Journal of Hydrology 562 (2018) 664–684

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Ensemble Kalman filter inference of spatially-varying Manning's *n* coefficients in the coastal ocean

Adil Siripatana[a], Talea Mayo[b], Omar Knio[c], Clint Dawson[d], Olivier Le Maître[e], Ibrahim Hoteit[a,*]

[a] Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia
[b] Department of Civil, Environmental and Construction Engineering, University of Central Florida, 4000 Central Florida Blvd, Orlando, 32816 FL, USA
[c] Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia
[d] Institute for Computational Engineering and Science, University of Texas at Austin, 201 E 24th St. Stop C0200, Austin, 78712-1229 TX, USA
[e] Mechanics and Energetic Department, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, LIMSI-CNRS Campus universitaire bât 508 Rue John von Neumann F-91405 Orsay cedex, Paris 91405, France

ARTICLE INFO

This manuscript was handled by G. Syme, Editor-in-Chief

Keywords:
Data assimilation
Singular evolutive interpolated Kalman filter
Manning's *n* coefficients
ADvanced CIRCulation (ADCIRC) model
Uncertainty quantification

ABSTRACT

Ensemble Kalman (EnKF) filtering is an established framework for large scale state estimation problems. EnKFs can also be used for state-parameter estimation, using the so-called "Joint-EnKF" approach. The idea is simply to augment the state vector with the parameters to be estimated and assign invariant dynamics for the time evolution of the parameters. In this contribution, we investigate the efficiency of the Joint-EnKF for estimating spatially-varying Manning's *n* coefficients used to define the bottom roughness in the Shallow Water Equations (SWEs) of a coastal ocean model.

Observation System Simulation Experiments (OSSEs) are conducted using the ADvanced CIRCulation (ADCIRC) model, which solves a modified form of the Shallow Water Equations. A deterministic EnKF, the Singular Evolutive Interpolated Kalman (SEIK) filter, is used to estimate a vector of Manning's *n* coefficients defined at the model nodal points by assimilating synthetic water elevation data. It is found that with reasonable ensemble size ($\boldsymbol{O}(10)$), the filter's estimate converges to the reference Manning's field. To enhance performance, we have further reduced the dimension of the parameter search space through a Karhunen-Loéve (KL) expansion. We have also iterated on the filter update step to better account for the nonlinearity of the parameter estimation problem. We study the sensitivity of the system to the ensemble size, localization scale, dimension of retained KL modes, and number of iterations. The performance of the proposed framework in term of estimation accuracy suggests that a well-tuned Joint-EnKF provides a promising robust approach to infer spatially varying seabed roughness parameters in the context of coastal ocean modeling.

1. Introduction

Simulation of ocean waves, tides, and estuarine and coastal floodplain inundation is crucial for various maritime-related activities, coastal resources management, planning, and sustenance (Yanagi, 1999). In particular, accurate storm-surge forecasting during extreme events may considerably improve the chance of protecting lives and coastal infrastructures, which ultimately benefit the global community, both economically and ecologically (e.g. Jelesnianski, 1966; Bunya et al., 2010; Dietrich et al., 2010; Dietrich et al., 2011).

The shallow water equations (SWEs), derived from depth-integrating the Navier-Stokes equations, have been widely used in coastal ocean modeling. They assume that the horizontal length scale of the problem domain is much larger than the vertical length scale under hydrostatic pressure (Luettich et al., 1992; Vreugdenhil, 1994). In real world applications, the numerical solution of the SWEs is subject to various sources of uncertainty, such as modeling errors, numerical discretization, inputs uncertainty, etc. In particular, the uncertainty associated with the poor characterization of the model parameters is considered a major source of error (Brummelhuis and Heemink, 1990; Yanagi, 1999; Sorensen and Madsen, 2006). A number of recent studies have therefore focused on quantifying and reducing the uncertainties associated with input parameters, aiming to achieve more reliable forecasts in fluid flow modeling (e.g. Ghanem and Red-Horse, 1999; Gharamti et al., 2014; Sraj et al., 2014; Giraldi et al., 2017). In coastal ocean modeling, the specification of a parameter called "the Manning's *n* coefficient of roughnes", used to define the bottom stress components in the SWEs, is particularly important (Ishii and Kawahara, 2006; Mayo

* Corresponding author.
E-mail addresses: adil.siripatana@kaust.edu.sa (A. Siripatana), talea.mayo@ucf.edu (T. Mayo), omar.knio@kaust.edu.sa (O. Knio), clint@ices.utexas.edu (C. Dawson), olm@limsi.fr (O.L. Maître), ibrahim.hoteit@kaust.edu.sa (I. Hoteit).

https://doi.org/10.1016/j.jhydrol.2018.05.021

Available online 16 May 2018
0022-1694/ © 2018 Elsevier B.V. All rights reserved.

et al., 2014; Siripatana et al., 2017).

The Manning's n coefficient is an empirically derived parameter, defined as the resistance to water flow due to bottom surface characteristics (e.g., sands, rocks and reefs etc.). It is used to describe multiple types of resistance, e.g. friction resistance, form resistance, wave resistance, and resistance of flow instability (Arcement and Schneider, 1989; Ding et al., 2004). It enters the SWEs via the momentum equations, and the amplitude of the water column at a given point in the model domain can be highly sensitive to its value (Mayo et al., 2014). The Manning's n coefficient cannot be measured directly (Budgell, 1987) and often exhibits spatially heterogeneous variability. It also depends on the ocean bottom surface characteristics; changes in the ocean floor during extreme events (such as storm surges and tsunamis) may further alter the near-shore Manning's n field. In such hazardous scenarios, it is critical that changes in ocean bottom stress be detected and updated to accurately predict water height. Unfortunately, the acquisition of the complete knowledge of Manning's n coefficients in realistic settings is not feasible. Parameter identification by trial-and-error, e.g. comparing the SWEs solution produced by different Manning's fields to observations, is tedious and impractical (Khatibi et al., 1997). As a consequence, parameter specification methods, based on established look-up tables for each land cover type and roughness, have been commonly used to parameterize Manning's n fields in large scale coastal ocean models (Luettich et al., 1992; Luettich and Westerink, 2005; Bunya et al., 2010). A more advanced specification method based on a random forest model was also proposed in Medeiros et al. (2015). In this paper, we resort to a well-established inverse modeling approach (Mayo et al., 2014) to infer spatially-varying Manning's n coefficients.

A number of approaches have been developed to solve parameter estimation problems in the context of meteorology and oceanography (e.g. Navon, 1998; Posselt and Bishop, 2012; Gharamti et al., 2013; Altaf et al., 2014). Many are originally motivated by optimal control theory (Lions, 1971), and are based on the minimization of a cost function penalizing discrepancies between model outputs and observations (Buckley, 1978; Gilbert and Lemaréchal, 1989; Nash, 2000; Altaf et al., 2016). However, this approach can be computationally demanding and typically requires the development of an adjoint model (Evensen, 2013; Altaf et al., 2013; Hoteit et al., 2013). Another popular approach for parameter inference is through the Bayesian framework (Ho and Lee, 1964; Besag et al., 1995), where the parameters are represented with probability density functions (pdfs) conditioned on available data. The parameter inference problem is then viewed as the transformation of a prior pdf to a posterior pdf by incorporating the likelihood of the observations (Tarantola, 2004). The posterior is rarely explicit and often needs to be sampled as a collection of realizations that are consistent with data and prior information (Hoteit, 2008). The most popular implementation of this method is the Markov Chain Monte Carlo (MCMC) method (e.g., Hastings, 1970; Gamerman and Lopes, 2006), which has become more practical in recent years with increases in computational power. The primary advantage of MCMC is the ability to produce a full approximation of the posterior distribution. As a result, MCMC is often treated as the benchmark to evaluate the performance of other parameter inference methods (Law and Stuart, 2012; Posselt and Bishop, 2012; Siripatana et al., 2017).

In order to obtain good resolution of the posterior distribution, a large number of samples are required (Sraj et al., 2014; Giraldi et al., 2017). This makes MCMC very computationally demanding, as each MCMC iteration requires a model evaluation in order to compute the likelihood. As a result, using MCMC for parameter estimation is often too costly for a realistic large scale inference problem. Even with model reduction techniques, e.g., Polynomial chaos, KL expansions, etc., parameter estimation in MCMC may still be quite computationally prohibitive.

Bayesian inference can also be cast as a filtering problem in which the posterior distribution is updated sequentially as data becomes available (Evensen, 2009), an approach known as data assimilation. A Bayesian filter operates as a succession of forecast steps to propagate the pdf of the unknowns forward in time, and update steps to incorporate data every time new observations become available. For parameter estimation, filtering schemes usually apply the standard augmented state-parameter technique (Derber and Rosati, 1989; Anderson, 2001; Ait-El-Fquih et al., 2016), that allow the state and parameters of the system to be estimated concurrently. Currently, the most popular approach for data assimilation into ocean models is the Ensemble Kalman Filter (EnKF) (Burgers et al., 1998; Evensen, 2009) and its deterministic versions (Bishop et al., 2001; Anderson, 2001; Hoteit et al., 2002; Tippett et al., 2003; Hoteit et al., 2015, to cite but a few). An EnKF follows a Monte Carlo framework to integrate an ensemble of model realizations in the forecast step and then applies a linear Kalman correction in the update step (Hoteit et al., 2008). The stochastic EnKF assimilates perturbed observations and this was shown to induce noise in the final solution when the filter is implemented with small ensembles (Altaf et al., 2014; Hoteit et al., 2015). Deterministic EnKFs, which avoid observations perturbations, mainly differ in the way they sample the new analysis ensemble after the filter update step. Various deterministic EnKFs were compared with a realistic setting of ADvanced CIRCulation (ADCIRC) model in the Gulf of Mexico (Altaf et al., 2014), showing that, with enough tuning, these filters performed closely well, all outperforming the stochastic EnKF.

The primary advantage of EnKF-type techniques over MCMC is the algorithmic ability to directly accommodate the estimation of large dimensional state-parameter vectors (Evensen, 1994; Naevdal et al., 2005; Hendricks Franssen and Kinzelbach, 2008; Tong et al., 2010; DeChant and Moradkhani, 2012; ElSheikh et al., 2013). Furthermore, these methods are non-intrusive, i.e. they require no modifications to the model code. Despite their empirical Gaussian framework (Hoteit et al., 2012; Liu et al., 2016), EnKF methods have been found to be efficient in terms of performance, computational cost, and robustness in handling ocean state estimation problems (e.g., Serafy and Mynett, 2008; Butler et al., 2012; Korres et al., 2012; Tamura et al., 2014; Altaf et al., 2014). There is now increasing interest in the coastal ocean community to apply EnKF methods to parameter estimation problems. Mayo et al. (2014) and Siripatana et al. (2017) have demonstrated that the EnKFs are able to provide very good estimates of low-dimensional parameterizations of Manning's n coefficients in the SWEs.

In this study, we are interested in the inference problem of a 2D spatially varying Manning's n coefficient. The approach we follow resembles that of ElSheikh et al. (2013), which consists of a sequence of methods to formulate the inference of parameters, including a statistical parameterization of the parameter search space, the construction of a synthetic parameter field, the generation of an initial (prior) ensemble, the implementation of a model reduction technique, and finally the application of a parameter inference method. We generate realizations of 2D spatial maps of Manning's n coefficients subjected to a few synthetic observations based on the sequential simulation algorithm of multi-Gaussian fields (Gómez-Hernández and Journel, 1993). A reference field and an initial ensemble are then selected from these realizations. Next, we apply the Singular Evolutive Interpolated Kalman (SEIK) filter, a deterministic EnKF (Pham, 2001; Hoteit et al., 2002; Butler et al., 2012), to estimate the reference Manning's n field using the Joint-EnKF. Localization (Houtekamer and Mitchell, 2001; Anderson, 2003) is also applied to enable efficient implementation of the SEIK with reasonable ensemble sizes and to remove any spurious correlations between distant points. To limit the parameter search space, and impose some regularization on the inferred model, a truncated Karhunen-Loéve (KL) series is constructed by applying a singular value decomposition on the covariance matrix of various realizations of Manning's n coefficients. The parameters are then updated through their coordinates in the reduced KL basis, instead of the large nodally defined parameter vector. The representation of the ensemble members in the KL basis is expected to better preserve the geostatistical characteristics of the parameter field in the filter update steps (Hoteit and

Köhl, 2006; Chang and Zhang, 2014). Finally, to enhance the filter's performance and better deal with the nonlinear parameter estimation problem, we introduce iterations to the SEIK update steps as in Luo and Hoteit (2014) and Gharamti et al. (2015). Numerical experiments are conducted to evaluate the performance of the iterative SEIK against the EnKF in a realistic coastal configuration using the ADCIRC model.

The rest of this paper is organized as follows. The problem formulation is described in Section 2. Section 3 summarizes the techniques used in our inference framework, including the sampling of multi-Gaussian realizations of the parameter field, the KL expansion, and the SEIK filter. Section 4 describes the details of the experimental setup. The experimental results, its significance, and implications are presented and discussed in Section 5. A summary of the work and conclusions are given in Section 6.

