Assessing an Ensemble Kalman Filter Inference of Manning's n coefficient of an Idealized Tidal Inlet against a Polynomial Chaos-Based MCMC

Adil Siripatana · Talea Mayo · Ihab Sraj ·
Omar Knio · Clint Dawson · Olivier Le Maitre ·
*Ibrahim Hoteit

Received: date / Accepted: date

Abstract Bayesian estimation/inversion is commonly used to quantify and reduce modeling uncertainties in coastal ocean model, especially in the framework of parameter estimation. Based on Bayes rule, the posterior probability distribution function (pdf) of the estimated quantities is obtained conditioned on available data. It can be computed either directly, using a Markov Chain Monte Carlo (MCMC) approach, or by sequentially processing the data following a data assimilation approach, which is heavily exploited in large dimensional state estimation problems. The advantage of data assimilation schemes over MCMC-type methods arises from the ability to algorithmically accommodate a large number of uncertain quantities without significant increase in the computational requirements. However, only approximate estimates are generally obtained by this approach due to the restricted Gaussian prior and noise assumptions that are generally imposed in these methods. This contribution aims at evaluating the effectiveness of utilizing an ensemble Kalman-based data assimilation method for parameter estimation of a coastal ocean model against an MCMC Polynomial Chaos (PC)-based scheme. We focus on quantifying the uncertainties of a coastal ocean ADCIRC model with respect to the Manning's n coefficients.

Based on a realistic framework of observation system simulation experiments (OSSEs), we apply an ensemble Kalman filter and the MCMC method employing a surrogate of AD-

Adil Siripatana
King Abdullah University of Science and Technology, Thuwal, KSA

Talea Mayo
Princeton University, New Jersey, USA

Ihab Sraj
King Abdullah University of Science and Technology, Thuwal, KSA

Omar Knio
King Abdullah University of Science and Technology, Thuwal, KSA

Clint Dawson
University of Texas at Austin, Texas, USA

Olivier Le Maitre
Laboratoire d'Informatique pour la Mecanique et les Sciences de l'Ingénieur, Paris, France

*Ibrahim Hoteit
4700 King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia
E-mail: ibrahim.hoteit@kaust.edu.sa

CIRC constructed by a non-intrusive PC expansion for evaluating the likelihood, and test both approaches under identical scenarios. We study the sensitivity of the estimated posteriors with respect to the parameters of the inference methods, including ensemble size, inflation factor, and PC order. A full analysis of both methods, in the context of coastal ocean model, suggests that an ensemble Kalman filter with appropriate ensemble size and well-tuned inflation provides reliable mean estimates and uncertainties of Manning's n coefficients compared to the full posterior distributions inferred by MCMC.

Keywords Coastal ocean model · Manning's n coefficients · Parameter estimation · Ensemble Kalman filter · Polynomial Chaos · MCMC

1 Introduction

Coastal and estuarine systems are home to more than half of the human population. Understanding and forecasting the dynamics of these systems is critical for human resources sustenance around the world, both economically and ecologically. Coastal inundation during extreme events, such as hurricanes and tsunamis, is a major cause of destruction to human lives and their habitats. Even in normal conditions, accurately forecasting coastal ocean events (e.g., tides and coastal flows) is important for marine activities (e.g., [81, 62]).

Coastal ocean modeling is widely utilized to simulate shallow water circulations for the purposes of conservation, contaminant transport modeling, development of coastal structures (e.g., bridges, dams and breakwaters), and emergency and economic planning. State-of-the-art coastal ocean models are generally based on the shallow water equations (SWEs), which are derived by depth-integrating the Navier-Stokes equations, assuming hydrostatic pressure and horizontal length scales that are large in comparison to the vertical length scales [58]. Even with the simplifications, such models contain many parameters, carrying uncertainties on which the accuracy of model prediction strongly depends [81]. Moreover, many of the parameters cannot be directly measured, or in some situations the relevant data to estimate these parameters are difficult to collect [13]. The values of these parameters must therefore be inferred from available data [21].

Quantifying and reducing uncertainties in the model outputs associated with uncertainties in the parameters is essential for reliable and robust coastal ocean simulations and predictions (e.g., [29, 62, 74]). Of particular importance is the Manning's n coefficient of roughness, introduced in the SWEs through the bottom stress components in the momentum equation [43, 46].

A general framework for inferring model parameter is Bayesian inversion [36, 42, 47, 72], which is usually performed numerically using Markov Chain Monte Carlo (MCMC) [34, 28, 66]. MCMC allows sampling the probability distribution of the unknown parameters conditioned on available observations, called the posterior, given a prior distribution and the likelihood of predicting the observations given a set of parameters. MCMC is however computationally prohibitive for state-of-the-art coastal ocean models. To avoid sampling the posterior, this approach is often cast as a variational least-squares problem derived using a Gaussian noise assumption [22, 77]. It is then solved deterministically, using an optimization algorithm to determine the parameters that best fit the model to the observations [4]. The disadvantage of variational methods is that they can be computationally demanding and are difficult to implement because they generally require the development of an adjoint model [26, 4]. Consequently, ensemble data assimilation methods which sequentially implement the Bayesian inversion problem have become more popular because they are

non-intrusive and can be implemented with reasonable computational requirements [6, 12, 39]. These methods may further provide information about the uncertainty associated with the predictions [18, 37, 40]. The Ensemble Kalman filter (EnKF) approach is the most popular among sequential methods, combining an integration of an ensemble of state vectors in the forecast step and a Kalman filter update whenever new data becomes available [15, 24, 26]. These methods were first developed in the context of state estimation, but have been extended to include the estimation of the model parameters as well, though this approach is still not yet widely applied in coastal ocean modeling [62].

There is currently a growing interest in using EnKF methods for parameter estimation of ocean models, as they have been found to be efficient and robust in many coastal ocean state estimation applications (e.g., [70, 16, 3, 5]). This is particularly needed as the predictability of coastal models may in certain situations depend more on the system parameters than the initial state. Parameter estimation within an EnKF framework is usually achieved through state space augmentation [19, 6]. In this approach, parameters are considered to be part of the system state, and they are updated via the cross-covariance between the parameters and the observations as sampled by the ensemble members, similar to the way state variables are updated. However, it has been suggested that parameter estimation using EnKF may not perform well with strongly nonlinear systems [49]. This is because EnKFs are formulated based on Gaussian prior and noise assumptions [7, 38, 55, 71], which may not hold for such problems.

Few recent studies have evaluated ensemble Kalman-based techniques against the more sophisticated but computationally demanding Markov Chain Monte Carlo (MCMC) methods. The comparison of the EnKF for state estimation of the two-dimensional incompressible Navier-Stokes equations in a periodic box against the gold-standard MCMC in [51] suggests that the bias in the Gaussian approximation may become significant and the covariance may misrepresent the dependence between the parameters and the data when the dynamics are sufficiently nonlinear. [66] also tested parameter estimation in the context of an idealized 1D deep convection model, comparing an ensemble transform Kalman smoother (ETKS) with a full Markov Chain Monte Carlo (MCMC) method. Similar conclusions were drawn; with proper parameter choices, evaluated filters perform well in producing the distribution's mean but not an accurate covariance because the model is strongly nonlinear.

To assess the relevance of EnKFs for parameters estimation in the context of realistic coastal ocean modeling, we consider the Advanced Circulation model (ADCIRC) [58] and conduct observation system simulation experiments (OSSEs) to attempt to recover the Manning's n coefficient from a set of initial guesses and available data. To evaluate the EnKF performance for parameter estimation, we compare the produced posterior distributions against those resulting from MCMC. Since the implementation of MCMC is too costly with realistic coastal ocean models, we resort to a polynomial chaos (PC) method to sample the MCMC posterior distribution with a computationally efficient surrogate model.

PC methods have become one of the standard approaches to propagate and quantify uncertainties in various disciplines, including mechanical systems (e.g., [53, 30, 52]), chemical systems (e.g., [65, 54, 63]), and more recently large scale oceanic systems (e.g., [2, 74, 75]). These methods represent the model parameters in terms of a spectral expansion in an orthogonal polynomial basis according to their probabilistic distributions. This enables the construction of extremely cost effective surrogate models that can then be used to efficiently sample the statistical properties of some Quantities of Interest (QoIs) of the model outputs [29, 50]. The advantage of using PC based inference over an EnKF is in its ability to provide the full probability distribution function of the estimated parameters, Gaussian or not. Nonetheless, PC suffers from 'the curse of dimensionality', which makes it difficult to

implement in problems with large number of stochastic parameters [2]. Here we show that with good tuning of the filter parameters, one can accurately recover the reference parameter, even when the posterior pdf is parameterized Gaussian. We further perform sensitivity experiments to various parameters (e.g., ensemble size and inflation factor) and study their impact on the posterior pdfs.

The rest of this paper is organized as follows. Section 2 establishes the Bayesian framework for parameter estimation, comprising the parameter inference methodology with the EnKF and PC-MCMC. Section 3 elaborates in detail the experimental setup. Inference results are presented and analyzed in section 4. Summary and conclusions are offered in section 5.

2 Bayesian inference

Bayes' rule states that the probability distribution function (pdf) of the unknown parameters θ conditioned on available data y , can be computed as [68,44]

$$p(\theta | y) = \frac{p(y | \theta)p_{pr}(\theta)}{p_Y(y)}, \quad (1)$$

where $p(\theta | y)$ is also known as the posterior pdf, $p(y | \theta)$ is the likelihood function of obtaining data y given a parameter value θ , $p_{pr}(\theta)$ is the prior distribution which defines the prior knowledge of the parameters, and $p_Y(y)$ is a scaling factor, which is the density function of all possible measurements.

