Environmental Modelling & Software 54 (2014) 24–38

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Ensemble based prediction of water levels and residual currents in Singapore regional waters for operational forecasting

Rama Rao Karri [a,*], Xuan Wang [a], Herman Gerritsen [b]

[a] Singapore-Delft Water Alliance, National University of Singapore, Singapore 117577
[b] Deltares, P.O. Box 177, 2600 MH Delft, The Netherlands

A R T I C L E   I N F O

Article history:
Received 24 July 2013
Received in revised form
10 December 2013
Accepted 13 December 2013
Available online 9 January 2014

Keywords:
Data assimilation
Ensemble Kalman filter
Steady state Kalman filter
Singapore regional model
Sea level anomaly
Malacca Strait

A B S T R A C T

Singapore Strait located between the South China Sea and Andaman Sea is driven by tides coming from both sides and the hydrodynamics in this area is complex. From the viewpoint of long term forecasting, however, models developed for this area suffer from limitations introduced by parametric uncertainty, absence of data for appropriate specification of forcing and lateral boundary conditions. For improving the model forecasts, a data assimilation technique based on ensemble Kalman filter is implemented and applied. Based on the latter, an ensemble based steady state Kalman filter is derived to address the computational limitation for daily operational forecasting. Via a twin experiment on a simulation period that includes a significant storm surge event (sea level anomaly) the skills of both data assimilation schemes are assessed and compared.

1. Introduction

Understanding ocean dynamics and the physics of its driving forces is important for safe navigation and offshore operations. With the development of information technology, numerical ocean models have become popular to describe the dynamics of water bodies in shelf seas as well. However, the numerical models tend to be less than perfect due to reasons such as incomplete representation of the ocean physics and, other simplifying assumptions, numerical approximations, insufficient resolution, uncertainty in the model parameters and uncertainty in the forcing terms (Babovic and Fuhrman, 2002; Sun et al., 2008).

The water body of interest, Singapore Strait is one of the busiest shipping routes in the world and its coastal area is heavily utilized for port activities or industrial facilities with rapid economic development. Providing accurate hydrodynamic information of the depth-integrated, barotropic regional water motion is important for scheduling of navigation, docking and minimizing other potential hazards. The hydrodynamics of Singapore regional waters, however, is complex since the water motion in Singapore Strait is driven by tides originating from South China Sea, Andaman Sea and Java Sea. With the objective to provide reliable hydrodynamic information for the region, the Singapore regional model (SRM) was developed by Kernkamp and Zijl (2004). The model was later improved by Kurniawan et al. (2010) by analysing the sensitivity of the tidal representation in the coastal waters around Singapore to various modelling parameters, leading to improved tidal representation. For a detailed description of model setup, discretization, parameters, etc., see Kernkamp and Zijl (2004), Ooi et al. (2009) or Kurniawan et al. (2011). While the SRM has been shown to be able to capture the key depth-integrated barotropic hydrodynamic phenomena and yields desirable forecasts in most scenarios, the model is less capable to accurately predict detailed flows and water levels due to the nonlinear nature of the uncertainties especially near the coast. To improve the accuracy of model predictions, various data assimilation procedures can be applied (Van Loon et al., 2000; Babovic and Fuhrman, 2002; Aguirre et al., 2005; Solomatine and Ostfeld, 2008; Van Velzen and Segers, 2010; Liu et al., 2012; Zijl et al., 2013). The methods are classified into four categories (Babovic et al., 2005):

(a) updating of input parameters,
(b) updating of model parameters,
(c) updating of state variables, and
(d) updating of output variables.

* Corresponding author.
E-mail addresses: kramarao.iitd@gmail.com, rama@nus.edu.sg (R.R. Karri).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.12.006

In method (a), the input variables are updated by certain iterative procedures to address the input uncertainties whereas; in method (b) the model parameters are updated through calibration techniques. The main difference between method (c) and method (d) is that in method (c) the data assimilation process is integrated with the model and the solution procedure has to be adjusted; while in method (d) the model solution procedure and the data assimilation procedure for the model are detached. Updating of state variables by Kalman filter techniques provides an optimal update for linear systems (Kalman and Bucy, 1961). For a nonlinear system the extended Kalman filter (EKF) is a natural choice, as it extends the basic algorithm to nonlinear problems by linearizing the nonlinear function around the current estimate. As such, it is found to be capable and efficient in handling high dimensional systems like ocean dynamics and meteorological weather forecasting (Heemink et al., 1997; Sorensen and Madsen, 2004; Karssenberg et al., 2010; Wei and Malanotte-Rizzoli, 2010). For strongly nonlinear systems EKF is known to fail to estimate unmeasured variables of nonlinear systems (Aguirre et al., 2005; Solomatine and Ostfeld, 2008). The Ensemble Kalman filter (EnKF), one of the most advanced sequential assimilation methods (Evensen, 1994, 2003; Hamill, 2006), extends the conventional Kalman filter using an ensemble forecast computed from the nonlinear model directly to estimate an error covariance matrix. This technique has been successfully applied in different applications with varying range of nonlinearity (Evensen, 1994; Houtekamer and Mitchell, 1998; Zang and Malanotte Rizzoli, 2003; Heemink et al., 2009; Zamani et al., 2010).