2. Problem formulation

2.1. ADvanced CIRCulation (ADCIRC) model

We use the ADvanced CIRCulation (ADCIRC) model, which solves the SWEs derived from the depth integration of the incompressible Navier-Stokes equations:

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(Q_x) + \frac{\partial}{\partial y}(Q_y) = 0, \tag{1}$$

$$\begin{aligned}
&\frac{\partial Q_x}{\partial t} + \frac{\partial UQ_x}{\partial x} + \frac{\partial VQ_x}{\partial y} - fQ_y = -gH\frac{\partial[\zeta + P_s/g\rho_0 - \alpha\eta]}{\partial x} + \frac{\tau_{sx}}{\rho_0} - \frac{\tau_{bx}}{\rho_0} + M_x - D_x - B_x,\\
&\frac{\partial Q_y}{\partial t} + \frac{\partial UQ_y}{\partial x} + \frac{\partial VQ_y}{\partial y} - fQ_x = -gH\frac{\partial[\zeta + P_s/g\rho_0 - \alpha\eta]}{\partial y} + \frac{\tau_{sy}}{\rho_0} - \frac{\tau_{by}}{\rho_0} + M_y - D_y - B_y.
\end{aligned} \tag{2}$$

Here, ζ is the free-surface elevation relative to the geoid, h is the bathymetric depth relative to geoid, $H = \zeta + h$ is the water depth, U and V are the depth-averaged horizontal velocity components, $Q_x = UH$ and $Q_y = VH$ are the flux per unit width in the x and y directions, f is the Coriolis parameter, g is acceleration due to gravity, P_s is the atmospheric pressure at the free surface, ρ_0 is the reference density of water, α is the Earth elasticity factor, η in the Newtonian equilibrium tide potential, τ_{sx} and τ_{sy} are the applied free surface stresses, τ_{bx} and τ_{by} are the bottom friction components, M_x and M_y are the vertically-integrated lateral stress gradients, D_x and D_y are the momentum dispersion, and B_x and B_y are the vertically-integrated baroclinic pressure gradients. In ADCIRC, the continuity equation is replaced by the second-order, hyperbolic generalized wave continuity equation (GWCE) to reduce spurious oscillations that occur in the original form. Manning's n coefficients arise in the bottom friction terms of (2). The explicit expression of the bottom friction components are $\frac{\tau_{bx}}{\rho_0} = \frac{K_{slip}Q_x}{H}$ and $\frac{\tau_{by}}{\rho_0} = \frac{K_{slip}Q_y}{H}$. The coefficient $K_{slip} = c_f|\mathbf{u}|$, where $c_f = \frac{gn^2}{H^{1/3}}$, represents a quadratic drag law. The scalar value, n, is the Manning's n coefficient. Since the Manning's n coefficients spatially vary, they are defined node-wise within the discretized physical domain, and are a piece-wise linear representation of the continuous bottom friction field.

The SWEs in ADCIRC are discretized spatially using a first-order continuous Galerkin finite element method with unstructured triangular elements. The time derivatives in the GWCE are approximated with centered finite differences, and forward differences are used for the time derivatives in the momentum equations. ADCIRC has been successfully implemented in many coastal ocean studies (e.g. Westerink et al., 2008; Bunya et al., 2010; Dietrich et al., 2010; Dietrich et al., 2011; Kennedy et al., 2011; Butler et al., 2012).

To simulate tides in an estuarine system, we adopted the same domain as that of Mayo et al. (2014) and Siripatana et al. (2017). This selected domain is an idealized coastal inlet with an ebb shoal, with an open ocean boundary on the left and a reflective boundary (representing the wall along the coastline) on the right as shown in Fig. 1. The domain is discretized into 1518 grid nodes and 2828 elements. Its dimension is 4500 m in the x-direction and 3000 m in the y-direction. Bathymetry is measured downward from the geoid to the ocean floor. The bathymetric depth increases linearly from 3.8 m at the open ocean boundary to 1 m at the mouth of the inlet on the west side of the domain. The shallowest area of the domain is on the mound in front of the west entrance of the inlet with a depth of 0.5 m below the geoid. The landlocked area has a constant bathymetry of 1 m. The diameter of the ebb shoal is 750 m. This configuration is considered to be a simplified version of a real-world ebb shoal system, which is a natural feature of many coastal ocean regions. We force ADCIRC by the $\mathbf{M}_2$ tidal constituent with an amplitude of 0.25 m (relative to the geoid) and a 2 s time step.

3. Parameter estimation framework

This section describes the techniques that are used in our parameters inference framework. These include: (3.1) a sampling scheme and a search space representation (sequential simulation algorithm), (3.2) reduction of the search space (Karhunen-Loève (KL) expansion), (3.3) an ensemble filtering inference scheme (Joint-SEIK for parameter inference), and (3.4) an iterative technique in the filter update step (iterative SEIK).

3.1. Sequential simulation algorithm

The generation of spatially-dependent fields of various variables is useful for the numerical simulation of many problems in geophysical fluid dynamics (Gómez-Hernández and Journel, 1993). Since the collected data is often limited, one must resort to algorithms capable of generating realizations of a full variable field, subject to available data and a suitable covariance model. One of the well-established techniques to generate spatially variant maps is the so-called 'sequential simulation algorithm' (Johnson, 2011). This method recursively draws realizations of variables from a multivariate pdf modeled from series of univariate conditional pdfs that are constrained by available data. For variables following joint Gaussian distributions, the prescribed covariance model, mean, and variance of the field are needed in order to solve for a set of coefficients in a simple kriging system (Anderson, 1984). These are then used to calculate the mean and variance that characterize the conditional density function of each variable, given the set of conditioning data. The covariance model is given by $Cov(h) = c - g(h)$ where h is the variable, $g(h)$ is the corresponding semi-variogram model and c is its sill. In this study, the Manning's field is assumed Gaussian and anisotropic, which can be sampled from Gaussian semi-variogram of the form

$$g(h) = c\cdot(1 - exp(-h^2)). \tag{3}$$

Here $h = \sqrt{(h_x/a_x)^2 + (h_y/a_y)^2}$, where $h_i, i = x,y$, is the lag distance between two locations in the i direction and $a_i, i = x,y$, is an appropriate range in the i direction.

3.2. Karhunen-Loève (KL) expansion

The KL expansion (Loeve, 1947; Karhunen, 1947), is a classical method for expressing stochastic processes as an orthonormal set of deterministic functions. It follows the result of Mercer's theorem (Xiu, 2010), which states that a symmetric positive definite matrix $\boldsymbol{C}(\mathbf{x_1},\mathbf{x_2})$ admits the spectral decomposition

$$\boldsymbol{C}(\mathbf{x_1},\mathbf{x_2}) = \sum_{k=1}^{\infty} \lambda_k \psi_k(\mathbf{x_1})\psi_k(\mathbf{x_2}), \tag{4}$$

where $\lambda_k > 0$ are the eigenvalues of $\boldsymbol{C}$ and ψ_k are the corresponding eigenvectors, i.e. the terms in (4) must satisfy

$$\int_\Omega \boldsymbol{C}(\mathbf{x_1},\mathbf{x_2})\psi_k(\mathbf{x_2})d\mathbf{x_2} = \lambda_k\psi_k(\mathbf{x_1}), \quad k = 1,2,... \tag{5}$$

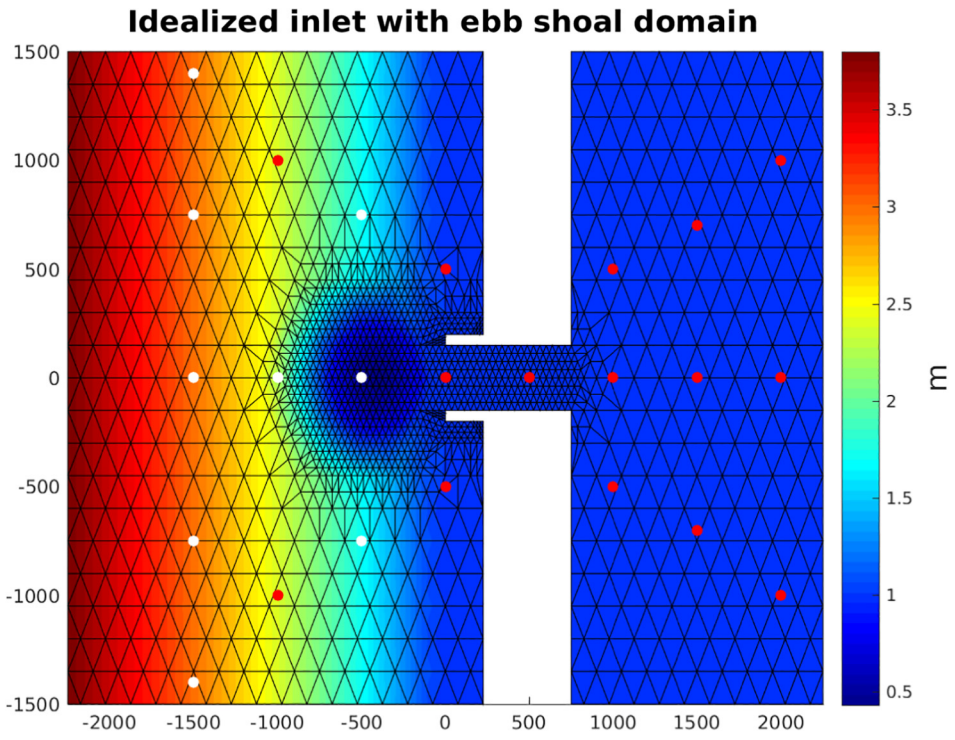

Fig. 1. Idealized inlet with ebb shoal domain. The discretization of the domain is represented. The first 15 observation stations used in the experiment are marked with red dots, and the additional 9 observation stations added later are marked with white. The color bar represents the bathymetry of the domain measuring down from the geoid (m).

The sequential simulation algorithm described in Subsection 3.1. produces realizations of a variable field with mean $\mu(\mathbf{x})$ and a discretization $\boldsymbol{C}(\mathbf{x_1},\mathbf{x_2})$ of the covariance function $Cov(h)$, where $\mathbf{x} \in \mathbb{R}^d$ is a vector of length d of the nodes of a discretized domain. The covariance function is then decomposed according to Mercer's theorem. Let $\boldsymbol{K}(\mathbf{x},\xi)$ be a stochastic function of a coordinate vector $\mathbf{x}$ and a random variable ξ. Every realization of $\boldsymbol{K}$ can then be expressed as

$$\boldsymbol{K}(\mathbf{x},\xi) = \mu(\mathbf{x}) + \sum_{k=1}^{\infty} \sqrt{\lambda_k}\xi_k\psi_k(\mathbf{x}). \tag{6}$$

In the case of a multi-Gaussian field generated by the sequential simulation algorithm, ξ_k is a Gaussian independent identically distributed random variable with zero mean and unit variance. The function $\boldsymbol{K}$ is fully characterized by a set of ξ_k when the basis ψ_k is known. Given a realization of $\boldsymbol{K}$, together with a known covariance matrix, the ξ_k can be obtained by evaluating the integral

$$\xi_k = \int (\boldsymbol{K}(\mathbf{x},\xi) - \mu(\mathbf{x}))\psi_k(\mathbf{x})d\mathbf{x}. \tag{7}$$

KL expansions represent highly spatially variant parameters as realizations of a stochastic process with only a few dominant modes by truncating the infinite series in (6) with a finite number of N terms. The size N essentially depends on the desired energy percentage to be retained by the KL modes $\sum_{k=1}^{N}\sqrt{\lambda_k} / \sum_{k=1}^{\infty}\sqrt{\lambda_k}$. This notion of "optimal" truncation is particularly useful for large scale parameter inference problems in order to alleviate computational burdens while retaining the essential features of the inference space (ElSheikh et al., 2013).

3.3. Joint-SEIK for parameter Inference

The Joint EnKF approach is widely used for parameter estimation by the subsurface modeling community (e.g. Naevdal et al., 2005; Tong et al., 2010; Gharamti et al., 2013; Katterbauer et al., 2015; Ait-El-Fquih et al., 2016). In the most general form, a vector of model parameters to be estimated, $\mathbf{w}$, is appended to the system state vector $\mathbf{x}_k$, to form the joint state-parameter vector

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{w} \end{bmatrix}. \tag{8}$$

Assuming stationary dynamics for the parameters, the augmented state-space model is then written as

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{w}_k \end{bmatrix} = \begin{bmatrix} \mathscr{M}(\mathbf{x}_{k-1}) \\ \mathbf{w}_{k-1} \end{bmatrix} + \begin{bmatrix} \eta_k \\ \mathbf{0} \end{bmatrix}, \tag{9}$$

where $\mathscr{M}$ is the dynamical operator describing the time evolution of the state vector from time $k-1$ to time k, and η_k is the model error with Gaussian of mean zero and covariance matrix Q_k. The observation $\mathbf{y}_k$ is then related to the augmented state vector as

$$\mathbf{y}_k = \mathbf{H}_k^{\mathbf{z}}(\mathbf{z}_k) + \varepsilon_k^{\mathbf{z}} = \mathscr{H}_k(\mathbf{x}_k) + \varepsilon_k, \tag{10}$$

where $\mathbf{H}_k$ is the linearized observation operator and ε_k the measurement noise.

Some studies pointed to some difficulties in estimating the model parameters with the ensemble Kalman filter (Yang and Delsole, 2009; Koyama and Watanabe, 2010; DelSole and Yang, 2010), but many more presented quite successful implementations, e.g. Aksoy et al. (2006), Franssen and Kinzelbach (2008), Gharamti et al. (2013), Mayo (2013); Siripatana et al., 2017 just to cite a few. Among the most reported issues were related to strong nonlinear relations between the observations and the estimated parameters (Kivman, 2003; Chen et al., 2009; Subramanian et al., 2012), the relevance of the assimilated information (Anderson, 2001), and the size of the problem (Moradkhani et al., 2005). These were however not problematic in our particular setting and the filter performances were deemed quite satisfactory in our numerical experiments presented in Section 5.

Here we follow Mayo et al. (2014) and Butler et al. (2012), and implement the Singular Evolutive Interpolated Kalman (SEIK) filter, which was found to be particularly efficient at enhancing the predictive capabilities of ADCIRC (Butler et al., 2012; Altaf et al., 2014) and also for parameters estimation (Mayo et al., 2014; Siripatana et al., 2017). Compared to the other deterministic EnKFs, SEIK involves a stochastic rotation in the resampling step to randomly spread the error variance in the ensemble space (Hoteit et al., 2002), which is suitable for strongly nonlinear dynamics that often arise during storm surges and was later suggested for the Ensemble Transform Kalman Filter (ETKF) (Wang et al., 2004; Nerger et al., 2012). SEIK algorithm can be split in three steps; given an initial ensemble ($\mathbf{z}_0^{a,i}, i = 1,...,N$).

3.3.1. Forecast step

The forecast step integrates the analyzed ensemble members, $\mathbf{z}_{k-1}^{a,i}$, with the model (9) to compute the forecast ensemble members, $\mathbf{z}_k^{f,i}$. One then takes the average of the $\mathbf{z}_k^{f,i}$ as the forecast state vector, $\mathbf{z}_k^f$, and their sample covariance as the forecast error covariance, $\mathbf{P}_k^f$. Assuming a perfect model ($Q_k = \mathbf{0}$), one can decompose

$$\mathbf{P}_k^f = \mathbf{L}_k \mathbf{U}_{k-1} \mathbf{L}_k^T, \tag{11}$$

with

$$\mathbf{L}_k = [\mathbf{z}_k^{f,1} - \mathbf{z}_k^f \quad \dots \quad \mathbf{z}_k^{f,r+1} - \mathbf{z}_k^f]\mathbf{T}, \tag{12}$$

and

$$\mathbf{U}_{k-1} = [(r-1)\mathbf{T}^T\mathbf{T}]^{-1}. \tag{13}$$

Here $\mathbf{T}$ is an $(r+1) \times r$ full rank orthogonal matrix with zero column sums. When the model error is not neglected, SEIK accommodates the model error by adding its covariance matrix to the right hand side of (11). Its algorithm remains mostly unchanged. However in this case, $\mathbf{P}_k^f$ will not remain of low-rank r, and re-approximating the forecast covariance matrix $\mathbf{P}_k^f$ will be required (Hoteit et al., 2007).

