

Dee, D. P. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, *123*, 1128–1145.

DHI (2001), MIKE 3 estuarine and coastal hydrodynamics and oceanography, user manual, DHI Water and Environ., Hørsholm, Denmark.

DHI (2002), MIKE 21 coastal hydraulics and oceanography, user manual, DHI Water and Environ., Hørsholm, Denmark.

Erichsen, A. C., and P. S. Rasch (2002), Two- and three-dimensional model system predicting the water quality of tomorrow, in *Proceedings of the Seventh International Conference on Estuarine and Coastal Modeling*, edited by M. L. Spaulding, pp. 165–184, Am. Soc. of Civ. Eng., Reston, Va.

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

Evensen, G. (2003), The Ensemble Kalman Filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367.

Fukumori, I., and P. Melanotte-Rizzoli (1995), An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model, *J. Geophys. Res.*, *100*, 6777–6793.

Fukumori, I., R. Raghunath, L.-L. Fu, and Y. Chao (1999), Assimilation of TOPEX/Poseidon altimeter data into a global ocean circulation model: How good are the results?, *J. Geophys. Res.*, *104*, 25,647–25,665.

Gerritsen, H., H. de Vries, and M. Philippart (1995), The Dutch Continental Shelf Model, in *Quantitative Skill Assessment for Coastal Ocean Models*, *Coastal Estuarine Stud.*, vol. 47, edited by D. R. Lynch and A. M. Davies, pp. 425–467, AGU, Washington, D. C.

Hamill, T. M., J. S. Whitaker, and C. Snyder (2001), Distance-dependent filtering of background error covariance estimates in an Ensemble Kalman Filter, *Mon. Weather Rev.*, *129*, 2776–2790.

Hastie, T., R. Tibshirani, and J. Friedman (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 533 pp., Springer-Verlag, New York.

Heemink, A. W., and H. Kloosterhuis (1990), Data assimilation for non-linear tidal models, *Int. J. Numer. Methods Fluids*, *11*, 1097–1112.

Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an Ensemble Kalman Filter technique, *Mon. Weather Rev.*, *126*, 796–811.

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

Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, 376 pp., Academic, San Diego, Calif.

Madsen, H., and R. Cañizares (1999), Comparison of extended and Ensemble Kalman Filters for data assimilation in coastal area modelling, *Int. J. Numer. Methods Fluids*, *31*, 961–981.

Pham, D. T., J. Verron, and M. C. Roubaud (1997), Singular evolutive Kalman filter with EOF initialization for data assimilation in oceanography, *J. Mar. Syst.*, *16*, 323–340.

Pham, D. T., J. Verron, and L. Gourdeau (1998), A singular evolutive Kalman filter for data assimilation in oceanography, *C. R. Acad. Sci., Ser. IIa Terre Planetes*, *326*, 255–260.

Pinardi, N., and J. Woods (2002), *Ocean Forecasting: Conceptual Basis and Applications*, 472 pp., Springer-Verlag, New York.

Pinardi, N., F. Auclair, C. Cesarini, E. Demirov, S. Fonda-Umani, P. Oddo, M. Tonani, G. Montanari, and M. Zavatarelli (2002), Toward marine environmental predictions in the Mediterranean Sea coastal areas: A monitoring approach, in *Ocean Forecasting: Conceptual Basis and Applications*, edited by N. Pinardi and J. Woods, pp. 339–376, Springer-Verlag, New York.

Rogers, E., T. Black, B. Ferrier, Y. Lin, D. Parrish, and G. DiMego (2001), Changes to the NCEP Meso Eta Analysis and Forecast System: Increase in resolution, new cloud microphysics, modified precipitation assimila-

tion, modified 3DVAR analysis, *NWS Tech. Procedures Bull. 488*, Natl. Weather Serv., Silver Spring, Md.

Sørensen, J. V. T., H. Madsen, and H. Madsen (2002), Towards an operational data assimilation system for a three-dimensional hydrodynamic model, in *Proceedings of the Fifth International Conference on Hydroinformatics*, vol. 2, edited by I. D. Cluckie et al., pp. 1204–1209, Inst. of Welsh Affairs, Cardiff, UK.

Sørensen, J. V. T., H. Madsen, and H. Madsen (2004a), Parameter sensitivity of three Kalman filter schemes for the assimilation of tide gauge data in coastal and shelf sea models, *Ocean Modell.*, in press.

Sørensen, J. V. T., H. Madsen, and H. Madsen (2004b), Data assimilation in hydrodynamic modelling: On the treatment of nonlinearity and bias, *Stochastic Environ. Res. Risk Assess.*, in press.

Verlaan, M., and A. W. Heemink (1997), Tidal flow forecasting using reduced rank square root filters, *Stochastic Hydrol. Hydraul.*, *11*, 349–368.

Vested, H. J., J. W. Nielsen, H. R. Jensen, and K. B. Kristensen (1995), Skill assessment of an operational hydrodynamic forecast system for the North Sea and Danish Belts, in *Quantitative Skill Assessment for Coastal Ocean Models*, *Coastal Estuarine Stud.*, vol. 47, edited by D. R. Lynch and A. M. Davies, pp. 373–396, AGU, Washington, D. C.

H. Madsen, DHI Water & Environment, Agern Allé 5, DK-2970 Hørsholm, Denmark. (hem@dhi.dk)

H. Madsen, Richard Petersens Plads, Building 321, DK-2800 Kongens Lyngby, Denmark. (hm@imm.dtu.dk)

J. V. T. Sørensen, DHI Water & Environment, Agern Allé 11, DK-2970 Hørsholm, Denmark. (jts@dhi.dk)