2.1 Parameter estimation with MCMC

The most straightforward way to sample the posterior distribution is to use an MCMC method [11,28]. These methods iteratively generate a Markov chain in which the parameters vector θ_i at a given iteration i depends only on the previous one θ_{i-1} . Suppose the current value of the chain is θ_i , one draws a proposal parameter θ' , for instance from a Gaussian function centered at the current parameter: $Q(\theta_i) \sim N(\theta_i, \beta^2 I)$, and then calculates the ratio of the likelihood evaluated for the proposed sample θ' and the previous sample θ_i :

$$r = \frac{L(\theta')}{L(\theta_i)}, \quad (2)$$

where $L(\theta) \doteq \pi(\theta | y)$. The new parameter θ_{i+1} is then chosen according to the following rule:

$$\theta_{i+1} = \begin{cases} \theta' & \text{if } \alpha < r, \text{ ACCEPT,} \\ \theta_i & \text{if Otherwise, REJECT.} \end{cases}$$

Here $\alpha \sim U(0, 1)$ and β^2 is the variance in the proposal drawing step. β^2 needs to be well-tuned in order to produce a well-mixed chain [34,61]. In this study, the value of β^2 was set by trial and error, but more sophisticated techniques could be also used such as adaptive MCMC [33,8] to achieve an average acceptance ratio between 0.40 and 0.50, which has been shown to be the optimal acceptance rate for inferring a 1-D Gaussian distribution [69].

2.2 Parameters estimation with an EnKF

Another mean of estimating $\pi(\theta | y)$ is a filtering approach in which the data are sequentially assimilated as they become available. The posterior is then computed conditioned on data available up to the estimation time only, and is updated with the Bayes' rule every time a new observation is available, using the most recent estimate as the prior [36].

The Kalman filter (KF) is optimal for state estimation when the system is linear and its noise is Gaussian [45]. The parameter estimation problem often arises, however, with a nonlinear system. One popular approach that has been developed in oceanography and meteorology to cope with the nonlinearity is the Ensemble Kalman filter (EnKF) and its variants (e.g., [23, 12, 6, 64, 39]). The EnKF uses an ensemble of realizations of the state vector, from which the first two moments are forecasted with the nonlinear model. The parameters inference is then performed recursively every time new observations become available, assuming Gaussian distributions.

EnKF methods can be stochastic, updating each ensemble member with perturbed observations, or deterministic, not requiring observations perturbations [15, 78, 3, 40]. The Singular Evolutive Interpolated Kalman (SEIK) filter belongs to the class of deterministic EnKFs and updates the sample mean and covariance of the forecast ensemble as in the KF. The SEIK filter is implemented in three steps: (1) a sampling step, (2) a forecast step, and (3) an analysis step. The sampling step generates an ensemble using a second-order resampling scheme that exactly matches the analyzed state and error covariance. The forecast step uses the nonlinear model to integrate the sampled ensemble members forward in time. The analysis step updates the forecast ensemble mean and covariance with incoming observations exactly as in the KF.

The filter starts from an estimate of the model state, $\mathbf{x}_{k-1}^a$ and a low-rank ($r \ll n$) error covariance, decomposed as $\mathbf{P}_{k-1}^a = \mathbf{L}_{k-1} \mathbf{U}_{k-1} \mathbf{L}_{k-1}^T$, where $\mathbf{L}_{k-1}$ is $n \times r$, and $\mathbf{U}_{k-1}$ is an $r \times r$ positive definite matrix. The SEIK filter then samples the analysis ensemble as

$$\mathbf{x}_{k-1}^{a,i} = \mathbf{x}_{k-1}^a + \sqrt{r+1} \mathbf{L}_{k-1} (\boldsymbol{\Omega}_{k-1} \mathbf{C}_{k-1}^{-1})^T, \quad i = 1, \dots, r+1,$$

where $r+1$ is the ensemble size and $\boldsymbol{\Omega}_{k-1}$ is $(r+1) \times r$ matrix with orthonormal columns and zero column sums generated using Householder matrices [39]. The analysis ensemble members are then integrated with the nonlinear model to obtain the forecast ensemble, $\mathbf{x}_k^{f,i}$. Taking the average $\overline{\mathbf{x}_k^{f,i}}$ of the $\mathbf{x}_k^{f,i}$ as the forecast state, and their sample covariance as the forecast error covariance, the latter can be decomposed as

$$\mathbf{P}_k^f = \mathbf{L}_k \mathbf{U}_{k-1} \mathbf{L}_k^T, \quad (3)$$

where

$$\mathbf{L}_k = \begin{bmatrix} \mathbf{x}_k^{f,1} & \dots & \mathbf{x}_k^{f,r+1} \end{bmatrix} \mathbf{T}, \quad (4)$$

and

$$\mathbf{U}_k^{-1} = \{\rho^{-1} [(r+1) \mathbf{T}^T \mathbf{T}]^{-1} + (\mathbf{L}_k^T \mathbf{L}_k)^{-1} \mathbf{L}_k^T \mathbf{Q}_k \mathbf{L}_k (\mathbf{L}_k^T \mathbf{L}_k)^{-1}\}^{-1} + [(\mathbf{H} \mathbf{L})_k]^T \mathbf{R}_k^{-1} (\mathbf{H} \mathbf{L})_k \quad (5)$$

Here, $\mathbf{T}$ is an $(r+1) \times r$ full rank matrix with zero column sums. $\mathbf{R}_k$ and $\mathbf{Q}_k$ are the observation and model covariance matrices, respectively. $\mathbf{H}$ is the observation matrix and ρ is an inflation factor, which increases the background error covariance and improves the forecast error when small ensemble sizes are used [64, 39]. The term $(\mathbf{L}_k^T \mathbf{L}_k)^{-1} \mathbf{L}_k^T \mathbf{Q}_k \mathbf{L}_k (\mathbf{L}_k^T \mathbf{L}_k)^{-1}$

represents the model error projection on the forecast ensemble. One may also account for the model error directly by perturbing the forecast model during the integration of the ensemble members [41].

When an observation y_k^i becomes available, the observation operator $\mathbf{H}_k$ is applied to each $\mathbf{x}_k^{f,i}$ in (4) to compute $(\mathbf{H}\mathbf{L})_k$. The forecast state is then updated to the analysis state as

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^o - \mathbf{H}_k\mathbf{x}_k^f), \quad (6)$$

where $\mathbf{K}_k$ is the Kalman gain computed as

$$\mathbf{K}_k = \mathbf{L}_k \mathbf{U}_k (\mathbf{H}\mathbf{L})_k^T \mathbf{R}_k^{-1},$$

The analysis error covariance can be expressed as

$$\mathbf{P}_k^a = \mathbf{L}_k \mathbf{U}_k \mathbf{L}_k^T. \quad (7)$$

The SEIK filter has recently been successfully applied to state estimation in storm surge forecasting [17, 3], providing good estimates of state and error statistics of the system using small ensembles ($\mathcal{O}(10)$).

[6] suggested that an EnKF could also be used for parameters estimation through joint state-parameters estimation, also commonly referred to as state augmentation [1, 25]. In this method, a vector of model parameters, $\mathbf{w}$, is appended to the system state, to form the joint state-parameter vector

$$\tilde{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{w} \end{bmatrix}.$$

Given that the evolution of the model parameters is a stationary process, a time evolution system for the state-parameter augmented state vector is thus constructed as

$$\tilde{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{w}_k \end{bmatrix} = \begin{bmatrix} \mathcal{M}(\mathbf{x}_{k-1}) \\ \mathbf{w}_{k-1} \end{bmatrix}. \quad (8)$$

The model parameters are not observed, so the corresponding observation operator is defined as

$$\tilde{\mathbf{H}}_k \tilde{\mathbf{x}}_k^f = \begin{bmatrix} \mathbf{H}_k \mathbf{x}_k^f \\ \mathbf{0} \end{bmatrix},$$

where $\tilde{\mathbf{H}}_k := [\mathbf{H}_k \quad \mathbf{0}]$, is the modified observation operator. Following the new augmented system one may directly apply the SEIK filter steps to jointly estimate both the model state and unknown parameters, and the associated error statistics.

Joint EnKF estimation has several desirable features. Its implementation is straightforward and its computational cost is very reasonable. In addition, it provides information about the uncertainty of the parameter's estimates. However, several issues in its implementation should be considered in practice. A specific issue concerns the assumption of stationary parameters. While the variance in the state variable increases during each forecast step, the variance in the parameter gradually decreases with the assimilation cycles. This may cause the ensembles to eventually collapse. The common practice to mitigate this problem is to use an inflation factor. In this approach, the posterior's standard deviation is inflated by

some factor to maintain a finite variance in the ensemble members. [1] suggested that inflation may not be always effective, depending on the characteristics of the parameter being estimated. Another inherent issue in the joint estimation method is that the parameters are only observed through the model state. Thus, the model state tends to converge to the true state much more rapidly than the estimated parameters. This suggests that increasing number of assimilation cycles may be required to obtain satisfactory estimates of the parameters [1]. Despite these issues, the joint-state EnKF approach has been successfully implemented and used in a wide range of problems [6, 9, 27, 1, 31].

2.3 Polynomial chaos-based MCMC (PC-MCMC)

MCMC is computationally prohibitive for realistic coastal ocean models, and can therefore only be implemented with some model reduction techniques. The polynomial chaos (PC) method is an efficient approach in this regard [50, 52]. Uncertainty quantification based on a PC framework could be implemented following different techniques [76]. Due to the complexity of our coastal ocean model, we adopt a non-intrusive approach that requires no modification to the model code.

The PC method assumes that the model output X to admit a spectral expansion of the form

$$X(\xi) = \sum_{k=0}^{\infty} c_k \Psi_k(\xi) \cong \sum_{k=0}^P c_k \Psi_k(\xi), \quad (9)$$

where $\xi \in \Omega^* \subseteq \mathbb{R}^d$ are independent random variables, with density $F_\xi : \mathbb{R}^d \mapsto \mathbb{R}^+$, parameterizing the random inputs, Ψ_k are d-variate orthogonal polynomials and c_k are the PC coefficients. In this work we use $\xi \stackrel{iid}{\sim} \mathcal{U}(-1, 1)^d$ so the Ψ_k are Legendre polynomials.