3.3.2. Analysis step

When a new observation $\mathbf{y}_k$ becomes available, The forecast state is updated to obtain the analysis state

$$\mathbf{z}_k^a = \mathbf{z}_k^f + \mathbf{K}_k(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{z}_k^f), \tag{14}$$

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

$$\mathbf{K}_k = \mathbf{L}_k \mathbf{U}_k (\mathbf{HL})_k^T \mathbf{R}_k^{-1}. \tag{15}$$

$(\mathbf{HL})_k$ is computed by applying $\mathbf{H}_k$ to the ensemble perturbations $\mathbf{z}_k^{f,i} - \mathbf{z}_k^f$,

$$(\mathbf{HL})_k = [\mathbf{H}_k(\mathbf{z}_k^{f,1} - \mathbf{z}_k^f) \quad \dots \quad \mathbf{H}_k(\mathbf{z}_k^{f,r+1} - \mathbf{z}_k^f)]\mathbf{T}, \tag{16}$$

and

$$\mathbf{U}_k^{-1} = \frac{1}{\rho}\mathbf{U}_{k-1}^{-1} + (\mathbf{HL})_k^T \mathbf{R}_k^{-1} (\mathbf{HL})_k. \tag{17}$$

The inflation factor, ρ, is used to inflate the forecast error covariance as a way to account for various sources of uncertainties in the system, e.g. model error, small ensembles, Gaussian assumption, etc (Anderson, 2001; Hoteit et al., 2002). The analysis error covariance can be expressed as $\mathbf{P}_k^a = \mathbf{L}_k \mathbf{U}_k \mathbf{L}_k^T$, but this is not needed for the filter algorithm.

3.3.3. Resampling step

New ensemble members need to be generated to start the next forecast cycle. These are sampled from the analysis mean and the covariance as

$$\mathbf{z}_{k-1}^{a,i} = \mathbf{z}_{k-1}^a + \sqrt{N}\mathbf{L}_{k-1}(\mathbf{\Omega}_{k-1}\mathbf{C}_{k-1}^{-1})^T, \quad i = 1,\dots,N \tag{18}$$

where $\mathbf{\Omega}_{k-1}$ is an $(r+1) \times r$ matrix with orthonormal columns and zero column sums generated using Householder matrices (Pham, 2001; Hoteit et al., 2002). In this study, we are only interested in estimating the parameters, i.e. $\mathbf{w}_k$.

3.4. Iterative SEIK

Parameter estimation with an EnKF can suffer from strong nonlinearities between the observed state and the parameters (Jardak et al., 2010). Iterating on the parameter update step has been shown to improve the accuracy of the filter estimates (Lorentzen and Naevdal, 2011; Gharamti et al., 2015).

Let $\mathbf{x}_k^{a,j}$ be the analyzed state at timestep k and iteration j. The iterative SEIK (ISEIK) seeks the solution of the nonlinear least squares problem:

$$\underset{x_k^{a,j+1}}{argmin} \|\mathbf{y}_k^o - \mathscr{H}(\mathbf{x}_k^{a,j}) - \mathbf{J}_k^j(\mathbf{x}_k^{a,j+1} - \mathbf{x}_k^{a,j})\|_{\mathbf{R}_k}^2 + \|\mathbf{x}_k^{a,j+1} - \mathbf{x}_k^{a,j}\|_{\mathbf{C}_k}^2, \tag{19}$$

where $\mathbf{J}$ is the Jacobian matrix of $\mathscr{H}$. The term $\mathscr{H}(\mathbf{x}) - \mathbf{J}_k^i(\mathbf{x}_k^{a,j+1} - \mathbf{x}_k^{a,j})$ is the first-order Taylor approximation of $\mathscr{H}(\mathbf{x}_k^{a,j+1})$ and $\mathbf{C}_k$ is a symmetric, positive semidefinite matrix. The solution $\mathbf{x}_k^{a,j+1}$ of (19) is derived in Engl et al. (2000) as

$$\mathbf{x}_k^{a,j+1} = \mathbf{x}_k^{a,j} + \mathbf{K}_k(\mathbf{y}_k^o - \mathscr{H}(\mathbf{x}_k^{a,j})). \tag{20}$$

We see that this equation is the iterative form of (14).

As the iterations advance, the inbreeding problem may cause the filter to increasingly underestimate the ensemble variance, ultimately degrading the filter's performance (Song et al., 2013). This problem is more pronounced when the parameter and the state are strongly nonlinearly related. To this end, we adopt a strategy that limits the size of the update term in the later iterations via a damping factor ω_j, as suggested by Hendricks Franssen and Kinzelbach (2008). The factor ω_j takes values between 0 to 1 and multiplies the update term (increment to the forecast) in (20). This helps to smooth the perturbation of Manning's n coefficients, which alleviates the impact of the state-parameter nonlinear relation and sampling errors.

The iterative scheme (20) can be directly applied to SEIK with minor modifications. In particular, if $\mathscr{H}$ is linear (as in this study) and $\mathbf{C}_k$ is taken as the covariance matrix $\mathbf{P}_k$, then one only needs to iterate on $(\mathbf{HL})_k$ to derive the iterative SEIK scheme. Moreover, since here we only update the parameters and not the state, we only need to iterate on the parameter ensemble mean in (20), while maintaining the ensemble variance during each assimilation cycle k. In this study, the iterations are stopped when the updates become small or a maximum iteration number is reached. For more sophisticated stopping criteria, readers may refer to Luo and Hoteit (2014).

4. Experimental setup

4.1. Generating synthetic multi-Gaussian Manning's n fields

Synthetic data of the Manning's n coefficients are first generated by taking a small number of samples from the uniform distribution $U(0.005,0.2)$ to simulate a scenario where a few point-wise Manning's n coefficients data are collected (or inferred from point-wise bottom surface characteristics). These data are assumed collected at 24 locations representing the observations stations as illustrated in Fig. 1. The synthetic Manning's n coefficient data are then integrated with the public domain ANSI-C code 'GCOSIM3D' developed in Gómez-Hernández and Journel (1993), to generate multi-Gaussian 2D Manning's fields for our idealized ebb shoal domain, based on the sequential simulation algorithm (Section 3.1). From this, any number of Manning's field realizations can be generated once the properties of the semi-variogram are set. We first generate 1000 realizations of nodally-defined parameter fields following the Gaussian semi-variogram as in (3), with a mean of 0.1025, a variance of 0.0002, and a correlation range of 180 m in the x-direction and 30 m in the y-direction. The variance is properly scaled so that the realizations of 2D multi-Gaussian fields fall within an appropriate range of Manning's n coefficients (0.005–0.2). The maximum and minimum Manning's n values of these realizations are 0.1879 and 0.0177, respectively. Examples of realizations of Manning's n fields generated by the sequential simulation algorithm are shown in Fig. 2. These realizations are used to select the initial ensemble, compute the KL modes, and define the reference field.

4.2. Observation system simulation experiments (OSSEs)

We let the 101^{st} realization generated by GCOSIM3D code in Section 4.1 be the reference Manning's n field, which we seek to infer (Fig. 2). Synthetic observations of water elevation are generated by ADCIRC integrated using the reference Manning's n field. The dimension of the

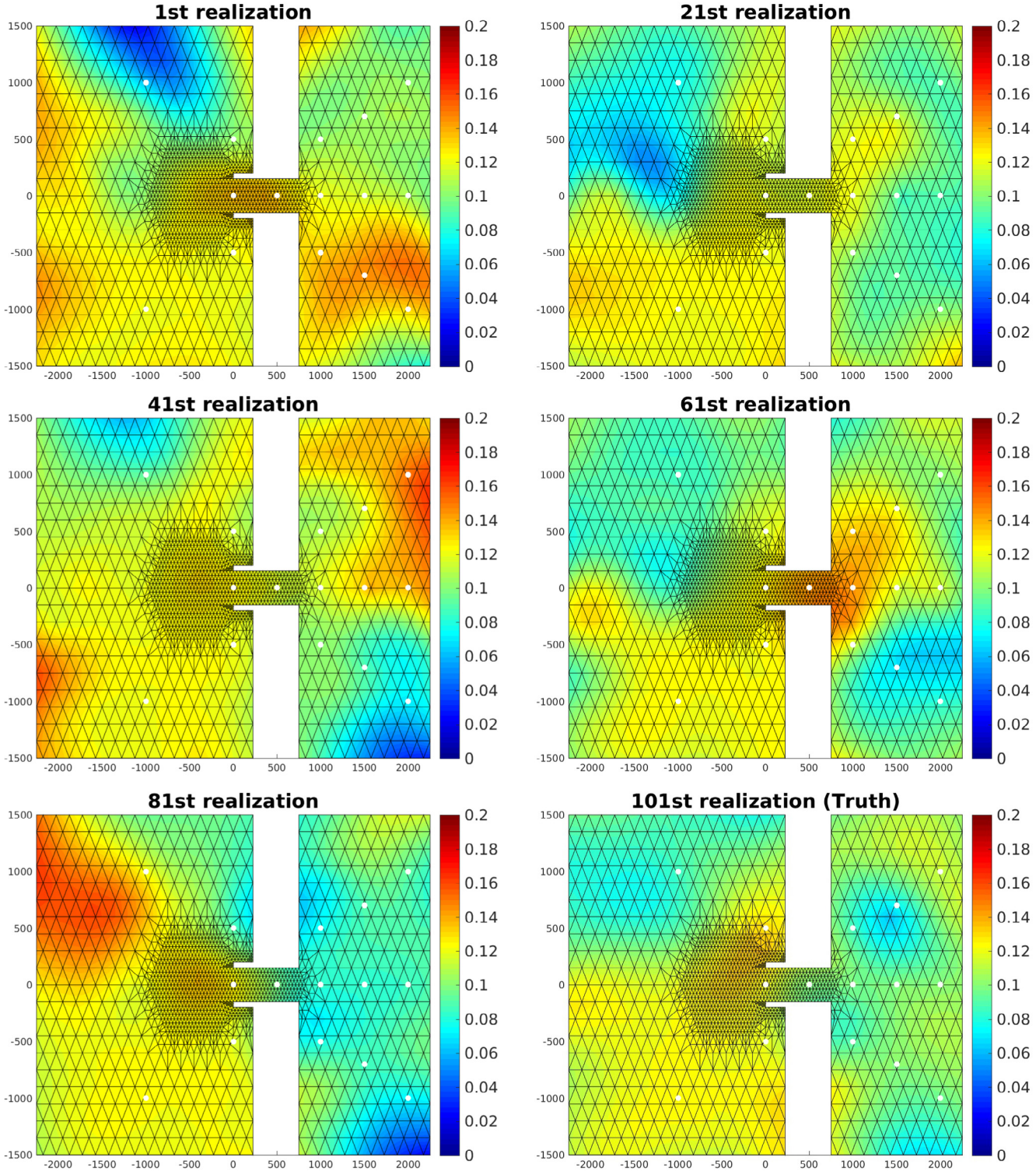

Fig. 2. A few realizations of Manning's n fields generated by the sequential simulation algorithm. The 101^{st} realization is taken as the reference field that the inference results are compared against.

observations is the number of observed locations multiplied by the number of assimilation time steps. Initially, we use 15 observation stations as shown in Fig. 1, and 108 assimilation timesteps (4.5 days with incoming data every 1 h). Later, we also increase the number of observation stations to 24 and total assimilation time to 20 days in some experiments to study the impact of the number of observations on parameter inference. Two-hundred Manning's n field realizations, excluding the reference realization, are taken as the initial members. We first let the simulation ramp-up for 12 days using the mean of the initial members before the first assimilation cycle starts. The observations are then assimilated by SEIK to infer the reference Manning's field. We test the filter with four different settings: 1) nodally-defined Manning's n values, 2) Manning's n field parameterized by the KL-reduced space, 3) Manning's n field inference in the KL space with perturbed variogram models, and 4) iterative SEIK filter inferring the Manning's field in both the full and reduced space.