To improve the SRM model further with the aim of day-to-day detailed forecasting, the application of data assimilation schemes based on ensemble Kalman filter is found to be very beneficial (Babovic et al., 2011; Karri et al., 2013). These assimilation methods integrate the model solution with the available observations to obtain a practically improved but theoretically suboptimal solution up to the present time epoch. However, for these highly nonlinear dynamic systems, the implementation of EnKF demands much computational effort. Therefore, a steady state Kalman gain derived from EnKF termed an ensemble steady state Kalman filter (EnSSKF) is proposed to reduce the computational demand without compromising on the accuracy of the forecast improvement.

The objective of the study is to evaluate and compare the skills of data assimilation schemes EnKF and EnSSKF to correct the model states of SRM and accurately predict the water levels and currents in Singapore regional waters. For a historic period that includes a significant surge event a twin experiment is defined. The advantage of such a twin experiment approach is that model based; synthetic observations used in the evaluation are available at all model points, allowing skill evaluations and comparisons at positions where actual measurements have not been made. Both data assimilation schemes are applied using the open source data assimilation framework OpenDA (www.openda.org) to assimilate the observations in the model states of SRM. Initially, the EnKF is applied with ensemble size $N_E = 64$ to derive the Kalman gain at each time step. Furthermore, a steady state gain is computed based on the average of all the gains over time. Using this steady state Kalman gain, the derived EnSSKF is also used for estimating the water levels and currents. In order to compare and evaluate the performances of both EnKF and EnSSKF, they are subjected to hindcast or reanalysis runs where observations are available at different tidal gauge stations to analyse the quality of the improved estimates. To statistically validate the forecast ability of these two data assimilation schemes for different time horizons, the corrected/updated states are used as initial states and the deterministic model is run into the future in (retrospective) forecast mode, for a prediction horizon 1–24 h.

2. Data assimilation schemes

2.1. Ensemble Kalman filter

The ensemble Kalman filter, which was first proposed by Evensen (1994), is a suboptimal estimator. The error statistics in this method are predicted by using a Monte Carlo integration to solve the Fokker–Planck equation (Gillijns et al., 2006). A brief description of the EnKF is included here and further details are available in references (Evensen, 2003; Hamill, 2006). The nonlinear model propagates in time to obtain a forecast (background) estimate of model states $x_{k/k-1}$ at time t_k. The initial condition of state vector $\hat{x}_{k|k-1}$ is the estimate at time t_k conditioned on the measurement until time t_k. At every time step t_k, the state vector ($x^i_{k|k-1}$) of each ensemble member i, is forced by the model error w^i_k and propagates in time. The model errors were drawn from a predefined distribution with zero mean and covariance matrix. The update of the state vector can be estimated through the mean of the ensemble as

$$\hat{x}_{k|k-1} = \frac{1}{N}\sum_{i=1}^{N} x^i_{k|k-1}, \tag{1}$$

where N is the maximum ensemble size.

The error covariance matrix of the update state vector, $P_{k|k-1}$ is calculated as a covariance matrix of the ensemble

$$\begin{aligned} P_{k|k-1} &= L_{k|k-1}L^T_{k|k-1} \\ L_{k|k-1} &= \tfrac{1}{\sqrt{N-1}}\sum_{i=1}^{N}\left(x^i_{k|k-1} - \hat{x}_{k/k-1}\right) \end{aligned} \tag{2}$$

The Kalman gain K_k, which represents the optimal weighting between the error covariance of the model states P_k and the errors of the observed state R_k, is calculated as

$$K_k = \frac{P_{k|k-1}H^T_k}{\left(H_k P_{k|k-1} H^T_k + R_k\right)} \tag{3}$$

The updated state vector for each ensemble is calculated as the gain weighted difference between the observed states z^i_k and the transformed model forecast:

$$x^i_{k|k} = x^i_{k|k-1} + K_k\left(z^i_k - H_k x^i_{k|k-1}\right) \tag{4}$$

where, H_k is the measurement operator that maps the state vector to the measurement domain.

2.2. Steady state Kalman filter

Canizares et al. (2001) noticed the fact that for high dimensional systems, the error covariance matrix often becomes nearly invariant after 1–2 days of simulation and in some instances even less depending on the nonlinearity of the system and the simulation time step. In order to make use of this feature, a constant Kalman gain (weighting matrix) can be calculated during 1–2 days simulation. This will be similar to an optimal interpolation method, although in this case the constant weighting matrix has been calculated from a time variant sequential data assimilation method. So, in the steady state Kalman filter, the constant Kalman gain K is computed offline to circumvent the error covariance propagation. This course of action makes the steady state Kalman filter the most cost efficient among the suboptimal routines.