The expansion (9) is truncated to a certain polynomial order m , such that

$$P = \frac{(d+m)!}{d!m!} - 1. \quad (10)$$

Since $\{\Psi_k\}_{k=0}^P$ forms an orthogonal system,

$$\langle X, \Psi_k \rangle = c_k \langle \Psi_k, \Psi_k \rangle, \quad (11)$$

with

$$c_k = \frac{\langle X, \Psi_k \rangle}{\langle \Psi_k^2 \rangle}. \quad (12)$$

Here the moments $\langle \Psi_k^2 \rangle$ of the multivariate Legendre polynomials in (12) can be evaluated analytically [52] while $\langle X, \Psi_k \rangle$ is obtained from the projection

$$\langle X, \Psi_k \rangle = \int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_\xi(\mathbf{s}), \quad k = 0, \dots, P.$$

Thus evaluating c_k involves computing the values of a set of $P+1$ integrals over $\Omega^* \subseteq \mathbb{R}^d$, where Ω^* is the image of the sample space over ξ , which can be discretized as finite sums using an appropriate quadrature formula:

$$\int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_\xi(\mathbf{s}) \doteq \sum_{j=1}^{N_q} w_j X(\xi_j) \Psi_k(\xi_j). \quad (13)$$

Here, $\xi_j \in \Omega^*$ and w_j are the nodes and weights of the corresponding X , and N_q is the number of quadrature nodes. The set of integration quadrature nodes comprises the so-called non-intrusive spectral projection (NISP) sample denoted by

$$\mathcal{S} = \{\xi_j\}_{j=1}^{N_q} \subset \Omega^*. \quad (14)$$

Thus, to evaluate (13), one needs to compute $X(\xi_q)$ for all $\xi_q \in \mathcal{S}$. Let $\Pi \in \mathbb{R}^{(P+1) \times N_q}$ be the NISP projection matrix

$$\Pi_{k,j} = \frac{w_j \Psi_k(\xi_j)}{\langle \Psi_k^2 \rangle}, \quad k = 0, \dots, P, \quad j = 1, \dots, N_q,$$

and ζ be the vector with coordinates $\zeta_j = X(\xi_j)$. Then the vector of PC coefficients $\mathbf{c}$ can be expressed as $\Pi\zeta$, or in the coordinate form

$$c_k = \sum_{j=1}^{N_q} \Pi_{kj} \zeta_j = \sum_{j=1}^{N_q} \Pi_{kj} X(\xi_j), \quad k = 0, \dots, P. \quad (15)$$

The complexity of NISP scales with N_q and can increase greatly with the number d of canonical random variables. Thus the application of this approach is usually computationally restricted to a limited number of parameters. For a more detailed mathematical formulation of the NISP method the reader is referred to [2].

3 Experimental setup

3.1 The ADvanced CIRCulation Model (ADCIRC)

The coastal ocean model ADCIRC [57] is used in this study. This model is commonly implemented for simulating coastal and estuarine systems, as well as for analyses of water elevations and currents [58,62]. The model solves a modified version of the shallow water equations (SWEs); the continuity equation is replaced by the second-order, hyperbolic generalized wave continuity equation (GWCE). This prevents spurious oscillations that often arise from the numerical solution of the original form [60,48]. Together, the reformulation of the continuity equation and the momentum equation form the governing equations solved by ADCIRC. This system of equations are spatially discretized on unstructured, triangular elements and solved using a first-order continuous Galerkin finite element scheme. The time derivatives are approximated using centered finite differences in the GWCE and forward differences in the momentum equations.

A variety of physical domain scales and complex bathymetry structures can be represented in ADCIRC, including the range of scales necessary to represent the deep ocean basins and continental shelves, and coastal inland areas [17]. The minimum required inputs for ADCIRC are the description of the finite element mesh and tidal forcing parameters [35].

ADCIRC has been implemented and validated in many hindcast studies of several storms, including hurricanes Betsy (1965), Ivan (2004), Dennis (2004), Katrina (2005), Rita (2005) [80,14,20], Gustav (2008) [21] and Ike (2008) [46]. The model accuracy heavily depends upon the accuracy of model inputs and parameters. In this study we focus on the characterization of the bottom friction parameterized via the Manning's n coefficient.

3.2 The model discretization

We adopt the experiment of [62], which considers an idealized inlet with an ebb shoal domain as shown in Fig. 1. The structure of the domain includes a bay that is connected to the open ocean on the west side through an inlet with twin jetties. This model domain is a simplified version of a real ebb shoal, which is a common natural feature of coastal inlets. The domain contains 1,518 grid nodes and 2,828 elements, and is 4500 m wide and 3000 m long. The bathymetry of the domain is measured from the geoid toward the ocean floor. The depth of the ocean floor increases linearly from 3.8 m at the left-most boundary to 1 m at the mouth of the inlet on the west side of the domain. The landlocked area of the domain has a constant bathymetry of 1 m. The diameter of the ebb shoal is 750 m and the maximum height at the center of the ebb shoal is 4 m. On the open ocean boundary, ADCIRC is forced with the principal lunar semi-diurnal M_2 tidal constituent with an amplitude of 0.25 m relative to the geoid, and 2 s time step.

Fig. 1 Idealized inlet with ebb shoal domain. The discretization of the domain is represented. Observation stations are marked with red dots. The color bar represents the bathymetry of the domain measuring down from the geoid (m).

3.3 Parameterizing a field of Manning's n coefficient

Instead of estimating the Manning's n coefficient on each and every node of the domain, we model the field with low-dimensional parameterizations. This allows us to reduce the dimension of the parameter estimation problem. Here, we conduct the experiments in two Manning's n coefficients settings: (1) The estimation of fields of constant Manning's n coefficients. In this case, we assign a single constant value of Manning's n coefficient to every node in the domain, as commonly implemented in coastal ocean models. (2) The estimation

of two parameters, α and β , which parametrically define a 2-D field of Manning's n coefficients. In this case, we parameterize the field with two piecewise constants of Manning values in the open ocean (α) and the bay (β). The two constants field is parameterized in such a way that all nodes in the open ocean and land-locked area are assigned constant Manning's n coefficients, and these increase or decrease linearly for the nodes within the inlet. This is expressed as

$$\mathbf{w} = \alpha \mathbf{E}_1 + \beta \mathbf{E}_2 + (\beta - \alpha) \mathbf{E}_3. \quad (16)$$

Here $\mathbf{w}$ is the model parameter vector containing node-wise Manning's n coefficients. $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$, of the same dimension as $\mathbf{w}$, are the parameterization coefficients vectors with values ranging between 0 and 1. These vectors are referred to as the basis functions in [62]. The parameterized two constants Manning's field is denoted by $n_{\alpha,\beta}$. Note that if $\alpha = \beta$, the two constants Manning's n field reduces to a constant Manning's n field.

These two settings of Manning's n fields are idealized and only reflect bathymetry surface characteristics of the artificial inlet with an ebb shoal. For a more realistic setting where such parameters vary in space, some data (e.g., bottom surface characteristics) need to be collected. More sophisticated static data integration techniques, such as kriging [32] and Gaussian process regression (GPR) [67], can then be used to generate a highly varying 2-D Manning's n field conditioned on the data.

3.4 Manning's n coefficients classification

The model sensitivity greatly varies with the value of the Manning's n coefficients [62]. To study the sensitivity of the inverted solution to the choice of the prior distribution, we adopted a systematic Manning's n coefficients classification approach proposed in [62].

For the case of estimating single constant Manning's n coefficients, we classify Manning's n coefficients into five classes. The classes are defined by first running the ADCIRC model with the smallest Mannings n coefficient, 0.005, as smaller friction coefficients generate tides with larger amplitudes. The largest mean amplitude of the tides generated at several locations throughout the domain is computed, and the Mannings n coefficients that generate tides with mean amplitudes less than 20%, 40%, 60%, 80%, and 100% of this value are then divided into five classes, each class represents a set of Mannings n coefficients that generate comparable water elevation data. These classes and their produced mean water elevation (e.g., at observation station 1) are summarized in Table 1.

Table 1 Mannings n coefficients in idealized inlet with ebb shoal classified by amplitude of corresponding tidal data at station 1.

Class	Mean of tidal amplitude (m)	Mannings n
A	0.267989-0.214391	0.005-0.025
B	0.214391-0.160793	0.03-0.04
C	0.160793-0.107195	0.045-0.07
D	0.107195-0.053598	0.075-0.135
E	0.053598-0	0.14-0.2

For the two parameters case, α and β , the classification of Manning's n coefficients cannot be achieved as was done in the single coefficient case without substantial computation since it is not obvious which pair of α and β yields the largest mean tidal amplitude.

Furthermore, different combinations of α and β may also result in similar maximum water elevations, and the solution of the ADCIRC is non-monotonic with respect to the two constants Manning's n coefficients. Instead, we select a few pairs of α and β values that produce maximum tidal amplitudes that are within 20% of those amplitudes that result from ADCIRC when it is integrated with the true two constants parameterized Manning's n coefficients as the initial guesses. We chose a single true field $n_{0.005,0.1}$ in which the coefficient is set to 0.005 in the deep water of the open ocean and 0.1 in the shallow bay area. The selected true Manning's n coefficients represent suitable bottom stress of the deep water ($\alpha = 0.005$) and shallow water ($\beta = 0.1$) that are commonly used in coastal ocean models, e.g. [10, 14].

3.5 Observation System Simulation Experiments (OSSEs)

Observation system simulation experiments (OSSEs) are conducted in this study. Synthetic observations are obtained by running ADCIRC simulations with specified initial conditions and Manning's n coefficients. The specified Manning's n coefficients are considered to be the true parameters, and we attempt to recover them from an inaccurate specification based on the extracted data. We initiate the model simulation with some given estimate of Manning's n coefficients, and then apply the SEIK filter and test for convergence of the updated parameters toward the true parameters after the subsequent assimilation cycles. We also apply the MCMC method using a PC surrogate of ADCIRC in the same OSSE setting to directly sample the posterior pdf of the Manning's n coefficients, and use its results as the reference solution to evaluate the SEIK estimates. We are interested not only in the parameter estimates, but also in their estimated uncertainties, and more generally in the posterior pdfs. We further investigate the OSSEs results with uniform and (more challenging) Gaussian priors.

3.5.1 Uniform prior

Here the prior $p_{pr}(\theta)$ in (1) follows a uniform distribution $\mathcal{U}(0.005, 0.2)$, corresponding to the allowed range of Manning's n coefficients of ADCIRC. The parameter inference schemes (i.e. MCMC and the SEIK filter) are initialized by sampling this prior.