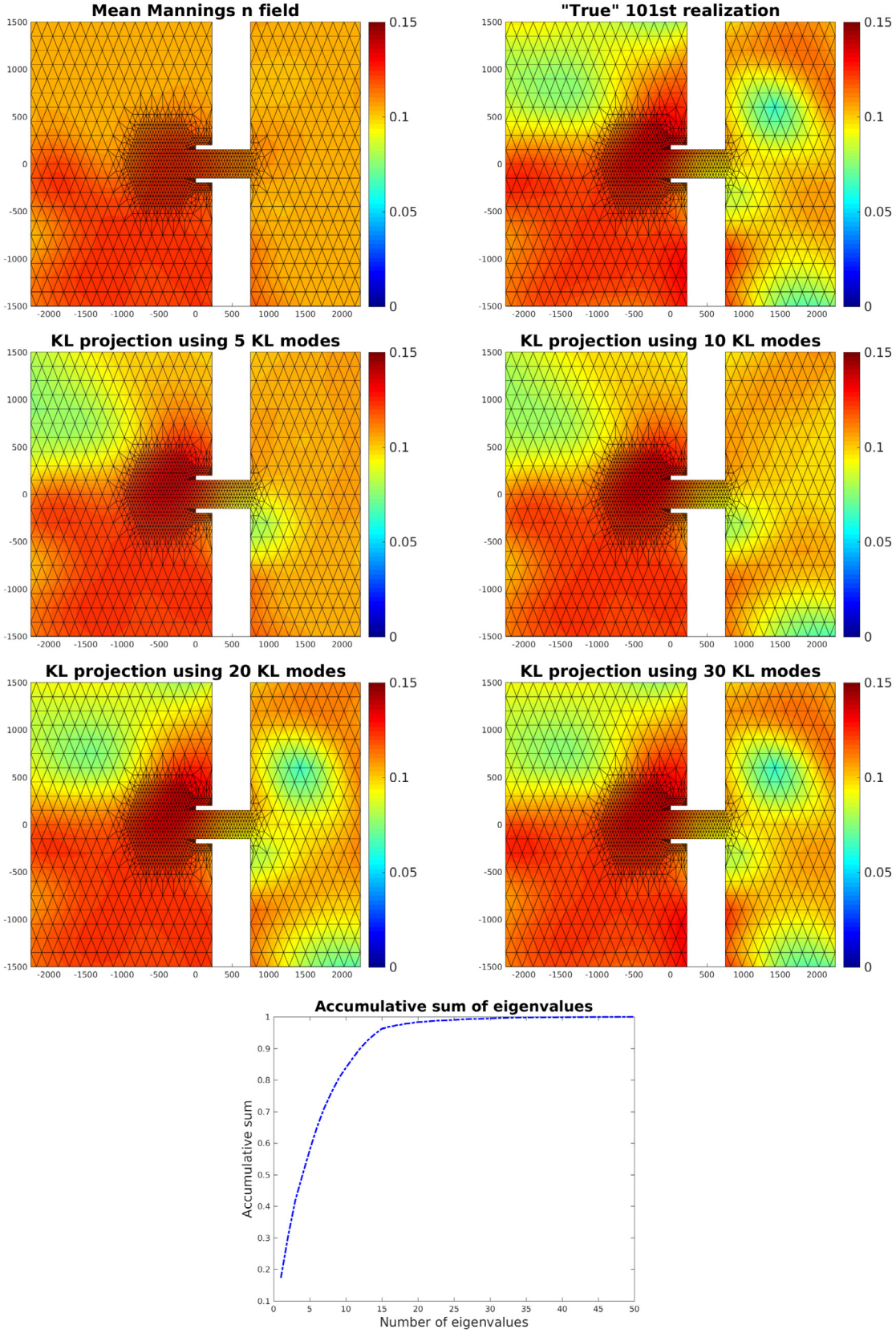

Fig. 3. The reconstruction of a Manning's n field using the truncated KL expansion for various retained KL modes. The top six subplots: spatial plots of the reconstructed Manning's n field. The bottommost subplot: the accumulative sum of the eigenvalues obtained for the KL decomposition.

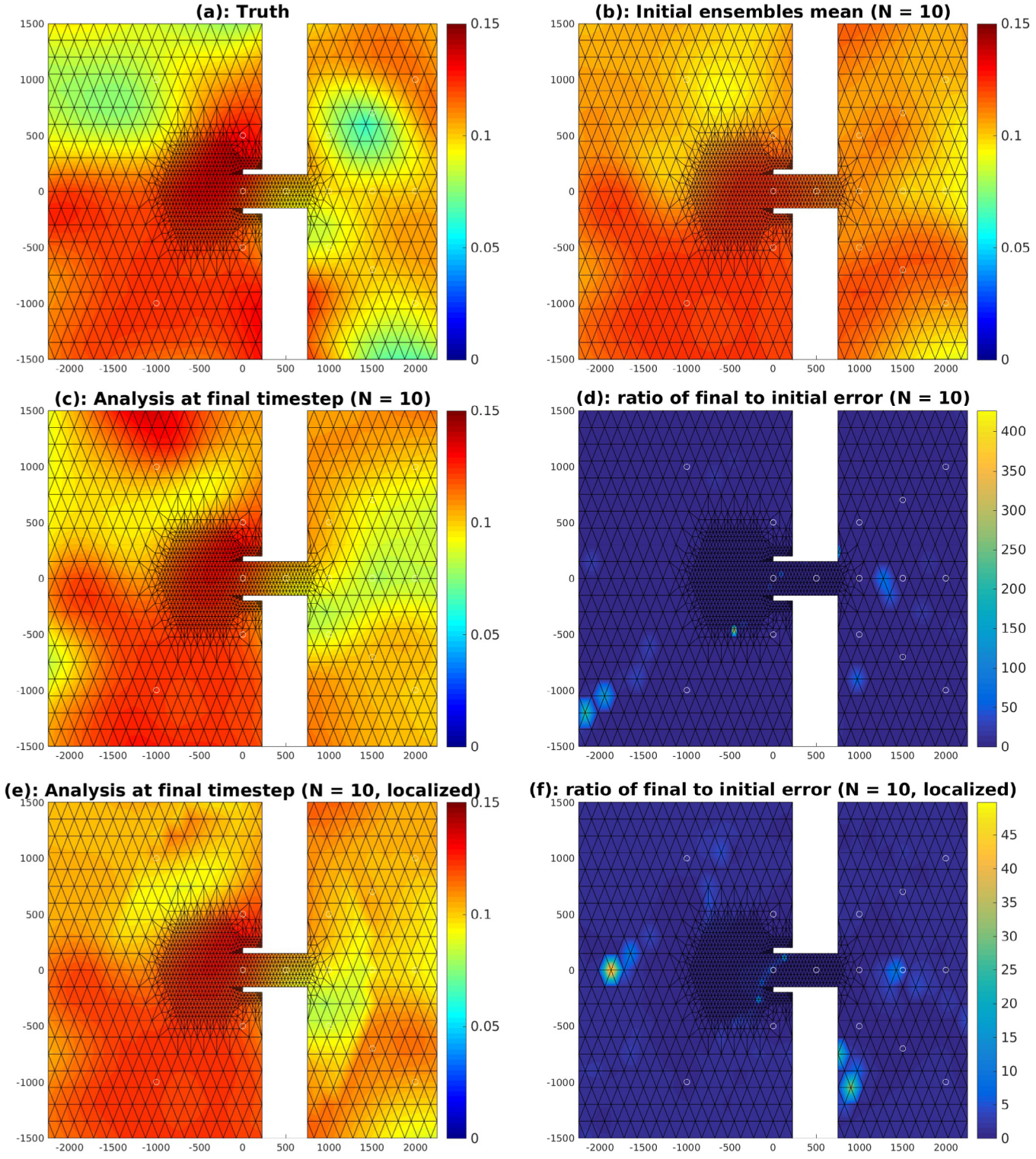

Fig. 4. The results of SEIK inference with 10 ensemble members, (a) the true field, (b) the initial ensemble mean, (c) the final analysis after 108 updates, (d) the ratio between the final error and the initial error, (e) the final analysis when applying localization, and (f) the ratio between the final and the initial error when applying localization.

4.3. KL basis construction

The sample covariance of the 1000 realizations generated in Subsection 4.1 is decomposed as in (4) to obtain the set of eigenpairs and the KL expansion of the parameter (i.e., the 2D Manning's n coefficients) as in (6). The cumulative sum of eigenvalues, which indicates the total variance retained by the KL expansion is plotted at the bottommost of Fig. 3. It shows that retaining 10 and 20 KL terms respectively preserve more than 83% and 98% of the total variance of the realizations. By increasing the truncation to 30 KL terms, more than 99% of the total variance is retained.

Fig. 3 also shows an example of the reconstruction of a Manning's field using a truncated KL expansion. The top row of the figure shows the mean of the 1000-realizations of the Manning's n field and the 101st

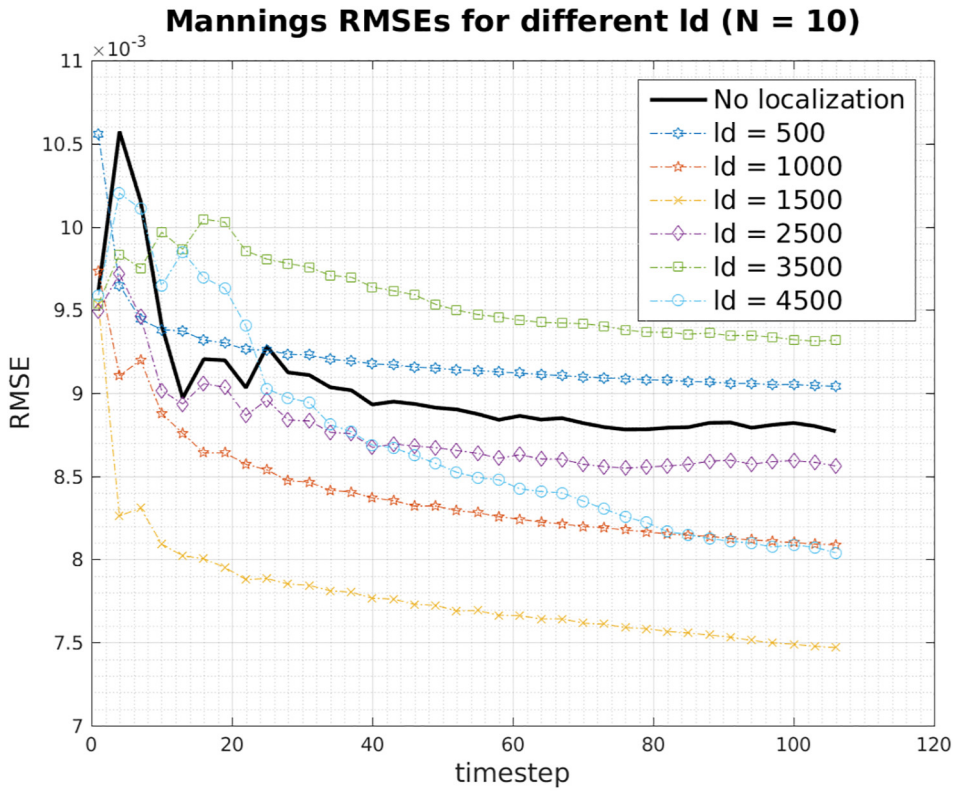

Fig. 5. Time series of the RMSEs of the Manning's n fields after each analysis step with respect to the true Manning's n field for varying localization distance (ld).

realization, respectively. The remaining subfigures are the reconstructions of the 101st realization as we increase the number of KL terms in (6). One can observe that with a small number of KL terms, for example, 3 KL modes, the reconstructed field resembles the mean, as the mean field dominates the modes. As we increase the number of KL terms, the reconstruction starts converging toward the target realization (the 101st realization in this figure).

5. Results and discussion

5.1. SEIK inference of nodally defined Manning's n values

Fig. 4(a)–(d) present the results of the Manning's n field SEIK inference using only 10 ensemble members compared to the reference field. The impact of the SEIK updates is clear from the final analysis, as the filter solution more accurately represents the reference Manning's n field. The 2D plot of the ratio between the final error and the initial error suggests reasonably small errors in most locations, except those where the Manning's n values of the initial ensemble vastly differ from the true values. With only 10 ensemble members, this set of results is considered as a preliminary test, upon which we make efforts to improve.

We then applied Local Analysis (LA) in an attempt to improve the SEIK filter performance (Nerger et al., 2006), later referred to as 'local SEIK'. This technique provides a straightforward way to cut the spurious long-range correlations in the covariance matrix of the filter's analysis step. In Fig. 5, we show the time-series of Root Mean Square Errors (RMSEs) of the analysis with respect to the reference field for various localization distances (in meters). Localization enhances the filter's performance in most cases, although the filter's behavior with 10 members is quite sensitive to the choice of the localization distances (ld); the smallest RMSE at the end of the assimilation window is attained using ld = 1500 m.

In Fig. 4(e) and (f), we show the impacts of applying localization to the SEIK filter. There is a clear improvement compared to those obtained without localization. The recovery of the Manning's n coefficients around the right side of the inlet and the top left corner of the domain is notably improved. The ratio between the final error and the initial error are close to zero in most areas, except the areas where the difference between the Manning's n values of the initial ensemble and the reference values were sizable.

5.1.1. Sensitivity to ensemble size

Increasing the ensemble size is generally expected to enhance the performance of an EnKF (Hoteit et al., 2002; Triantafyllou et al., 2003; Hoteit et al., 2013). In Siripatana et al. (2017), increasing the ensemble size from 10 to 100 drastically improved the estimation of a 1D Manning's n coefficient in ADCIRC. However, this raises the issue of determining a good trade-off between filter performance and computational costs. Doubling the size of the ensemble means twice as many model runs are required. In the case of a complex model such as ADCIRC, this can result in a tremendous increase in computational time. Our first aim is to determine the ensemble size that yields satisfying filter performance with reasonable computational cost.

We thus assess the filter's performance with increasing numbers of ensemble members: 10, 50, 100, and 200, respectively, using the same localization distance (ld = 1500 m). The best localization distance in term of reducing the RMSE can be dependent on the ensemble size, however, we found that the localization length scale of 1500 m provides the lowest RMSE for most of our experiments (the top panel of Fig. 6). More in-depth discussions on the choice of the localization length scale can be found in Nerger et al. (2006). The time-series RMSE results of these runs, including the runs from Subsection 5.1., are shown in the bottom panel of Fig. 6. Here we see that increasing the number of ensemble members to 100 greatly reduces the discrepancy between the estimates and the reference field. However, increasing the ensemble size beyond 100 members does not significantly boost the performance of the filter, although it drastically increases the computational cost.

Fig. 7(c) and (d) summarize the results obtained by implementing local SEIK with 100 ensemble members. Improvements over the case

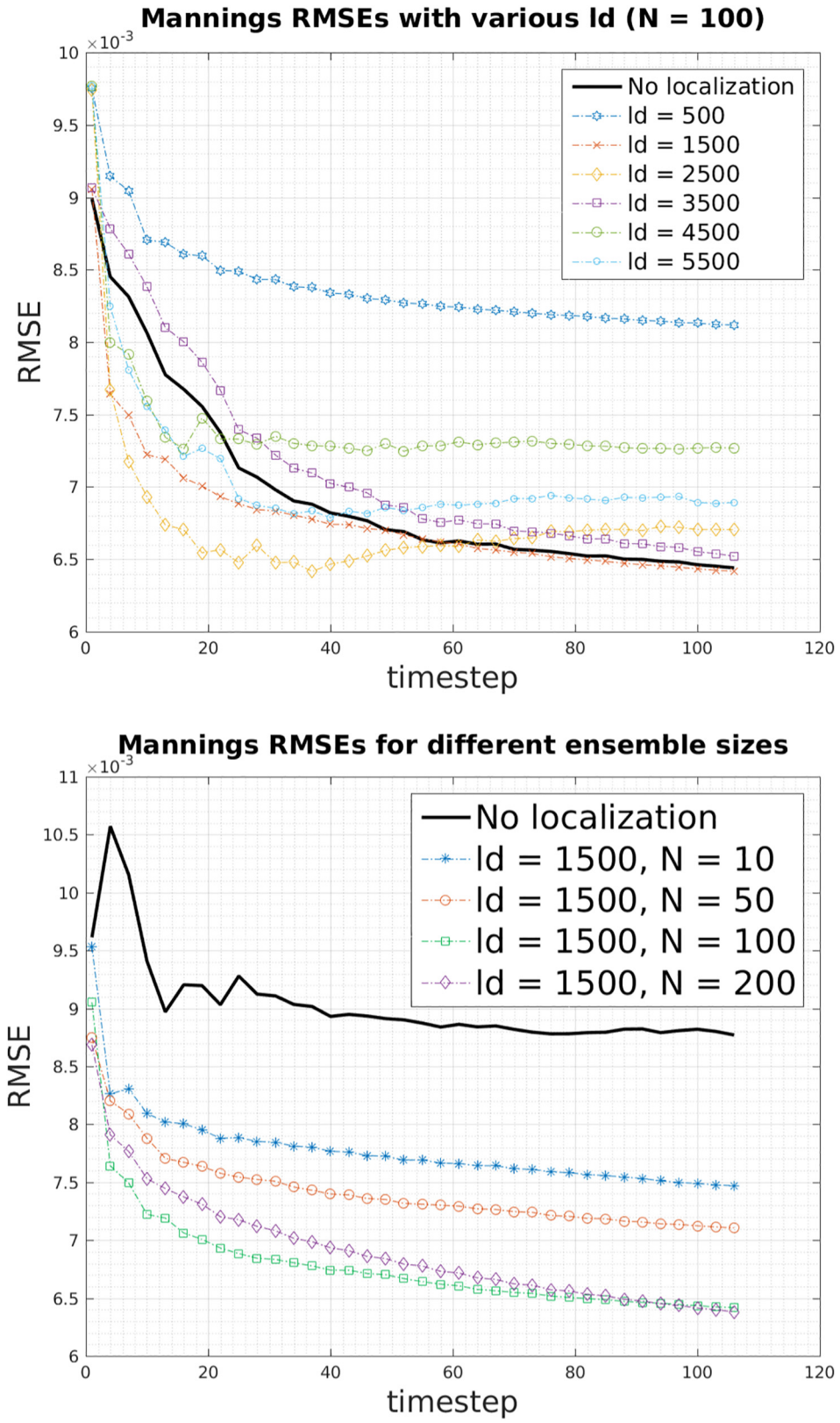

Fig. 6. Time series of the RMSEs of the Manning's *n* field after every analysis step. Top panel: the RMSEs for varying localization distances (ld) with 100 members. Bottom panel: the RMSEs for various ensemble sizes (N) with the same localization.

with 10 ensemble members are clear. The analysis at the end of the assimilation window accurately recovers the 2D Manning's *n* coefficients at the right side of the inlet. The pattern of small Manning's *n* values around the upper-right corner of the domain is also well recovered compared to the case with only 10 members. With this improvement, the ratios of the final to the initial error are close to zero and less than one in most areas, indicating that the local SEIK solution converges toward the reference field at almost every point in the

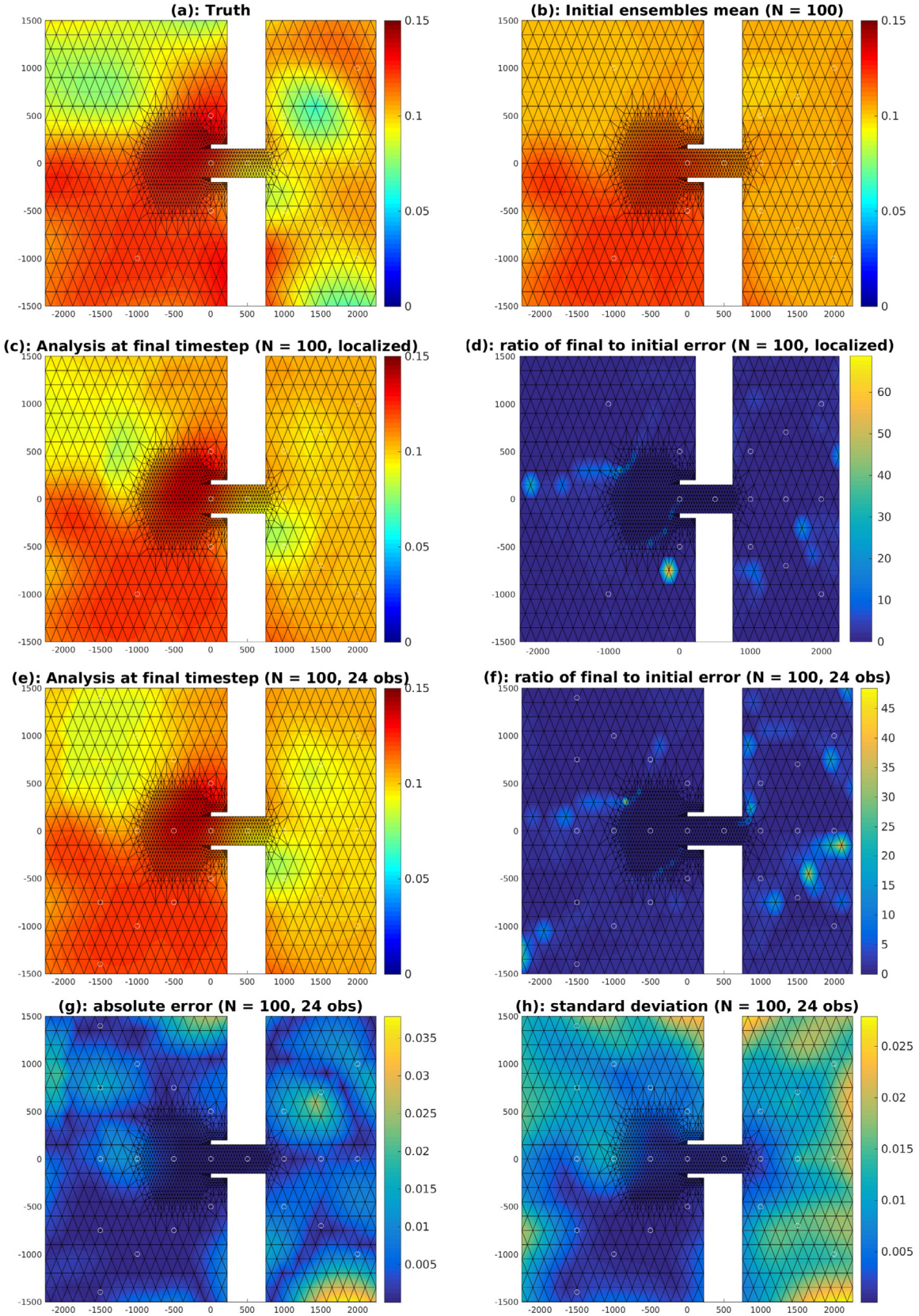

Fig. 7. The results of SEIK inference with 100 ensemble members, (a) the true field, (b) the initial ensemble mean, (c) the final analysis with 15 observation points, (d) the ratio between the final error and the initial error, (e) the final analysis with 24 observation points and 468 assimilation cycles, and (f) the ratio between the final error and the initial error, (g) the absolute error between the estimate and the truth for 24 observation points case, and (h) standard deviation of the ensembles at the final analysis step.

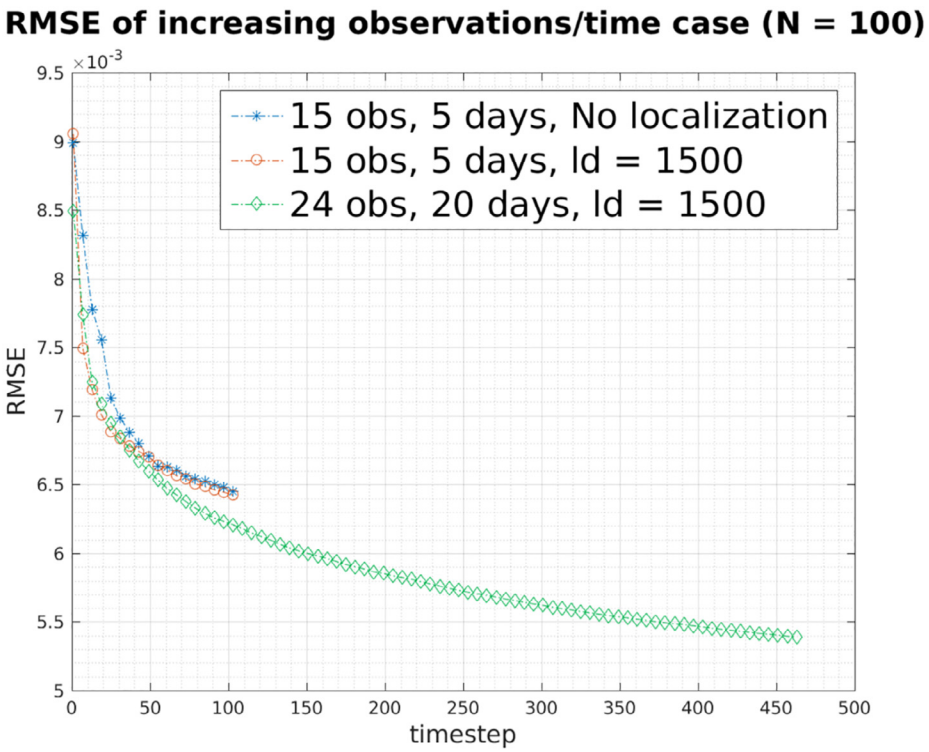

Fig. 8. Time series of the RMSEs of the Manning's n fields after each analysis step for 100 ensemble members, different localizations (ld) and different number of observations with respect to the true Manning's n field.

domain.

5.1.2. Sensitivity to the number of observations and assimilation cycle

In general, it is preferable to assimilate as many observations as possible to compute reliable estimates. Here we explore the behavior of the system with an increasing number of observations, both spatially and temporally. We first introduce 9 additional observation stations (indicated with white dots in Fig. 1) to the domain, increasing the total number of observation stations from 15 to 24. The observations locations are sampled to evenly span the spatial domain. We also increase the simulation time to 20 days, which equates to 468 total assimilation cycles.

Fig. 7(e) and (f) outline the results of this experiment, where 100 ensemble members are used. The SEIK filter successfully recovers most of the features of the Manning's n coefficients shown in the reference field. The node-wise ratios between the final error and the initial error are small and close to zero in most areas. The pattern of low Manning's n coefficients in the right land-locked area to the left area near the open ocean area is almost fully recovered. The only area where there is difficulty recovering the Manning's n features is the bottom-right corner of the domain. This can be attributed to the absence of observations in this area. In addition, we analyze the misfit between the filter estimate and the truth in Fig. 7(g) in relation to the predicted variance of the error as estimated by the ensemble standard deviation (STD) in Fig. 7(h), as resulting from the filter. Overall, both statistics are of the same order despite relatively larger STDs along eastern and northern boundaries. The plots further reveal similar spatial structures, e.g., large error and STD values at the bottom-right corner of the domain (highlighted in red) where the observations are scarce, contrasting with small errors and STD around the center of the open ocean area (highlighted in dark blue), where the observations are more abundant. Similar consistencies between the final (misfit between the truth and final estimate) and predicted (filter error variance) estimation errors were obtained in the rest of our experiments, indicating that with large enough ensembles and good tuning of the localization radius, the estimation of Manning's n coefficients with the EnKF does not suffer from any divergence problem in our particular setting.

In Fig. 8 we see the time-evolution of the RMSE of the estimates with respect to the reference field, based on three different implementations of SEIK with 100 ensemble members: regular SEIK, local SEIK, and local SEIK with additional observations and assimilation cycles, respectively. The discrepancy between the estimate and the truth visibly decreases as more observations are assimilated into the system, with a decreasing RMSE trend that suggests further improvements might be obtained with more assimilation cycles. Another conclusion one can draw from the time-evolution of the RMSE is that the filter does not really benefit in terms of estimation accuracy from localization when implemented with 100 ensemble members. Hereafter, we will consider the regular SEIK solution with 100 members as a reference to evaluate the performance of various tested filtering schemes.

5.2. SEIK inference in KL space

Instead of using SEIK to update the nodally-defined parameter, here we update the KL coefficients, ξ, that represent a specific realization of the Manning's n field in the KL space, using the same filtering procedure for parameter estimation described in Subsection 3.3. The number of KL coefficients to be updated by the filter is the number of terms retained in the KL expansion. Here, we study the sensitivity of the performance of SEIK for parameter estimation in the KL space, later referred to as SEIK-KL, to both the number of retained terms and the ensemble size.