For the single coefficient case, we set the true Manning's n coefficient to a value from the middle of each of the classes as described in the previous section. The values we attempt to estimate are 0.015, 0.035, 0.06, 0.105 and 0.17. For each target Manning's n coefficient, we first generate the synthetic water elevation data at 15 pre-defined locations representing the 15 observation stations and perturb them with Gaussian noise of variance 0.01 m. The observation error covariance matrix $\mathbf{R} = \delta^2 \mathbf{Id}$, with $\delta^2 = 0.01$ and $\mathbf{Id}$ is the identity matrix. For PC-MCMC inference, we initiate the MCMC chain from the middle value of Manning's n coefficient $n = 0.1025$ and sample directly from the uniform prior $\mathcal{U}(0.005, 0.2)$. For the EnKF inference, we initiate the model using $n = 0.1025$ and the initial ensemble members are drawn from the same uniform prior as PC-MCMC. Assimilation starts after a 12 hour ramp up period. We assume the data are available every hour over a 5 days period.

In the two coefficients case, α and β , the OSSEs are performed in the same manner as in the single coefficient case described above. We generate synthetic water elevation data from a given setting of Manning's n coefficients and consider these data as the truth. We then try to recover these values starting from different initial guesses. The inferences are initialized with uniform priors of the two constants Manning's n coefficients α and β as outlined in Table 2.

Table 2 The random input parameters for ADCIRC.

Parameter	Description	Distribution
n_1	Manning's n coefficient in the open ocean	$\mathcal{U}(0.005, 0.2)$
n_2	Manning's n coefficient in the landlocked area	$\mathcal{U}(0.005, 0.2)$

3.5.2 Gaussian prior

In the Gaussian prior setting, which is more common in a filtering/assimilation framework, we invert the same perturbed data generated by the true Manning's n coefficients as in the previous subsection. We follow [62] and initialize the assimilation from Gaussian priors with the initial guesses as means and a small standard deviation equal to 0.01. This choice of prior imitates the situation where the distribution and the support of the prior are not well-known and one has to initialize the inference from some initial guesses. In this case, we allow more flexibility for the choice of prior and attempt to improve the accuracy of the estimation by online tuning of the inference parameters during assimilation.

In the single coefficient case, to estimate a true Manning's n coefficients of a specific class, we initialize the model using a value from any of the four remaining classes as the initial guess. For instance, to recover the true constant Manning's coefficient of 0.17, we may initialize the model with Manning's n coefficients = 0.015, 0.035, 0.06 or 0.105, respectively. This means 4 experiments for every single true constant coefficient we attempt to estimate, for a total of 20 experiments for the 5 considered true Manning's n values, as schematized in Fig. 2.

Fig. 2 A schematic representation of one of the true constant Manning's n coefficients ($n = 0.17$) and the different Gaussian priors (with initial guesses from remaining classes as the means) we use in PC-MCMC and SEIK.

For the two parameters case, we synthesize the observations of water elevation by running ADCIRC with $n_{0.005, 0.1}$ and attempt to recover this true parameter field while initializ-

ing our simulation with different initial guesses. The selected initial guesses are $n_{0.005,0.005}$, $n_{0.1,0.1}$, $n_{0.06,0.06}$ and $n_{0.1,0.005}$. Thus we evaluate the results from incorrect initial guesses of either α , β , or α and β .

3.6 Construction of the PC surrogate

To cast ADCIRC uncertainties as stochastic variables, we let $\mathbf{n} = (n_1, n_2)^T$ be the vector of random model inputs having uniform distribution as specified in Table 2. The inputs n_i are parameterized by $\xi_i \sim \mathcal{U}(-1, 1)$ through

$$n_i(\xi) = \mu_i + \sigma_i \xi_i, \quad i = 1, 2 \quad (17)$$

where $\mu_i = (1/2)(a_i + b_i)$, and $\sigma_i = (1/2)(b_i - a_i)$, so that $p_i \sim \mathcal{U}(a_i, b_i)$ as in Table 2. Here a_i and b_i represent the pre-determined minimum and maximum values of a particular parameter, respectively. At a given time t and a point $\mathbf{x} \in \mathcal{G}$ and given a vector of random inputs $\mathbf{n}(\xi)$, the model output is represented as

$$\mathbf{X}(t, \mathbf{x}, \xi) = \mathbf{A}(t, \mathbf{x}; \mathbf{n}(\xi)), \quad (18)$$

where $\mathbf{A}(t, \mathbf{x}; \mathbf{n}(\xi))$ is the output of a deterministic ADCIRC simulation solved at time t and point $\mathbf{x}$ with the input parameters $\mathbf{n}(\xi)$. Here, $\mathbf{X}$ may correspond to any QoIs (e.g., water height at observation station).

We use NISP with 6 points Gaussian quadrature rule to compute the spectral expansion of the model output in the PC basis. With full tensorization in two stochastic dimensions (i.e. piece-wise Manning's n coefficients), the construction of the PC model requires 36 model runs. To confirm that the quadrature ensures good representation of ADCIRC in this specific case, we generate the PC approximation of the model for different PC order 1 to 6 and then assess the convergence of these representations. This is shown in Fig. 3 along with the convergence to the PC coefficients at a selected station and time. According to these results, the distribution seems to level off at the PC order level $m = 6$, suggesting that a sixth-order expansion is sufficient. In addition, we also compute the RMSE between the water elevation from ADCIRC and its PC-surrogate. The RMSE is less than 5% of the wave amplitude for all varying Manning's n coefficients. The error values are summarized in Table 3. Note that when using PC methods, it is useful to ensure that the posterior distribution is not in a low probability region of the prior, as the quality of the surrogate may deteriorate in such region. This is not an issue in our case because the prior is uniform, and the approximation is suitable for the entire range of the prior.

Table 3 RMSEs between the true model and the surrogate model (water elevations) for the whole simulation window at a specific observation station and varying Manning's n coefficients.

Manning's n coefficients	RMSE (at station 1)
Constant Manning's $n = 0.005$	0.0053
two constants Manning's $\alpha = 0.063$ and $\beta = 0.030$	0.0011
two constants Manning's $\alpha = 0.175$ and $\beta = 0.063$	0.00036
two constants Manning's $\alpha = 0.030$ and $\beta = 0.142$	0.00052
two constants Manning's $\alpha = 0.142$ and $\beta = 0.175$	0.00054
Constant Manning's $n = 0.2$	0.0024

Fig. 3 Left panels: Distribution of water elevation generated by random Manning's n coefficients at the observed locations at several time steps t , as indicated. Different scales are used in each panels to facilitate comparison of the different curves. p denotes the highest polynomial degree in the truncated expansion. In each case, the PC expansion was sampled 10^6 times to generate the distribution curve. Right panels: PC coefficients at observed locations at several time steps t .

4 Numerical Results

We present the PC-MCMC and SEIK inference results for the two different prior settings; uniform $\mathcal{U}(0.005, 0.2)$ and Gaussian with shifted means and a small variance. For both settings, we start by assuming the results of the PC-MCMC inference and then use these as the reference to evaluate those of SEIK.

4.1 Manning's n inference from uniform prior

4.1.1 PC-based MCMC Inference with uniform prior

We first infer the Manning's n coefficients using PC-surrogate ADCIRC with MCMC. The posterior distribution of estimated Manning's n coefficients are generated using 200,000 MCMC iterations. The results of estimating single constant Manning's n coefficients are presented in Fig. 4. It is shown that MCMC accurately recovers the true Manning's n coefficient (Table 4). The mode of the pdf well represents the true Manning's value. The posterior pdf as sampled by MCMC exhibits a shape close to a single mode Gaussian pdf. The true Manning's n coefficients falls within the 95% confidence intervals as inferred from these pdfs.

The inference of two constants piece-wise Manning's n coefficients are shown in Fig. 5. As with the single coefficient case, PC-MCMC accurately estimates the true Manning's n coefficients for all tested values of α and β , as outlined in Table 5. The estimated posterior for α is not Gaussian but positively skewed. The mass of the distribution is concentrated near $\alpha = 0.005$ in all cases. This shape results from the construction of PC surrogate which assumes a uniform prior with a minimum value of 0.005. In all test cases, the estimated posteriors for β approximately exhibits Gaussian shape with the true Manning's n coefficients falling within the 95% confidence intervals of the inferred pdfs, with exceptions of some cases where the pdf exhibits bimodal distribution (e.g., when the initial guess for β is 0.1, which is the true β). For both the single coefficient and two coefficients cases, the estimates produced by PC-MCMC is clearly not sensitive to the choice of initial guesses, which is expected for long enough MCMC chains.

4.1.2 SEIK Inference with uniform prior

We infer the same set of true Manning's n coefficients as in the previous section using the SEIK filter with initial ensembles sampled from the uniform prior $\mathcal{U}(0.005, 0.2)$. After the model ramp up, these initial ensembles are updated following the SEIK filter procedure as described in section 2.2. Ensembles of 10 members were used in these experiments. The results of the SEIK filter inference compared to PC-MCMC for the single and two constants coefficients cases are shown in Fig. 6 and 7, respectively.

For the single coefficient case, the SEIK filter accurately recovers the true Manning's n coefficients for all classes, except for class D (i.e. $n = 0.105$) where the SEIK filter slightly underestimates the truth. The filter estimate is however within class D and is therefore considered acceptable. The posterior pdfs inferred by the SEIK filter are narrow compared to those computed by PC-MCMC. This suggests that the SEIK inference underestimates the uncertainty about the inferred parameter compared to the gold-standard MCMC, which is optimal given the same prior.

The inference of the two constant coefficients case shows similar results. $\alpha = 0.005$ and $\beta = 0.1$ are accurately recovered, but with a narrow posterior pdfs compared to the posterior pdfs resulting from PC-MCMC. In addition, because of the Gaussian assumption, the SEIK inferred posterior fails to capture the MCMC posterior's heavy tail and further assigns positive probability for $\alpha < 0.005$.

In summary, this section shows that initializing the SEIK filter with uniform priors that cover the true Manning's n coefficients allows to accurately recovering the true value of the Manning's n coefficient in the single and two coefficients cases using only 10 ensemble members. The SEIK filter however underestimates the posteriors spread in all cases (as

Fig. 4 Posterior distributions of the constant Manning's n coefficients as estimated by PC-MCMC. The green dashed line represents the true value and the blue dashed line the estimated Manning's n coefficients, as the mode of the posterior pdf (red curve). Note that the estimates do not depend on the choice of initialized value of the MCMC chains if the MCMC chains are sufficiently long.

shown in Tables 4 and 5). The underestimation of the filter solution uncertainty will be addressed in section 4.3.