In Fig. 9, we plot the time-evolution of the RMSE of the analyzed Manning's n field with varying ensemble size and the number of KL terms. Each individual curve represents the RMSE of a single SEIK run with a fixed ensemble size and a specific number of preserved KL terms. The left column of subfigures (Fig. 9(a), (b) and (c)) show the RMSE of SEIK-KL using different numbers of KL terms for a specified ensemble size. Conversely, the right column of subfigures (Fig. 9(d), (e) and (f)) show the RMSE for a specified number of KL terms and varying ensemble sizes. First, we examine the results of SEIK inference with 10

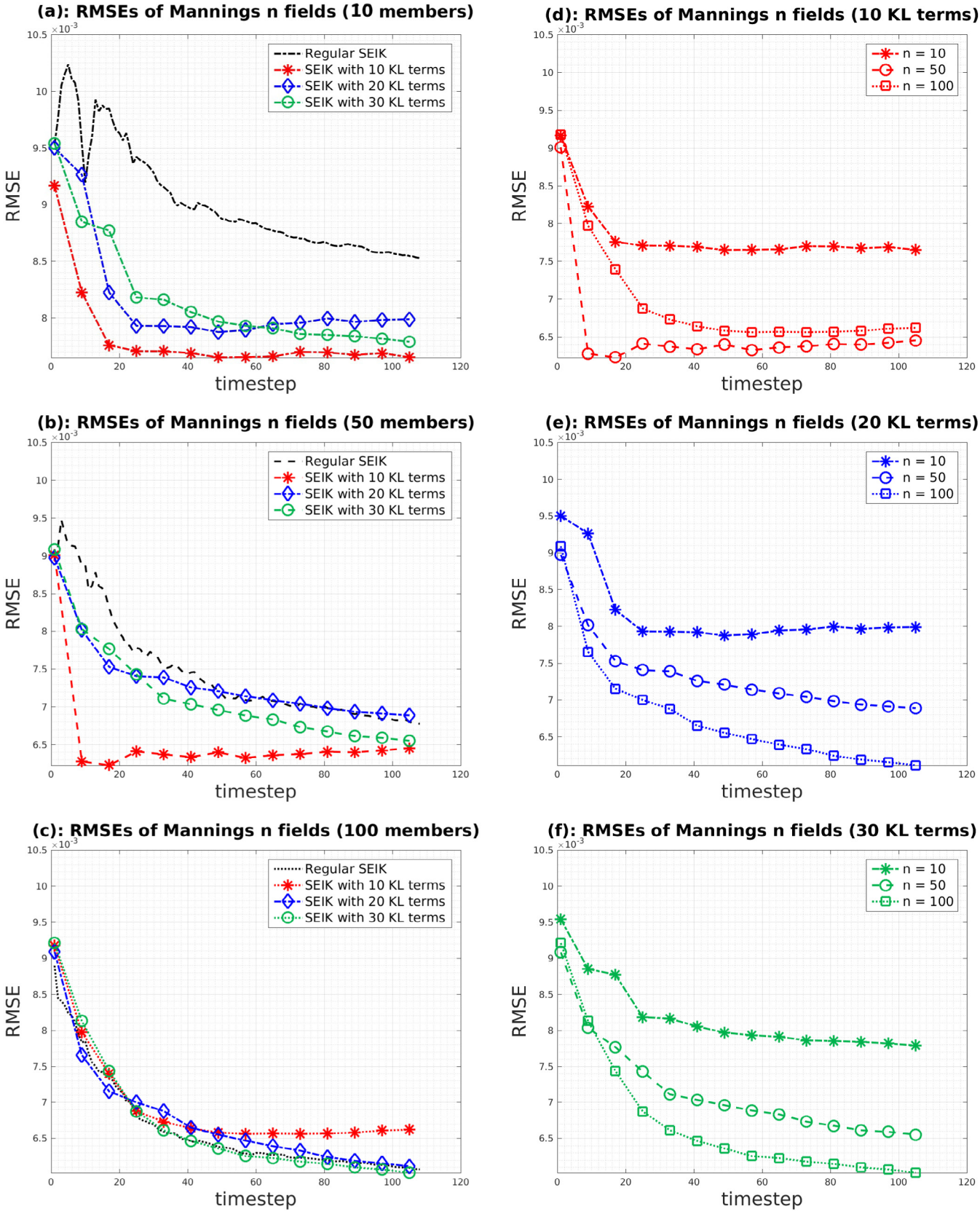

Fig. 9. Time series of the RMSEs as results from SEIK and SEIK-KL using varying ensemble sizes and numbers of retained KL terms. Left column: each figure represents a fixed ensemble size but varying number of retained KL terms. Right column: each figure represents a fixed number of retained KL terms but varying ensemble size

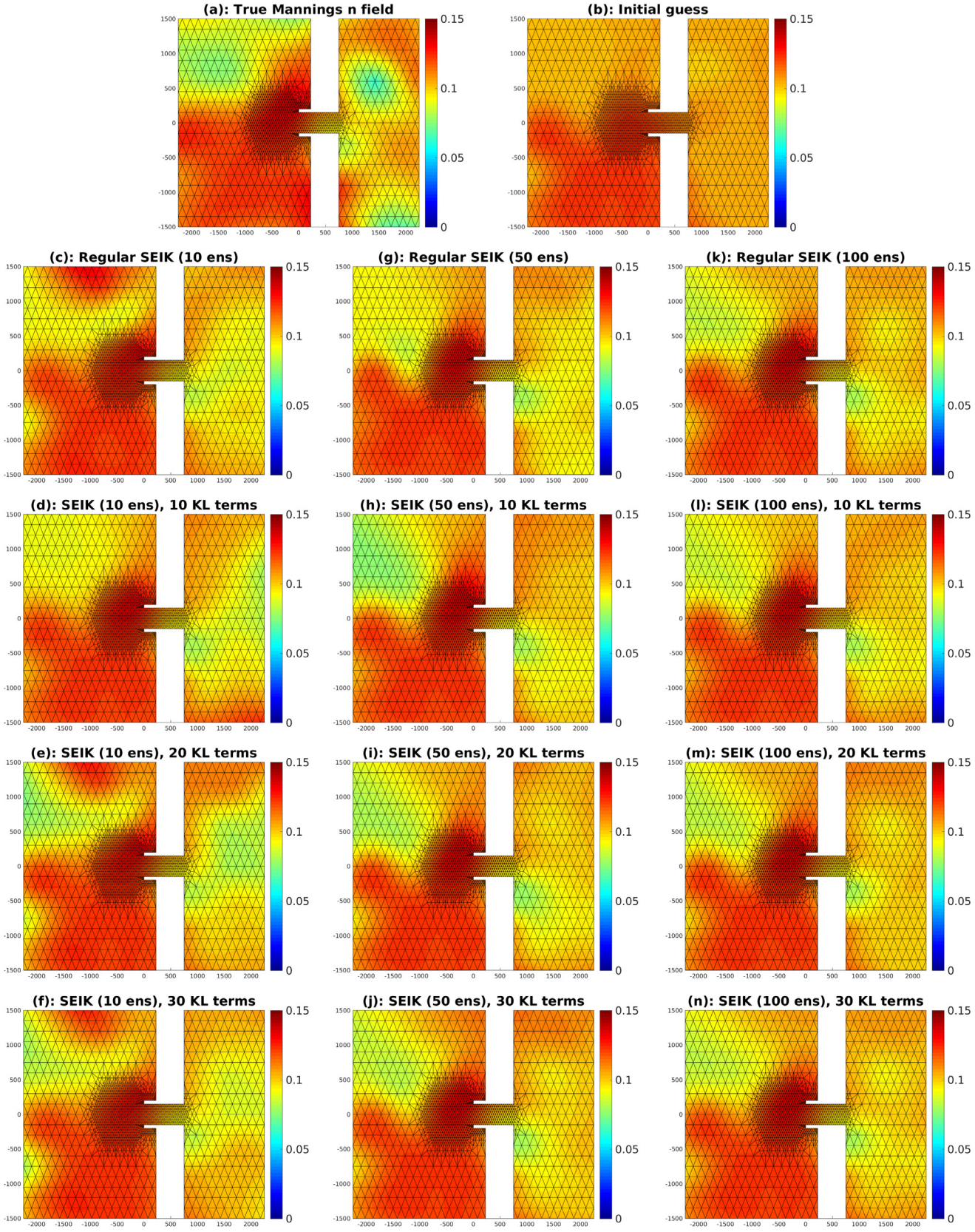

Fig. 10. Manning's n field estimates as inferred by different ensemble sizes and number of retained KL modes.

ensemble members (Fig. 9(a)). In all cases, SEIK-KL efficiently reduces the RMSE over time and leads to better final estimates than the regular SEIK. This suggests that the ensembles in the KL space exploit the statistical information retained by the KL modes to better span the parameter search space as compared to the full space spanned by limited ensembles.

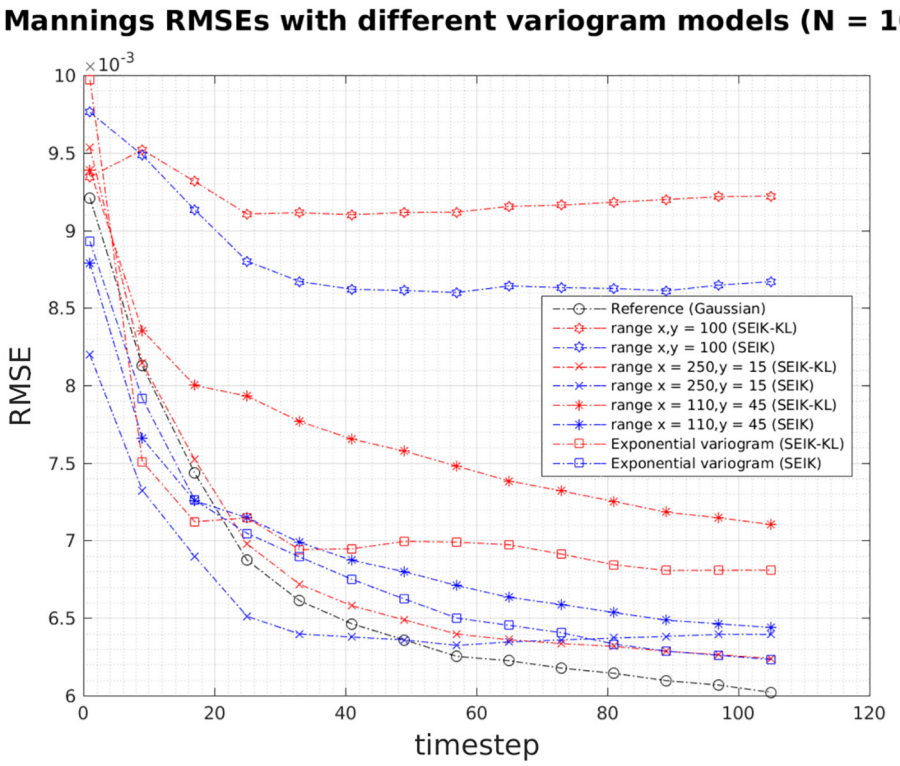

Fig. 11. Time series of the RMSEs of the Manning's n fields inferred from initial ensembles generated from various variogram models.

The convergence rate of SEIK-KL estimates to the truth is sensitive to the number of retained KL terms. For instance, when using 10 KL terms, the analysis converges rapidly toward the solution but quickly levels off after a few assimilation cycles. Increasing the simulation time does not improve the estimates when the ensemble size and number of retained KL terms are small. This is because a few KL terms are not enough to completely describe the variability of the search space. The filter stops improving after a few assimilation cycles due to the limited search directions, as also observed in Hoteit and Köhl (2006). Increasing the ensemble size in this case does not help much as a relatively small ensemble (often suggested to be of rank equal to the search space (Pham, 2001)) should be enough for efficient filtering. When the number of KL terms is increased (e.g., to 20 and 30 terms), the stagnant RMSEs pattern in the previous case is less pronounced, and the analysis starts to converge gradually, but slowly, toward the reference solution. Including more KL terms enables more search directions in the parameter space to be explored. As a result, more assimilation cycles may help SEIK-KL recover the reference field. Given a sufficiently large assimilation window, the SEIK-KL with 20 and 30 KL terms outperforms that of 10 KL terms.

Increasing the ensemble size (Fig. 9(b) and (c)), further reduces the RMSEs. However, the difference is not significant in the case of 10 KL terms. SEIK-KL inference with larger numbers of retained KL terms outperforms the cases with fewer KL terms for larger ensemble sizes. Fig. 9(b) shows that using 50 ensemble members, SEIK-KL with 30 KL terms starts to outperform the 10 KL-terms case at the end of the assimilation window. When 100 members are used (Fig. 9(c)), SEIK-KL with 20 and 30 KL terms leads to notably better estimates than those obtained using 10 KL terms. Due to less inherent variability, SEIK-KL with 20 KL terms performs poorer than the SEIK-KL with 30 KL terms for all tested ensemble sizes (Fig. 9(a), (b) and (c)). Also, SEIK with the full, nodally-defined parameter vector outperforms the SEIK-KL with 10-KL terms. Applying regular SEIK using 100 members leads approximately to the same level of RMSE as that of the best KL case (i.e., the 30-KL-terms case).

The sensitivity of the performance of SEIK-KL inference to the ensemble size is presented in the three plots in the right column of Fig. 9. In general, we see that as the ensemble size increases, the RMSE decreases, with an exception of the 10-KL terms case shown in Fig. 9(d); the RMSE produced by SEIK-KL using 50 members is smaller than that using 100 members. Again, this is the effect of using a few KL terms, which insufficiently describe the search space. This observation is consistent with Hoteit and Köhl (2006) and ElSheikh et al. (2013), who found that using 40 KL modes, small ensembles initially performed better, but are eventually outperformed by larger ensembles later in the simulation, as larger ensembles provide more exhaustive search directions.

In Fig. 10, we show the spatial plots of the inferred Manning's n coefficients using SEIK-KL with varying ensemble sizes and numbers of KL modes. The top row depicts the spatial structure of the true Manning's n field (Fig. 10(a)) and the initial guess (Fig. 10(b)), respectively. From the second row downward, each column represents an ensemble size and each row represents a number of retained KL modes. We notice that when using 10 ensemble members, the SEIK filter faces difficulty in recovering the reference field, even with a large number of KL terms; the best result is obtained using 10 KL modes (Fig. 10(d)). When using 50 ensemble members, all SEIK-KL inferences are better than the regular SEIK in recovering the Manning's n field. This is particularly clear in the area of low Manning's n values (cooler colors). When using 100 ensemble members, the filter's estimate is more accurate in all cases. The main Manning's n structures of the true parameter field are clearly recovered.