4.2 Inference sensitivity to a more challenging Gaussian priors

The OSSEs are conducted here as described in section 3.5.2. For each true Manning's n coefficient, we first generate ADCIRC outputs from which the observations are extracted. These data are then used both in the PC-MCMC and the SEIK filter to infer the truth starting from inaccurate initial guesses.

Fig. 5 Posterior distributions of the piece-wise constant Manning's n coefficients as estimated by PC-MCMC. The green dashed line represents the true value and the blue dashed line the estimated Manning's n coefficients, as the mode of the posterior pdf (red curve). Note that the estimates do not depend on the choice of initialized value of the MCMC chains if the MCMC chains are sufficiently long.

Table 4 Final estimates, RMSEs between the estimates and the truths and pdfs variance of constant Manning's n coefficients as estimated by PC-MCMC.

Measures	Truth	Inference method	
		PC-MCMC	SEIK
Final estimate	0.015	0.014998	0.015011
	0.035	0.035344	0.035068
	0.06	0.061107	0.059852
	0.105	0.105986	0.101681
	0.17	0.171362	0.171591
RMSE	0.015	0.001072	1.1363e-05
	0.035	0.000456	6.8020e-05
	0.06	0.001092	1.4765e-04
	0.105	0.001013	0.003318
	0.17	0.002637	0.001591
Variance	0.015	5.4278e-05	7.0412e-09
	0.035	3.6691e-05	4.0865e-08
	0.06	0.000165	1.2371e-07
	0.105	0.000140	5.0084e-07
	0.17	0.000555	1.7129e-07

4.2.1 PC-based MCMC inference with Gaussian priors

We apply PC-MCMC to infer the same set of true Manning's n coefficients as in section 4.1.1, but with Gaussian priors of variance 0.0001. The posterior pdfs of the single and two coefficients cases are presented in Fig. 8 and 9, respectively.

For the single coefficient case, PC-MCMC successfully recovers the truth starting from any initial guess (i.e. prior's mean) and regardless of the class of the true Manning's n coefficient. Our results suggest that considering a sufficiently long MCMC chain (and enough prior variance), PC-MCMC with Gaussian prior converges to the true solution. The posteriors spread, though small, depends on how far the prior mean is from the truth. This suggests that MCMC is sensitive to the choice of the Gaussian prior. In other experiments, not shown here, when the prior is far from the truth and the variance of the prior is very

Fig. 6 Posterior distributions of the constant Manning's n coefficients as estimated by SEIK with 10 ensemble members compared to PC-MCMC. The prior is $\mathcal{U}(0.005, 0.2)$ for both schemes. The green dashed line represents the true value. The black curve is the posterior pdf as estimated by SEIK and the red curve the posterior pdf as estimated by PC-MCMC.

small, MCMC failed to recover the truth. Larger inference window and more iterations may improve MCMC results in this case. We also found that, although the truth is accurately estimated for all classes, the posterior spread is larger when the true parameters belong to the large classes (for instance, the posterior spread of estimating the truth = 0.17 are much larger than those of truth = 0.015). This is due to the different sensitivity of the model to the class of constant Manning's n coefficients as discussed in subsection 3.4.

Similar results are obtained for the two parameters cases (Fig. 9). The truth $n_{0.005, 0.1}$ is accurately estimated with all initial guesses. However, in this case, the posterior spreads are approximately the same. This is because the initial guesses are selected from a set of Manning's n coefficients that produce similar average maximum water elevation as the truth

Fig. 7 Posterior distributions of the two constants Manning's n coefficients as estimated by SEIK with 10 ensemble members compared to PC-MCMC. The prior is $\mathcal{U}(0.005, 0.2)$ for α and β for both schemes. The green dashed line represents the true value. The black curve is the posterior pdf as estimated by SEIK and the red curve the posterior pdf as estimated by PC-MCMC.

Table 5 Final estimates, RMSEs between the estimates and the truths and pdfs variance of piecewise Manning's n coefficients used to parameterize two constants Manning's field as estimated by PC-MCMC. $n_{\alpha, \beta}$ represents the field parameterized by α and β .

Measures	Truth	Initial guess	
		PC-MCMC	SEIK
Final estimate	$\alpha = 0.005$	0.005505	0.005227
	$\beta = 0.1$	0.098813	0.098526
RMSE	$\alpha = 0.005$	0.000605	2.2783e-04
	$\beta = 0.1$	0.000186	0.001473
Variance	$\alpha = 0.005$	7.3667e-06	3.5574e-08
	$\beta = 0.1$	4.8089e-05	3.8540e-07

(i.e. within 20% of the truth). As a result, the effect of the sensitivity of PC-MCMC to the prior in the two coefficients case is not as pronounced as in the single coefficient case.

4.2.2 SEIK inference with Gaussian priors

We apply the SEIK filter to estimate the Manning's n coefficients starting from different initial guesses using 10 ensemble members. The results of the OSSEs are shown in Fig. 10. The SEIK filter accurately recovers the Manning's n coefficient from the correct class in the cases where the truth lies in any of the three smallest classes, A, B, and C, regardless of the class of the initial guesses. Values in class D are only recovered when the initial guesses are from class C and E. For the estimation of a value from class E, the SEIK filter is able to recover this value only when the initial guess is from the neighboring class (i.e. class D).

In SEIK, the initial ensemble is generated by adding Gaussian noise to the initial guess. Then the ensemble is updated after every analysis step. By construction, the filter ensemble represents, approximately and based on Gaussian assumptions, the pdf of the parameter conditional on the available observation. We plot the evolution of the 95% percentile of Manning's n coefficient along with their filter estimates, taken as the mean of the ensemble members (Fig. 10). It is clear that the uncertainties in the parameters decrease as the filter assimilates more data in time, as one may expect. The truth falls within the 95% estimated

Fig. 8 Posterior distributions of the constant Manning's n coefficients as estimated by PC-MCMC initialized with Gaussian priors with different means. Black dashed lines represent the true values. Colored dashed lines are the estimated Manning's n coefficients, or the modes of the posterior distributions. Gray lines and curves are respectively the posterior distributions from uniform prior and their modes. The estimates are not sensitive to the choice of the prior if the MCMC chains are sufficiently long. The spread of the posterior distribution is clearly sensitive to the choice of the prior mean.

confidence interval for low Manning's n coefficients (e.g., 0.015, 0.035 and 0.06). To estimate the parameter in larger classes, a larger ensemble and/or longer simulation window may be required, which we will examine later.

In the second set of OSSEs we estimate a two constants parameterized field of Manning's n coefficients. The point estimation of the two constants parameterized Manning's n coefficients and their uncertainties are shown in Fig. 11. We obtain accurate estimates of the parameters for all initial guesses except for the initial guess $n_{0.005,0.005}$. In this case $\beta = 0.1$ is significantly underestimated. This is similar to the constant coefficient case where the SEIK filter is not very successful at recovering the large Manning's value starting from

Fig. 9 Posterior distributions of the two constants Manning's n coefficients as estimated by PC-MCMC initialized with Gaussian priors with different means. Black dashed lines represent the true values. Colored dashed lines are the estimated Manning's n coefficients, or the modes of the posterior distributions. Gray lines and curves are respectively the posterior distributions from uniform prior and their modes. The estimates are not sensitive to the choice of the prior when the MCMC chains are sufficiently long.

the smallest guesses. However, for each initial estimate, the initial errors with respect to the truth are still reduced by at least 80%. In contrast to PC-MCMC, the posterior pdf produced by SEIK assigns some probability for α values below 0.005 due to Gaussian assumption on the posterior. However, we observe zero probability for negative values, since the pdfs spread around the mean estimates are small. For the two constants parameterized case, the choice of the initial guesses hardly affect the final estimates even when α and/or β for each guesses are vastly differ. We found no specific relation between the estimation accuracy and the choice of initial guess for the two constants coefficients case. This is because all initial guesses produce an average maximum water elevation within 20% of those obtained when the model is integrated with the true parameters. In this case, one may consider all initial guesses to be within the same class as the truth in term of produced water elevation, in analogy to the classification of the single constant coefficients OSSEs.

The accuracy of the estimated water elevation (by MCMC and SEIK) in the single coefficient case is directly related to the accuracy of the inferred roughness parameters. This is due to the monotonic dependence of the water levels on the Mannings n coefficients [62]. Accurate estimates of the Mannings n coefficient produce accurate estimates of water elevations, and vice versa (not shown). Likewise, we noticed that accurate estimates of the water elevation in the two coefficients OSSEs were only obtained in the cases where the Mannings n coefficients α and β were well recovered. Similar findings were reported in [62] to which we refer the reader for more detailed discussion and analysis.

Fig. 12 and 13 compare the posterior pdfs of the Manning's n coefficient as estimated by SEIK with 10 ensemble members and MCMC for a constant field and two constants parameterized field, respectively. The pdfs produced by SEIK are much narrower in comparison to those estimated by MCMC. SEIK also underestimates the truth in most cases compared to MCMC which experiences no difficulty in recovering the true parameters. Furthermore, one can clearly observe the dominant Gaussian shape of the pdfs estimated by SEIK, as one might expect. The MCMC posterior distributions cover the wide range of Manning's n coefficients and their means are clearly centered at the true values. In this specific setting, the SEIK filter with small ensemble sizes does not provide a good description of the uncertainty of the estimated parameters. It also estimates values below the range of possible values as

Fig. 10 Estimates of various Manning's n fields parameterized by a single value using 10 ensemble members. From top to bottom, left to right the true Manning's n value is 0.015, 0.035, 0.06, 0.105 and 0.17. The uncertainty of the estimates are represented by opaque color band (95th percentile) around the mean.

specified by the prior due to its Gaussian assumption, even though we do not observe likely probability for negative values.

4.3 SEIK tuning

4.3.1 Effect of increasing ensemble size

The SEIK filter should benefit from increased ensemble size [39, 79]. However, increasing the number of ensemble members means increased computational cost; for instance doubling the size of the ensemble means doubling the number of required model runs, which in

Fig. 11 Estimates of two constants piecewise Manning’s n coefficients, α (left) and β (right). The uncertainty of the estimates are represented by opaque color band (95th percentile) around the mean.

the case of ADCIRC significantly increases the computational load. This section examines the SEIK filter performance with respect to the ensemble size.

For the constant Manning’s n coefficient case, we use the SEIK filter with ensemble sizes of 10, 20, 100, 200, 300 and 400, and compare the parameter posterior focusing on both the mean of the ensemble and the uncertainty as estimated from the spread of the ensemble and represented by the 95th percentile of the parameter. The posterior distributions estimated with different ensemble sizes and initial guess equal to 0.06 are plotted in Fig. 14.

With 100 ensemble members, the filter performance significantly improves in comparison to using only 10 members. This can be for example clearly seen in the case where we choose the initial guess from class C to estimate the truth in class E (Fig. 14). The final estimate has increased from 0.1510 with 10 members to 0.1645 with 100 members, which has less than 3.5% relative error (Table 7). However, with 100 members, the truth in class E still lies outside the 95% confidence interval of the pdf as estimated from class C. An ensemble size of 300 members or more is needed in order for the estimate to fall within the 95% confidence interval of the estimate (when the initial guess is 0.06). No other significant improvement is observed for increasing ensemble size beyond 300 members. In the case of estimating small Manning’s n coefficients with the guess from a larger class, increasing the ensemble improves the estimate. Nevertheless, using 100 ensemble members is sufficient to recover the truth in a smaller class. Clearly, recovering the truth in larger classes from smaller classes is more accurate with increased ensemble size. We also notice faster convergence toward the truth with increased ensemble sizes. In term of computational complexity, using 100 members costs approximately ten times the execution time of running the SEIK filter using 10 ensemble members. This could be consequent for a large scale model such as ADCIRC. Thus, increasing the ensemble size does not seem to be the most efficient strategy to improve the estimation of low dimensional representations of Manning’s n coefficients. Also, the computational time of PC-MCMC with 200,000 iterations is approximately twice the computational time of SEIK with 10 members. In more realistic applications with a spatially varying Mannings n coefficients, the PC method may become unpractical, and an EnKF-like inference scheme will be more attractive in this case. A summary of the execution time of SEIK inference with 10 and 100 members, and PC-MCMC is provided in Table 6.

For the case of two constants parameterized Manning’s n field, we compare the estimated α and β obtained with 10 and 100 ensemble members. The results are shown in Fig. 15. As expected, we found clear improvement in estimating the parameters, especially β

Fig. 12 Comparison of the posterior pdfs of the smallest and the largest class of constant Manning's n coefficients as estimated by SEIK with 10 ensemble members (at the end of the assimilation period) and MCMC.

Fig. 13 Comparison of the posterior pdfs of two constants Manning's n coefficients as estimated by SEIK (at the end of the assimilation period) and MCMC. Distributions of α are plotted in the left panels and those of β in the right panels

Table 6 The computational complexity of different inference methods. The PC-MCMC and filtering codes are written in matlab. The codes are executed on Intel Xeon(R) CPU X5550 @2.67GHz $\times$ 8 processor and the memory of 23.5 GiB.

Inference method	Execution time
PC-MCMC	21,414.58 sec
SEIK (10 members)	11,176.25 sec
SEIK (100 members)	103,611.67 sec

Table 7 Final estimates, RMSEs between the estimates and the truths and pdfs variance of constant Manning's n coefficients as estimated by SEIK with 100 members starting from initial guess = 0.06.

Measures	Initial guess			
	0.015	0.035	0.105	0.017
Final estimate	0.015003	0.034949	0.093383	0.164566
RMSE	3.3684e-06	5.0988e-05	0.011616	0.005433
Variance	6.3893e-09	3.3541e-08	3.7095e-07	1.1795e-06

when using 100 ensembles, in comparison to using only 10 ensembles, which underestimates the truth for all initial guesses (Table 8). However, the spread of the pdfs in the case of 100 members does not significantly differ from those obtained in the 10 members case. Both estimate a variance of the same order of magnitude (e.g., $O(10^{-8})$ for α and $O(10^{-7})$ for β). PC-MCMC posterior pdfs exhibit much larger variances ($O(10^{-5})$). Compared to PC-MCMC, the SEIK filter accurately recovers the truth, but underestimates the associated uncertainty by as much as two orders of magnitude. This indicates that the SEIK filter fails to accurately describe the uncertainty of the inferred parameters. This likely explains the cause of SEIK's inability to recover some parameter, e.g., Manning's n values from class E. Covariance inflation is expected to improve the SEIK filter performance in that regard. This will be investigated in the next section.

Table 8 Final estimates, RMSEs between the estimates and the truths and pdfs variance of piecewise Manning's n coefficients used to parameterize two constants Manning's field as estimated by SEIK with 100 members. $n_{\alpha,\beta}$ represents the field parameterized by α and β .

Measures	Truth	Initial guess			
		$n_{0.005,0.005}$	$n_{0.06,0.06}$	$n_{0.1,0.1}$	$n_{0.1,0.005}$
Final estimate	$\alpha = 0.005$	0.005481	0.005081	0.005103	0.005211
	$\beta = 0.1$	0.093857	0.099733	0.099327	0.099168
RMSE	$\alpha = 0.005$	0.000481	8.1378e-05	0.000103	0.000211
	$\beta = 0.1$	0.088857	0.094733	0.094327	0.094168
Variance	$\alpha = 0.005$	3.9019e-08	4.8122e-08	4.7767e-08	4.5681e-08
	$\beta = 0.1$	2.4972e-07	4.4970e-07	4.2556e-07	4.1742e-07

Fig. 14 Comparison of the posterior pdfs as estimated by SEIK for constant Manning's n coefficient with three different ensemble sizes and initial guess = 0.06. Left panels: the 95th percentile over time. Right panels: the posterior pdfs at the end of the assimilation period.

4.3.2 The effect of inflation

In the previous section we showed that increasing the ensemble size is effective, however large ensembles are needed to accurately recover the true parameter, resulting in increased computational cost. Another way to enhance the robustness of an EnKF is to inflate the forecast error covariance [6,59]. Covariance inflation is now considered to be an important auxiliary technique for successful implementation of an EnKF.

We perform the same OSSEs as in the previous section for the idealized inlet ebb shoal using 10 ensemble members, with inflation factors, $\rho = 1.1, 1.2, 1.5$ and 2 . Fig. 16 shows the performance of the SEIK filter in recovering the true constant Manning's n coefficient 0.17 from an initial guess of 0.015 with several inflation factors. Significant improvement is observed when using inflation and the largest class of Manning's n coefficient are now accurately recovered from an initial guess from the smallest class. With inflation, the estimates further converge much faster toward the true values; convergence is reached before the 60th assimilation cycle in all cases except for the case with the smallest inflation, $\rho = 1.1$. Inflating the covariance also widens the ensemble spread around the mean, as expected. The range of the posterior pdf remains positive, which is consistent with the range of the Manning's n coefficients in all tested cases. On top of accurately recovering the true Manning's n coefficient, the SEIK filter with a well-tuned inflation significantly improves the posterior pdf. The variances of the posterior pdfs produced by the SEIK filter with inflation are approximately the same order of magnitude as those produced by PC-MCMC. An infla-

Fig. 15 Top row: sensitivity of the posterior distribution as it results from SEIK for the case of piecewise Manning's n coefficient in the open ocean (α) to different ensemble sizes. Left column: the 95th percentile over time, Right panels: the pdf of the posterior at the end of the simulations. Bottom row: comparison of the posterior distribution as it results from SEIK for piecewise Manning's n coefficient in the landlocked area (β) with different ensemble sizes. Left panels: the 95th percentile over time, Right panels: the pdf of the posterior at the end of the simulations. All experiments start from the initial guess $\alpha = 0.1$ and $\beta = 0.1$.

tion factor of 1.2 is sufficient to recover the posterior pdf, i.e. for the truth to fall within the 95% confidence interval of the estimated posterior.

The results of estimating the two constants parameterized Manning's n field using the SEIK filter with inflation are shown compared to using SEIK without inflation in Fig. 17 and 18. We see similar improvements as in the constant coefficient cases. Fig. 19 and 20 compare the posterior pdfs of the Manning's n coefficients as estimated by the SEIK filter with an inflation factor $\rho = 1.2, 1.5$ and 2.0 for the single and two constants parameterized fields. An inflation factor of 1.2 is sufficient for the SEIK filter to accurately recover the true Manning's coefficients from all classes. However, an inflation factor of 2.0 is required in most cases to obtain posterior's variance that is of the same order of magnitude as those produced by PC-MCMC. Using inflation imposes larger ensemble spread and also improves the mean estimates in all cases. However, in the two constant coefficients case, the SEIK filter with inflation assigns probability to negative values of α , because of its Gaussian framework.

5 Discussion and Conclusions

We investigate the effectiveness of estimating the Manning's n coefficients in the coastal ocean model, ADCIRC, using an ensemble Kalman filter, namely the SEIK filter. The fil-

ter performance was evaluated by comparing its estimated posterior distributions to those resulting from an MCMC method. To enable the implementation of MCMC with an idealized ADCIRC coastal ocean model, we resorted to the polynomial chaos (PC) approach to construct a surrogate model of ADCIRC. PC allows likelihoods to be evaluated at a fraction of the time required for a full evaluation with ADCIRC. A non-intrusive spectral projection scheme was used to derive the PC representation of the water elevation of ADCIRC. Using the PC expansion, we created a surrogate model for an idealized inlet with ebb shoal configuration. A degree 6 PC model was found to be adequate for producing the desired model response to changes in the Manning's n coefficients. PC-MCMC avoids the restriction of the Gaussian posteriors required in ensemble Kalman filters.

We conduct OSSEs in which synthetic water elevation data is generated from various fields of Manning's n coefficients. We then attempt to recover these fields from incorrect initial guesses using both PC-MCMC and EnKF methods. We first investigate the sensitivity of the inference methods to the choice of prior. For the uniform prior case with a support covering the range of possible Manning's n coefficients, both PC-MCMC and SEIK successfully recover the true coefficients in the single and two constants Manning's settings. The SEIK filter however produces much smaller posterior's spread compared to PC-MCMC, implying an underestimation of the parameter's uncertainty.