5.3. SEIK-KL sensitivity to inaccurate covariance model

The initial ensembles of the SEIK and the SEIK-KL have thus far been constructed based on the same covariance model from which the reference Manning's n field was generated. In many real-world applications, however, the initial covariance model might be poorly known. Here we examine the sensitivity of the performance of SEIK and SEIK-KL to perturbations in the covariance model used to generate both the initial ensembles of Manning's n coefficients and KL modes and

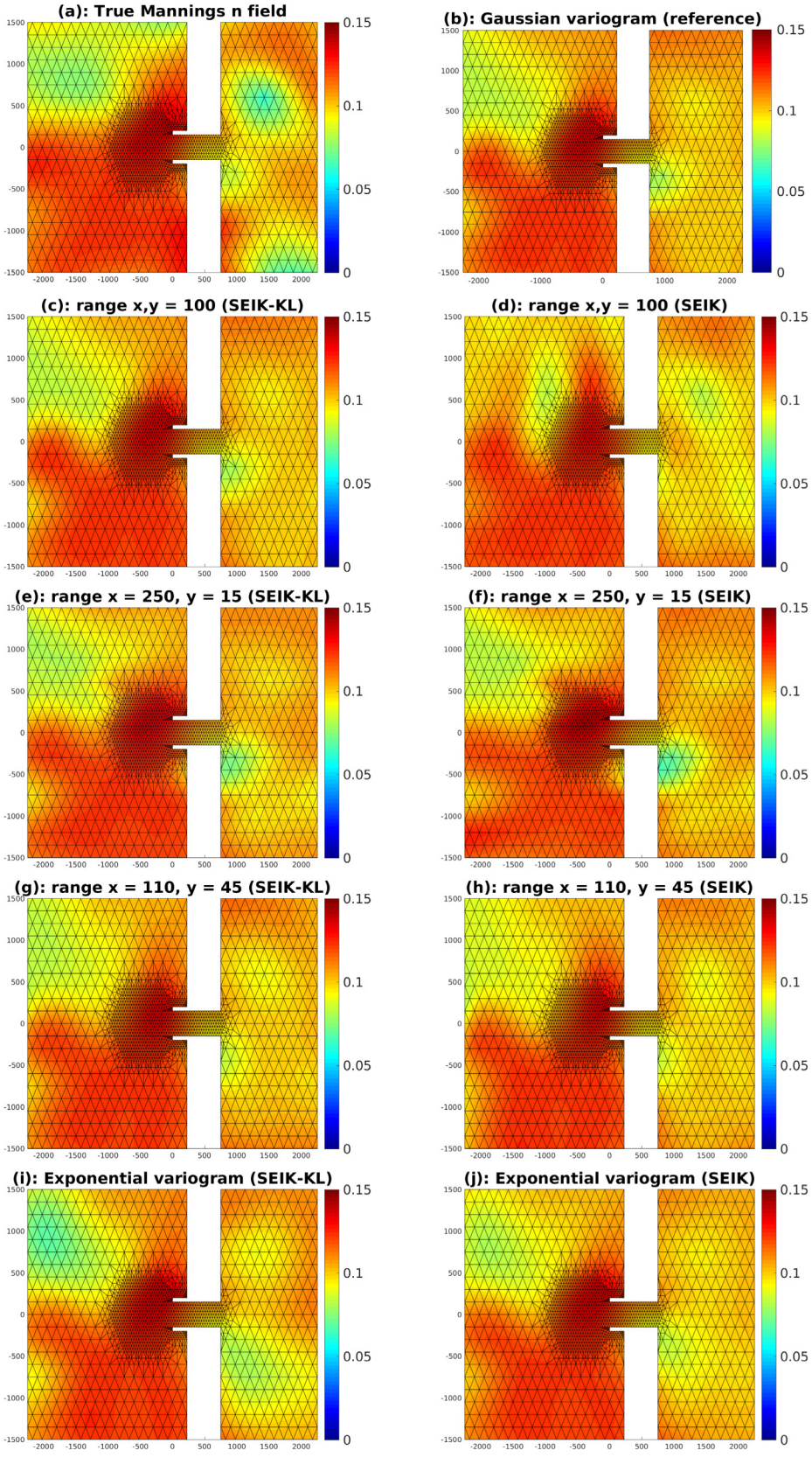

Fig. 12. Inferred Manning's *n* fields when various variogram models are used to generate the initial ensembles in KL space.

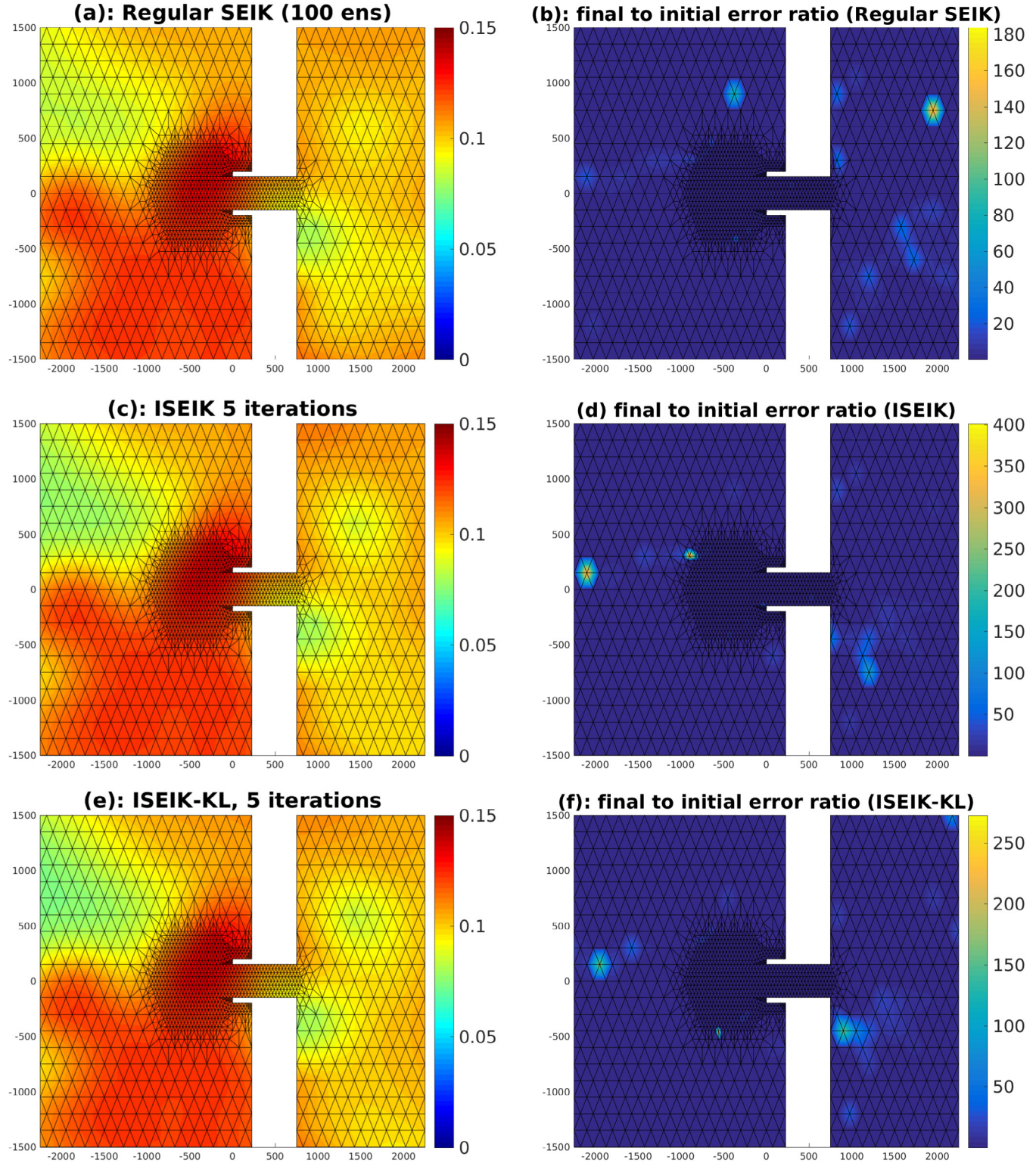

Fig. 13. Inferred Manning's n fields. 1st row: spatial plots of regular SEIK inference, 2nd row: spatial plots of ISEIK inference with 5 iterations, 3rd row: spatial plots of ISEIK-KL inference with 5 iterations.

compare the results against those obtained using the true (unperturbed) covariance model.

We first generate new realizations of Manning's n coefficients from different variogram models by perturbing some parameters in GCOSIM3D. 1) we use a Gaussian variogram with a range of 100 m in the x and y-directions (a perfect circle), 2) a Gaussian variogram with a range of 250 m in the x-direction and 15 m in the y-direction (i.e. the reference variogram is stretched in the x-direction and shrunk in y-direction), 3) a Gaussian variogram with range of 110 m in the x-direction and 45 m in the y-direction (i.e. the reference variogram is shrunk in x-direction and stretched in y-direction), and 4) an Exponential-type variogram.

In Fig. 11, we plot the time-series of the RMSE of the analyzed Manning's n field as estimated by SEIK and SEIK-KL (with 30 KL modes)

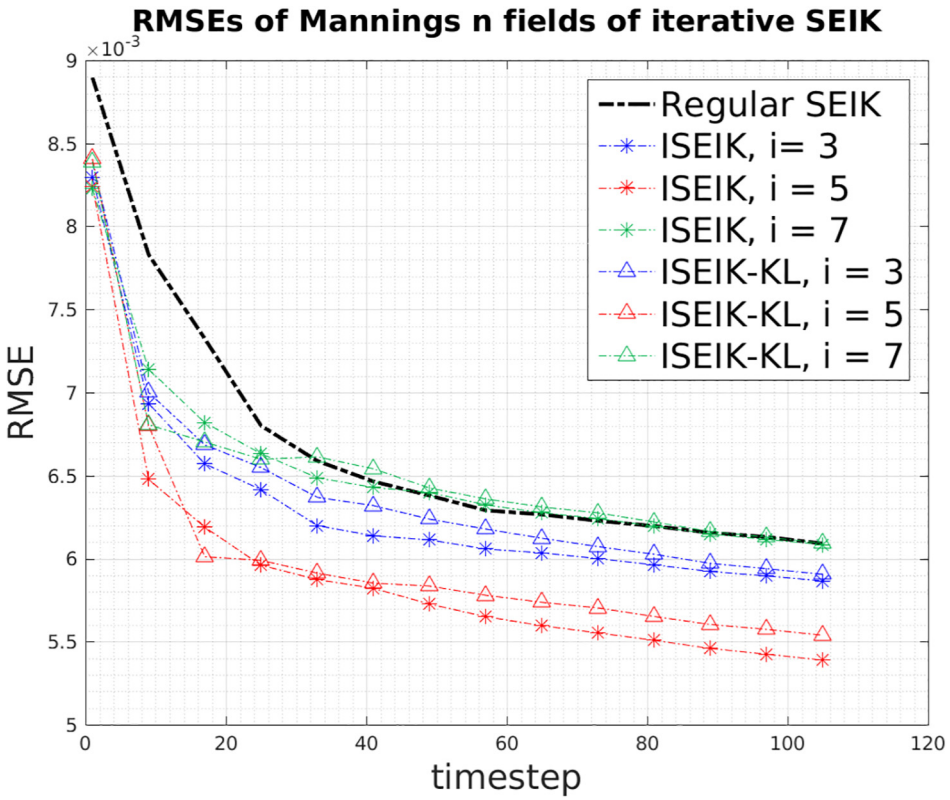

Fig. 14. Time-evolution of the RMSE of Manning's n field as inferred by the regular SEIK (in black) and iterative SEIK with various stopping criteria.

from the covariance models described above using 100 ensemble members. The first observation we make is that the performances of SEIK and SEIK-KL degrade when using any of the perturbed variograms. However, the degree at which the final RMSEs of the perturbed covariance cases differ from the reference case depends on the form of the perturbed covariance. For instance, using an Exponential variogram instead of Gaussian variogram (with the same mean, variance and correlation length) does not significantly alter the structure of the variogram, and as a result, the final RMSEs for the Exponential variogram cases are close to the reference case. The same can be said for the case in which we stretch or shrink the correlation range in the x- and y-directions (cases (2) and (3)). For the case where the perturbed variogram vastly differs from the reference case (i.e. case (1)), the resulting final RMSE is considerably larger than that of the reference case. The second observation is that in most cases with perturbed covariance models, the SEIK performs better than SEIK-KL, with the exception of case 3. The overall better performance obtained by SEIK is consistent with the results of Section 5.2, where the SEIK-KL was shown to outperform SEIK with small ensembles only ($N < 50$).

In Fig. 12, we show the spatial plots of inferred Manning's n fields as estimated by the SEIK and SEIK-KL for different covariance models. The results suggest that for all tested covariance models, the filter successfully recovers the main patterns of the true Manning's n coefficients over the studied domain, even for the perfect circle variogram case (1) where the inferred field exhibits the largest RMSE compared to the other cases. With sufficiently large ensembles, assimilated observations, and retained KL terms, SEIK-KL is capable of successfully capturing the main spatial structures of the reference Manning's n field, even when its reduced basis is constructed with imperfect KL-modes.

5.4. Iterative SEIK (ISEIK) in the full and KL spaces

Based on the above results, ISEIK is implemented using 100 members, 24 observation stations, and 5 days of simulation. The damping factor ω is chosen such that $\omega_{j+1} = \omega_j/2$, where j is the iteration number, and $\omega_0 = 1$. We start by studying the sensitivity of ISEIK to the number of iterations by performing 3, 5 and 7 iterations. The results of this experiment in terms of RMSE are presented in Fig. 14. ISEIK outperforms SEIK in all cases, notably reducing the RMSE for both the full-vector and the KL cases. The lowest RMSEs were obtained with 5 iterations.

In Fig. 13, we show the spatial plots of the estimates obtained using ISEIK. We notice particularly improved parameter recovery in the area of low Manning's n values compared to the regular SEIK. ISEIK also greatly reduces the ratio between the final error and the initial error in the right land-locked area. ISEIK in the full or KL spaces performs equally well in terms of reducing RMSE in all cases (Fig. 14). However, ISEIK-KL seems to better recover the spatial patterns of the reference Manning's field. This can be clearly observed in the area of low Manning's n values, colored in green (Fig. 13(e)): the recovered Manning's n structure as estimated by ISEIK in the KL space is more consistent with the reference than those produced in the full space. The computational cost of ISEIK is approximately the same as the regular SEIK. Thus, only modest increases in computational cost, ISEIK with a well-tuned damping factor performs comparably to the regular SEIK when using a much larger simulation window; with only 5 days of assimilation time, the final RMSE produced by ISEIK is as small as that obtained with the SEIK over 20 days of assimilation window.

6. Conclusions

We proposed a sequential data assimilation framework to estimate a 2D field of spatially varying Manning's n coefficients in the context of coastal ocean modeling. The proposed framework combines a deterministic ensemble Kalman filter (called SEIK), KL decomposition, and an iterative update scheme to improve the accuracy of estimation over an unaltered/baseline SEIK filter. Multi-Gaussian initial realizations of the Manning's n coefficients field are generated using a

sequential simulation algorithm. An empirical covariance matrix is computed from a sufficiently large number of realizations of Manning's n fields and used to construct KL coordinates representing the parameter in a reduced KL space. The KL expansion enhances the parameter search space and helps preserve the geostatistical characteristics of the parameter in the filter updates when the filter is implemented with a small number of ensembles.

Observation System Simulation Experiments (OSSEs) are conducted to evaluate the performance of the proposed framework. Synthetic water elevation data are generated by running ADCIRC with a reference Manning's n field, considered as the truth. SEIK is then implemented to estimate the Manning's n coefficients, both in the full nodally-defined and KL parameter space cases. We first study SEIK sensitivity to the ensemble size using the full parameter space and find that 100 ensemble members provide a reasonable trade-off between the filter performance and computational burden. Local analysis is also applied to alleviate the effect of spurious correlations between distant points. Increasing the number of observation stations from 15 stations to 24 stations further improves the filter performances. SEIK with the full nodally-defined parameter vector proves to be successful at recovering the main patterns of the true Manning's n field in our idealized setting.