We further explore the sensitivity of the inference to more challenging Gaussian priors. In this case, PC-MCMC still accurately recover the truth in all tested scenarios. However, the posterior's variance varies with the priors, depending on how far the prior's mean is from the truth. The MCMC chains required larger number of iterations to converge when the prior was further away from the truth. For the constant field, the SEIK filter has difficulty recovering the Manning's field of large values from small initial guesses. We show that increasing the number of ensemble members improves the filter results. The accuracy of the filter estimates increases by more than 200% when the ensemble size is increased from 10 to 100 and the truth lies within the 95% confidence interval of the estimate when we use 300 members (or more). For the two constants parameterized field of both domains, the use of a small ensemble size is enough to recover the true parameters in every test case. Since the increase of ensemble size requires more computation, we also investigate the impact of inflation of the background error covariance on the filter performance. In this case, the estimates converge to the truth much faster, even when the filter is implemented with 10 members. The estimates are also accurate regardless of the choice of the initial guesses. The use of inflation widens the posterior distribution of the parameter produced by the SEIK filter and makes it more comparable to that of PC-MCMC.

The SEIK filter shows good capability to accurately recover the true parameters, consistent with the results of [51]. Using inflation improves the estimation, convergence and distribution spread. However, over-inflation can cause the SEIK filter to enable negative Manning's n coefficients due to the Gaussian assumption, and thus the filter needs to be well tuned. In contrast to the study of [66], who estimated 10 parameters of a simple deep convection model using 10,000 ensemble members, we successfully obtained good approximations of posterior pdfs using small ensemble sizes with SEIK (only 10 members) and reasonable inflation.

To conclude, while the PC-based approach coupled with MCMC is well-suited for parameter estimation problems with low dimension stochastic systems, its efficiency may still be limited to real-time coastal ocean forecasting, which often includes a large number of parameters. Sequential data assimilation schemes could provide a robust alternative to compute the pdfs. Despite their inherent Gaussian assumption, our results suggest that the SEIK filter with enough tuning of the ensemble size and the inflation factor may produce reliable

estimates of the pdf mean and variance that are close to those of MCMC. This supports the idea of using an ensemble Kalman-type filter for parameter estimation in the context of large-scale realistic coastal ocean models, and for providing useful information about the associated uncertainties at reasonable computational cost. The parameter estimation problem of large dimensional spatially varying parameters with the SEIK filter will be explored. We will assess the relative merits of this approach by contrasting with more elaborate PC-based methods [56,73].

Acknowledgements This work was supported by the King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia grant number CRG3-2016. C. Dawson also acknowledges support of the Gulf of Mexico Research Initiative Center for Advanced Research on Transport of Hydrocarbons in the Environment (CARTHE).

References

1. Aksoy, A., Zhang, F., Nielsen-Gammon, J.: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Monthly weather review* **134**(10), 2951–2970 (2006)
2. Alexanderian, A., Winokur, J., Sraj, I., Srinivasan, A., Iskandarani, M., Thacker, W., Knio, O.: Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach **16**(3), 757–778 (2012). DOI 10.1007/s10596-012-9286-2
3. Altaf, M., Butler, T., Luo, X., Dawson, C., Mayo, T., Hoteit, I.: Improving short-range ensemble kalman storm surge forecasting using robust adaptive inflation. *Monthly Weather Review* **141**(8), 2705–2720 (2013)
4. Altaf, M., Gharamti, M.E., Heemink, A., Hoteit, I.: A reduced adjoint approach to variational data assimilation. *Computer Methods in Applied Mechanics and Engineering* **254**, 1 – 13 (2013). DOI <http://dx.doi.org/10.1016/j.cma.2012.10.003>
5. Altaf, M., Raboudi, N., Gharamti, M., Dawson, C., McCabe, M., Hoteit, I.: Hybrid vs adaptive ensemble kalman filtering for storm surge forecasting. *AGU Fall Meeting Abstracts* **1**, 3352 (2014)
6. Anderson, J.: An ensemble adjustment kalman filter for data assimilation. *Monthly weather review* **129**(12), 2884–2903 (2001)
7. Anderson, J., Anderson, S.: A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review* **127**(12), 2741–2758 (1999). DOI 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2
8. Andrieu, C., Moulines, É.: On the ergodicity properties of some adaptive mcmc algorithms. *The Annals of Applied Probability* **16**(3), 1462–1505 (2006)
9. Annan, J., Hargreaves, J., Edwards, N., Marsh, R.: Parameter estimation in an intermediate complexity earth system model using an ensemble kalman filter. *Ocean Modelling* **8**(1), 135–154 (2005)
10. Arcement, G.J., Schneider, V.R.: Guide for selecting manning's roughness coefficients for natural channels and flood plains. Tech. rep., USGPO; For sale by the Books and Open-File Reports Section, US Geological Survey, (1989)
11. Besag, J., Green, P., Higdon, D., Mengersen, K.: Bayesian computation and stochastic systems. *Statistical science* pp. 3–41 (1995)
12. Bishop, C.H., Etherton, B.J., Majumdar, S.J.: Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review* **129**(3), 420–436 (2001)
13. Budgell, W.P.: Stochastic filtering of linear shallow water wave processes. *SIAM Journal on Scientific and Statistical Computing* **8**(2), 152–170 (1987). DOI 10.1137/0908027
14. Bunya, S., Dietrich, J., Westerink, J., Ebersole, B., Smith, J., Atkinson, J., Jensen, R., Resio, D., Luetich, R., Dawson, C., et.al.: A high-resolution coupled riverine flow, tide, wind, wind wave, and storm surge model for southern louisiana and mississippi. part i: Model development and validation. *Monthly Weather Review* **138**(2), 345–377 (2010)
15. Burgers, G., van Leeuwen, P.J., Evensen, G.: Analysis scheme in the ensemble kalman filter. *Monthly Weather Review* **126**(6), 1719–1724 (1998). DOI 10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2
16. Butler, T., Altaf, M., Dawson, C., Hoteit, I., Luo, X., Mayo, T.: Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. *Monthly Weather Review* **140**(7), 2215–2231 (2012)

17. Butler, T., Altaf, M., Dawson, C., Hoteit, I., Luo, X., Mayo, T.: Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. *Monthly Weather Review* **140**, 2215–2231 (2012)
18. DeChant, C.M., Moradkhani, H.: Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research* **48**(4) (2012)
19. Derber, J., Rosati, A.: A global oceanic data assimilation system. *Journal of Physical Oceanography* **19**(9), 1333–1347 (1989)
20. Dietrich, J., Bunya, S., Westerink, J., Ebersole, B., Smith, J., Atkinson, J., Jensen, R., Resio, D., Luettich, R., Dawson, C., et al.: A high-resolution coupled riverine flow, tide, wind, wind wave, and storm surge model for southern louisiana and mississippi. part ii: Synoptic description and analysis of hurricanes katrina and rita. *Monthly Weather Review* **138**(2), 378–404 (2010)
21. Dietrich, J., Westerink, J., Kennedy, A., Smith, J., Jensen, R., Zijlema, M., Holthuijsen, L., Dawson, C., Jr, R.L., Powell, M., et al.: Hurricane gustav (2008) waves and storm surge: Hindcast, synoptic analysis, and validation in southern louisiana. *Monthly Weather Review* **139**(8), 2488–2522 (2011)
22. Dimet, F.X.L., Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* **38A**(2), 97–110 (1986). DOI 10.1111/j.1600-0870.1986.tb00459.x
23. Evensen, G.: Inverse methods and data assimilation in nonlinear ocean models. *Physica (D)* **77**, 108–129 (1994)
24. Evensen, G.: Data assimilation: the ensemble Kalman filter. Springer, Verlag (2009)
25. Evensen, G.: The ensemble kalman filter for combined state and parameter estimation. *Control Systems, IEEE* **29**(3), 83–104 (2009)
26. Evensen, G.: The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics* **53**(4), 343–367 (2013)
27. Franssen, H.H., Kinzelbach, W.: Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research* **44**(9) (2008)
28. Gamerman, D., Lopes, H.: Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall/CRC, Boca Raton (2006)
29. Ghanem, R., Red-Horse, J.: Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach. *Fluid Dyn. Res.* **133**, 137–144 (1999)
30. Ghanem, R., Spanos, P.: Stochastic Finite Elements: A Spectral Approach, 2nd edn. Dover, New York (2002)
31. Gharamti, M.E., Hoteit, I., Valstar, J.: Dual states estimation of a subsurface flow-transport coupled model using ensemble kalman filtering. *Advances in Water Resources* **60**, 75 – 88 (2013). DOI <http://dx.doi.org/10.1016/j.advwatres.2013.07.011>
32. Gómez-Hernández, J.J., Journel, A.G.: Joint Sequential Simulation of MultiGaussian Fields, pp. 85–94. Springer Netherlands, Dordrecht (1993). DOI 10.1007/978-94-011-1739-5-8. URL <http://dx.doi.org/10.1007/978-94-011-1739-5-8>
33. Haario, H., Saksman, E., Tamminen, J.: An adaptive metropolis algorithm. *Bernoulli* **7**(2), 223–242 (2001)
34. Hastings, W.K.: Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**(1), 97–109 (1970). DOI 10.1093/biomet/57.1.97
35. Hill, D.: Tidal modeling of glacier bay, alaska - methodology, results, and applications. Tech. rep., Department of Civil, Environmental Engineering, The Pennsylvania State University (2007)
36. Ho, Y., Lee, R.: A bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control* **9**(4), 333–339 (1964). DOI 10.1109/TAC.1964.1105763
37. Höllt, T., Altaf, M.U., Mandli, K.T., Hadwiger, M., Dawson, C.N., Hoteit, I.: Visualizing uncertainties in a storm surge ensemble data assimilation and forecasting system. *Natural Hazards* **77**(1), 317–336 (2015)
38. Hoteit, I., Hoar, T., Collins, N., Anderson, J., Cornuelle, B., Heimbach, P.: A mitgcm/dart ocean analysis and prediction system with application to the gulf of mexico. *AGU Fall Meeting Abstracts* **1**, 1288 (2008)
39. Hoteit, I., Pham, D.T., Blum, J.: A simplified reduced order kalman filtering and application to altimetric data assimilation in tropical pacific. *Journal of Marine Systems* **36**(12), 101 – 127 (2002). DOI [http://dx.doi.org/10.1016/S0924-7963\(02\)00129-X](http://dx.doi.org/10.1016/S0924-7963(02)00129-X)
40. Hoteit, I., Pham, D.T., Gharamti, M., Luo, X.: Mitigating observation perturbation sampling errors in the stochastic enkf. *Monthly Weather Review* (2015) (2015)
41. Hoteit, I., Triantafyllou, G., Korres, G.: Using low-rank ensemble kalman filters for data assimilation with high dimensional imperfect models. *JNAIAM* **2**(1-2), 67–78 (2007)

42. J. Besag P. Green, D.H., Mengersen, K.: Bayesian computation and stochastic systems. *Stat. Sci.* **10**, 3–41 (1995)
43. Jelesnianski, C.P.: Numerical computations of storm surges without bottom stress. *Monthly Weather Review* **94**(6), 379–394 (1966)
44. Kaipio, J.P., Somersalo, E.: Statistical inversion theory. *Statistical and Computational Inverse Problems* pp. 49–114 (2005)
45. Kalman, R.: a new approach to linear filtering and prediction problems 1. *Transactions of the ASME-Journal of Basic Engineering*, **82**(Series D.), 35–45 (1960)
46. Kennedy, A.B., Gravois, U., Zachry, B.C., Westerink, J.J., Hope, M.E., Dietrich, J.C., Powell, M.D., Cox, A.T., Luettich, R.A., Dean, R.G.: Origin of the hurricane ike forerunner surge. *Geophysical Research Letters* **38**(8) (2011)
47. Kennedy, M., OHagan, A.: Bayesian calibration of computer models. *J. of the Royal Stat. Soc. Series B (Stat. Method.)* **63** (2001) **63**, 425–464 (2011)
48. Kinnmark, I.P., Gray, W.G.: The 2x-test: A tool for analyzing spurious oscillations. *Advances in Water Resources* **8**(3), 129 – 135 (1985). DOI [http://dx.doi.org/10.1016/0309-1708\(85\)90053-3](http://dx.doi.org/10.1016/0309-1708(85)90053-3)
49. Kivman, G.A.: Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics* **10**(3), 253–259 (2003)
50. Knio, O., Maitre, O.: Uncertainty propagation in cfd using polynomial chaos decomposition. *Physica D* **38**, 616–640 (2006)
51. Law, K., Stuart, A.: Evaluating data assimilation algorithms. *Monthly weather review* **140**(11), 37573782 (2012)
52. Le Maitre, O., Knio, O.: *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*. Springer, Berlin (2010)
53. Le Maitre, O., Knio, O., Najm, H., Ghanem, R.: A stochastic projection method for fluid flow. i. basic formulation. *Journal of Computational Physics* **173**, 481–511 (2001)
54. Le Maitre, O., Najm, H., Pbay, P., Ghanem, R., Knio, O.: Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM Journal Scientific Computation* **29**(2), 864–889 (2007)
55. van Leeuwen, P.J.: Particle filtering in geophysical systems. *Monthly Weather Review* **137**(12), 4089–4114 (2009)
56. Li, G., Iskandarani, M., Hénaff, M.L., Winokur, J., Le Maître, O.P., Knio, O.M.: Quantifying initial and wind forcing uncertainties in the gulf of mexico. *Computational Geosciences* pp. 1–21 (2016). DOI [10.1007/s10596-016-9581-4](https://doi.org/10.1007/s10596-016-9581-4)
57. Luettich, R., Westerink, J.: Formulation and numerical implementation of the 2D/3D ADCIRC finite element model version 44. XX. R. Luettich (2004)
58. Luettich Jr, R., Westerink, J., Scheffner, N.W.: Adcirc: An advanced three-dimensional circulation model for shelves, coasts, and estuaries. report 1. theory and methodology of adcirc-2ddi and adcirc-3dl. Tech. rep., COASTAL ENGINEERING RESEARCH CENTER VICKSBURG MS (1992)
59. Luo, X., Hoteit, I.: Robust ensemble filtering and its relation to covariance inflation in the ensemble kalman filter. *Monthly Weather Reviews* **139**, 39383953 (2011)
60. Lynch, D.R., Gray, W.G.: A wave equation model for finite element tidal computations. *Computers & Fluids* **7**(3), 207 – 228 (1979). DOI [http://dx.doi.org/10.1016/0045-7930\(79\)90037-9](http://dx.doi.org/10.1016/0045-7930(79)90037-9)
61. Martino, L., Míguez, J.: A generalization of the adaptive rejection sampling algorithm. *Statistics and Computing* **21**(4), 633–647 (2010). DOI [10.1007/s11222-010-9197-9](https://doi.org/10.1007/s11222-010-9197-9)
62. Mayo, T., Butler, T., Dawson, C., Hoteit, I.: Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of mannings friction coefficient. *Ocean Modelling* **76**, 43–58 (2014)
63. Najm, H., Debusschere, B., Marzouk, Y., Widmer, S., Matre, O.L.: Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *International Journal of Numerical Methods in Engineering* **80**(6), 789–814 (2009)
64. Pham, D.: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly weather review* **12**(5), 1194–1207 (2001)
65. Phenix, B., Dinaro, J., Tatang, M., Tester, J., Howard, J., McRae, G.: Incorporation of parametric uncertainty into complex kinetic mechanisms: application to hydrogen oxidation in supercritical water. *Combustion and Flame* **112**, 132–146 (1998)
66. Posselt, D.J., Bishop, C.H.: Nonlinear parameter estimation: Comparison of an ensemble kalman smoother with a markov chain monte carlo algorithm. *Monthly Weather Review* **140**(6), 1957–1974 (2012)
67. Rasmussen, C.E.: *Gaussian processes for machine learning*. MIT Press (2006)
68. Robert P. Casella, G.: Monte carlo statistical methods. *Applied Mathematical Sciences* **160**(ISBN 0-387-22073-9), 344 (2004)

69. Roberts, G.O., Gelman, A., Gilks, W.R.: Weak convergence and optimal scaling of random walk metropolis algorithms. *Ann. Appl. Probab.* **7**(1), 110–120 (1997). DOI 10.1214/aoap/1034625254
70. Serafy, G.Y.H.E., Mynett, A.E.: Improving the operational forecasting system of the stratified flow in osaka bay using an ensemble kalman filter-based steady state kalman filter. *Water Resource Research* **44**, W06,416, doi:10.1029 2006WR005,412 (2008)
71. Song, H., Hoteit, I., Cornuelle, B.D., Luo, X., Subramanian, A.C.: An adjoint-based adaptive ensemble kalman filter. *Monthly Weather Review* **141**(10), 3343–3359 (2013)
72. Sraj, I., Iskandarani, M., Srinivasan, A., Thacker, W.C., Winokur, J., Alexanderian, A., Lee, C.Y., Chen, S.S., Knio, O.M.: Bayesian inference of drag parameters using axbt data from typhoon fanapi. *Monthly Weather Review* **141**(7), 2347–2367 (2013)
73. Sraj, I., Le Maître, O., Knio, O., Hoteit, I.: Coordinate transformation and polynomial chaos for the bayesian inference of a gaussian process with parametrized prior covariance function. *Computer Methods in Applied Mechanics and Engineering* **298**, 205–228 (2016)
74. Sraj, I., Mandli, K., Knio, O., Dawson, C., Hoteit, I.: Uncertainty quantification and inference of manning's friction coefficients using dart buoy data during the tōhoku tsunami. *Ocean Modelling* **83**, 82–97 (2014)
75. Sraj, I., Zedler, S., Knio, O., Jackson, C., Hoteit, I.: Polynomial chaos-based bayesian inference of k-profile parameterization in a general circulation model of the tropical pacific. *Monthly Weather Review* **0**(0), null (2016). DOI 10.1175/MWR-D-15-0394.1
76. Tagade, P.M., Choia, H.L.: A generalized polynomial chaos-based method for efficient bayesian calibration of uncertain computational models. *Inverse Problems in Science and Engineering* **22**(4), 602–624 (2014)
77. Tarantola, A.: Inverse problem theory and methods for model parameter estimation. *siam* (2005)
78. Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M., Whitaker, J.S.: Ensemble square root filters. *Monthly Weather Review* **131**(7), 1485–1490 (2003)
79. Triantafyllou, G., Hoteit, I., Petihakis, G.: A singular evolutive interpolated kalman filter for efficient data assimilation in a 3-d complex physical–biogeochemical model of the cretan sea. *Journal of Marine Systems* **40**, 213–231 (2003)
80. Westerink, J.J., Luettich, R.A., Feyen, J.C., Atkinson, J.H., Dawson, C., Roberts, H.J., Powell, M.D., Dunion, J.P., Kubatko, E.J., Pourtaheri, H.: A basin-to channel-scale unstructured grid hurricane storm surge model applied to southern louisiana. *Monthly Weather Review* **136**(3), 833–864 (2008)
81. Yanagi, T.: Coastal oceanography, Vol. 1. Springer (1999)

Fig. 16 Comparison of the posterior distribution as it results using SEIK for constant Manning's n coefficient between SEIK without and with inflation factor (1.1, 1.2, 1.5 and 2.0 respectively), initial guess = 0.015. Left panels: the 95th percentile over time, Right panels: the pdf of the posterior at the end of the simulations.

Fig. 17 Comparison of the posterior distribution as it results using SEIK for two constant piecewise Manning's n coefficients between SEIK with and without inflation in the open ocean area. Left panels: the 95th percentile over time, Right panels: the posterior pdfs at the end of the assimilation period.

Fig. 18 Comparison of the posterior pdfs as it results using SEIK for two constant piecewise Manning's n coefficients between SEIK with and without inflation in the land-locked area. Left panels: the 95th percentile over time, Right panels: the posterior pdfs at the end of the assimilation period.

Fig. 19 Comparison of the posterior pdfs of classes of constant Manning's n coefficients as estimated by SEIK with different inflation (at the end of assimilation period) and MCMC.

Fig. 20 Comparison of the posterior pdfs of two constant piecewise Manning's n coefficients as estimated by SEIK (at the end of the assimilation period) and MCMC. Right panels: the distributions of α , Left panels: the distribution of β .