We then conduct the SEIK inference in the KL space. For small ensembles (e.g., 10 ensembles) and only 10 terms in the KL expansion, a significant improvement is observed compared to the results obtained using the regular SEIK filter. We also find that increasing the ensemble size requires increasing the number of KL terms in order for the KL-SEIK to outperform the regular SEIK. For the case with 100 ensemble members, 30 KL modes are required. However, the sensitivity of the filter performance to the number of KL modes and the ensemble size is nonlinear. In all cases, the KL-SEIK consistently outperforms the regular SEIK when the Manning's n field is represented using 30 KL terms, which preserves almost 100% of the total variance of the parameter space.

Finally, iterative SEIK (ISEIK) is implemented at almost no additional computational cost to enhance the SEIK performances. We apply ISEIK to both the nodally-defined parameter vector and KL cases. Even with a small number of iterations (e.g., 3 iterations), improvements are clearly observed, with the best results obtained using 5 iterations for both the nodally defined and KL cases were.

Overall, our results demonstrate the relevance of sequential ensemble data assimilation filtering schemes for estimating spatially varying parameters in the context of coastal ocean modeling. Future work will focus on exploring approaches to further improve our inference framework by developing efficient schemes to update and evolve in the KL basis of the parameter search space based on incoming data along the method proposed by Sraj et al. (2016).

Acknowledgements

This work was supported by the King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia Grant No. CRG3-2016.

References

Yanagi, T., 1999. Coastal Oceanography, vol. 1 Springer.
Jelesnianski, C.P., 1966. Numerical computations of storm surges without bottom stress. Mon. Weather Rev. 94 (6), 379–394.
Bunya, S., Dietrich, J., Westerink, J., Ebersole, B., Smith, J., Atkinson, J., Jensen, R., Resio, D., Luettich, R., Dawson, C., et al., 2010. 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. Mon. Weather Rev. 138 (2), 345–377.
Dietrich, J., Bunya, S., Westerink, J., Ebersole, B., Smith, J., Atkinson, J., Jensen, R., Resio, D., Luettich, R., Dawson, C., et al., 2010. 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. Mon. Weather Rev. 138 (2), 378–404.
Dietrich, J., Westerink, J., Kennedy, A., Smith, J., Jensen, R., Zijlema, M., Holthuijsen, L., Dawson, C., Luettich Jr., R.L., Powell, M., et al., 2011. Hurricane gustav (2008) waves and storm surge: Hindcast, synoptic analysis, and validation in southern louisiana. Mon. Weather Rev. 139 (8), 2488–2522.
Luettich, R., Jr., Westerink, J., Scheffner, N.W., 1992. 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.
Vreugdenhil, C., 1994. Numerical Methods for Shallow-Water Flow. Water Science and Technology Library, Springer, Netherlands. URLhttps://books.google.com.sa/books?id=8uohwO1mgaEC.
Brummelhuis, P.T., Heemink, A., 1990. Parameter identification in tidal models with uncertain boundary conditions. Stochastic Hydrol. Hydraulics 4 (43), 193–208.
Sorensen, J.V.T., Madsen, H., 2006. Parameter sensitivity of three kalman filter schemes for assimilation of water levels in shelf sea models. Ocean Model. 11, 441–463.
Ghanem, R., Red-Horse, J., 1999. Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach. Fluid Dyn. Res. 133, 137–144.
Gharamti, M.E., Kadoura, A., Valstar, J., Sun, S., Hoteit, I., 2014. Constraining a compositional flow model with flow-chemical data using an ensemble-based kalman filter. Water Resour. Res. 50 (3), 2444–2467. http://dx.doi.org/10.1002/2013WR014830.
Sraj, I., Mandli, K., Knio, O., Dawson, C., Hoteit, I., 2014. Uncertainty quantification and inference of manning's friction coefficients using dart buoy data during the tōhoku tsunami. Ocean Model. 83, 82–97.
Giraldi, L., Le Maı̂tre, O.P., Mandli, K.T., Dawson, C.N., Hoteit, I., Knio, O.M., 2017. Bayesian inference of earthquake parameters from buoy data using a polynomial chaos-based surrogate. Comput. Geosci. 21 (4), 683–699. http://dx.doi.org/10.1007/s10596-017-9646-z.
Ishii, A., Kawahara, M., 2006. Parameter identification of manning roughness coefficient using analysis of hydrauric jump with sediment transport. Proceeding of the first Asian-Pacific Congress on Computational Mechanics 2, 1071–1076.
Mayo, T., Butler, T., Dawson, C., Hoteit, I., 2014. Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of manning's friction coefficient. Ocean Model. 76, 43–58.
Siripatana, A., Mayo, T., Sraj, I., Knio, O., Dawson, C., Le Maitre, O., Hoteit, I., 2017. Assessing an ensemble kalman filter inference of manning's n coefficient of an idealized tidal inlet against a polynomial chaos-based MCMC. Ocean Dyn. 1–28. http://dx.doi.org/10.1007/s10236-017-1074-z.
Arcement, G., Jr., Schneider, V., 1989. Guide for selecting manning's roughness coefficients for natural channels and flood plains 2339.
Ding, Y., Jia, Y., Wang, S.S.Y., 2004. Identification of manning's roughness coefficients in shallow water flows. J. Hydraul. Eng. 130 (6), 501–510. http://dx.doi.org/10.1061/(ASCE)0733-9429(2004) 130:6(501).
Budgell, W.P., 1987. Stochastic filtering of linear shallow water wave processes. SIAM J. Sci. Stat. Comput. 8 (2), 152–170. http://dx.doi.org/10.1137/0908027.
Khatibi, R.H., Williams, J.J., Wormleaton, P.R., 1997. Identification problem of open-channel friction parameters. J. Hydraul. Eng. 123 (12), 1078–1088.
Luettich, R., Westerink, J., 2005. ADCIRC: A Parallel Advanced Circulation Model for Oceanic, Coastal and Estuarine Waters. User's Manual for Version 48. R. Luettich.
Medeiros, S.C., Hagen, S.C., Weishampel, J.F., 2015. A random forest model based on lidar and field measurements for parameterizing surface roughness in coastal modeling. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (4), 1582–1590. http://dx.doi.org/10.1109/JSTARS.2015.2419817.
Navon, I., 1998. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. Dyn. Atmos. Oceans 27 (1), 55–79.
Posselt, D.J., Bishop, C.H., 2012. Nonlinear parameter estimation: comparison of an ensemble kalman smoother with a markov chain monte carlo algorithm. Mon. Weather Rev. 140 (6), 1957–1974.
Gharamti, M.E., Hoteit, I., Valstar, J., 2013. Dual states estimation of a subsurface flow-transport coupled model using ensemble kalman filtering. Adv. Water Resour. 60, 75–88. http://dx.doi.org/10.1016/j.advwatres.2013.07.011.
Altaf, M., Butler, T., Mayo, T., Luo, X., Dawson, C., Heemink, A., Hoteit, I., 2014. A comparison of ensemble kalman filters for storm surge assimilation. Mon. Weather Rev. 142 (8), 2899–2914.
Lions, J.L., 1971. Optimal Control of Systems Governed by Partial Differential Equations. Springer-Verlag, Berlin-Heidelberg.
Buckley, A.G., 1978. A combined conjugate-gradient quasi-newton minimization algorithm. Math. Program. 15 (1), 200–210. http://dx.doi.org/10.1007/BF01609018.
Gilbert, J.C., Lemaréchal, C., 1989. Some numerical experiments with variable-storage quasi-newton algorithms. Math. Program. 45 (1), 407–435. http://dx.doi.org/10.1007/BF01589113.
Nash, S.G., 2000. A survey of truncated-newton methods. J. Comput. Appl. Math. 124 (1–2), 45–59. http://dx.doi.org/10.1016/S0377-0427(00)00426-X. numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.
Altaf, M.U., Ambrozic, M., McCabe, M.F., Hoteit, I., 2016. A study of reduced-order 4dvar with a finite element shallow water model. Int. J. Numer. Meth. Fluids 80 (11), 631–647. http://dx.doi.org/10.1002/fld.4167. fld.4167.
Evensen, G., 2013. The ensemble kalman filter: theoretical formulation and practical implementation. Ocean Dyn. 53 (4), 343–367.
Altaf, M., Gharamti, M.E., Heemink, A., Hoteit, I., 2013. A reduced adjoint approach to variational data assimilation. Comput. Methods Appl. Mech. Eng. 254, 1–13. http://dx.doi.org/10.1016/j.cma.2012.10.003.
Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B., Khl, A., Heimbach, P., 2013. A mitgcm/dart ensemble analysis and prediction system with application to the gulf of mexico. Dyn. Atmos. Oceans 63, 1–23. http://dx.doi.org/10.1016/j.dynatmoce.2013.03.002.. URLhttp://www.sciencedirect.com/science/

article/pii/S0377026513000249.
Ho, Y., Lee, R., 1964. A bayesian approach to problems in stochastic estimation and control. IEEE Trans. Autom. Control 9 (4), 333–339. http://dx.doi.org/10.1109/TAC.1964.1105763.
Besag, D.H.J., Green, P., Mengersen, K., 1995. Bayesian computation and stochastic systems. Stat. Sci. 10, 3–41.
Tarantola, A., 2004. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
Hoteit, I., 2008. A reduced-order simulated annealing approach for four-dimensional variational data assimilation in meteorology and oceanography. Int. J. Numer. Methods Fluids 58 (11), 1181–1199.
Hastings, W.K., 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57 (1), 97–109. http://dx.doi.org/10.1093/biomet/57.1.97.. arXiv: http://biomet.oxfordjournals.org/content/57/1/97.full.pdf+html.
Gamerman, D., Lopes, H., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall/CRC, Boca Raton.
Law, K., Stuart, A., 2012. Evaluating data assimilation algorithms. Mon. Weather Rev. 140 (11), 3757–3782.
Evensen, G., 2009. Data Assimilation: The Ensemble Kalman Filter. Springer, Verlag.
Derber, J., Rosati, A., 1989. A global oceanic data assimilation system. J. Phys. Oceanogr. 19 (9), 1333–1347.
Anderson, J., 2001. An ensemble adjustment kalman filter for data assimilation. Mon. Weather Rev. 129 (12), 2884–2903.
Ait-El-Fquih, B., El Gharamti, M., Hoteit, I., 2016. A bayesian consistent dual ensemble kalman filter for state-parameter estimation in subsurface hydrology. Hydrol. Earth Syst. Sci. 20 (8), 3289–3307. http://dx.doi.org/10.5194/hess-20-3289-2016.. URLhttp://www.hydrol-earth-syst-sci.net/20/3289/2016/.
Burgers, G., van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble kalman filter. Mon. Weather Rev. 126 (6), 1719–1724 arXiv:doi: 10.1175/1520-0493(1998) 1261719:ASITEK2.0.CO;2, doi:10.1175/1520-0493(1998) 1261719:ASITEK2.0.CO;2.
Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. Mon. Weather Rev. 129 (3), 420–436.
Hoteit, I., Pham, D.-T., Blum, J., 2002. A simplified reduced order kalman filtering and application to altimetric data assimilation in tropical pacific. J. Mar. Syst. 36 (1–2), 101–127. http://dx.doi.org/10.1016/S0924-7963(02)00129-X.
Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M., Whitaker, J.S., 2003. Ensemble square root filters. Mon. Weather Rev. 131 (7), 1485–1490.
Hoteit, I., Pham, D.-T., Gharamti, M., Luo, X., 2015. Mitigating observation perturbation sampling errors in the stochastic enkf. Mon. Weather Rev.
Hoteit, I., Hoar, T., Collins, N., Anderson, J., Cornuelle, B., Heimbach, P., 2008. A MITgcm/dart ocean analysis and prediction system with application to the gulf of Mexico. AGU Fall Meeting Abstracts 1, 1288.
Evensen, G., 1994. Inverse methods and data assimilation in nonlinear ocean models. Physica (D) 77, 108–129.
Naevdal, G., Johnsen, L.M., Aanonsen, S.I., Vefring, E.H., 2005. Reservoir monitoring and continuous model updating using ensemble kalman filter. SPE 28 (2), 135–147. http://dx.doi.org/10.2118/84372-PA.
Hendricks Franssen, H.J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44 (9). http://dx.doi.org/10.1029/2007WR006505. w09408.
Tong, J., Hu, B.X., Yang, J., 2010. Using data assimilation method to calibrate a heterogeneous conductivity field conditioning on transient flow test data. Stoch. Env. Res. Risk Assess. 24 (8), 1211–1223. http://dx.doi.org/10.1007/s00477-010-0392-1.
DeChant, C.M., Moradkhani, H., 2012. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. Water Resour. Res. 48 (4). http://dx.doi.org/10.1029/2011WR011011. w04518.
ElSheikh, A.H., Pain, C.C., Fang, F., Gomes, J.L.M.A., Navon, I.M., 2013. Parameter estimation of subsurface flow models using iterative regularized ensemble kalman filter. Stoch. Env. Res. Risk Assess. 27 (4), 877–897. http://dx.doi.org/10.1007/s00477-012-0613-x.
Hoteit, I., Luo, X., Pham, D.-T., 2012. Particle kalman filtering: a nonlinear bayesian framework for ensemble kalman filters. Mon. Weather Rev. 140 (2), 528–542. http://dx.doi.org/10.1175/2011MWR3640.1.
Liu, B., Ait-El-Fquih, B., Hoteit, I., 2016. Efficient kernel-based ensemble gaussian mixture filtering. Mon. Weather Rev. 144 (2), 781–800. http://dx.doi.org/10.1175/MWR-D-14-00292.1.
Serafy, G.Y.H.E., Mynett, A.E., 2008. Improving the operational forecasting system of the stratified flow in osaka bay using an ensemble kalman filter-based steady state kalman filter. Water Resour. Res. 44, W06416 10.10292006WR005412.
Butler, T., Altaf, M., Dawson, C., Hoteit, I., Luo, X., Mayo, T., 2012. Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. Mon. Weather Rev. 140 (7), 2215–2231.
Korres, G., Triantafyllou, G., Petihakis, G., Raitsos, D., Hoteit, I., Pollani, A., Colella, S., Tsiaras, K., 2012. A data assimilation tool for the pagasitikos gulf ecosystem dynamics: methods and benefits. J. Mar. Syst. 94, S102–S117. http://dx.doi.org/10.1016/j.jmarsys.2011.11.004.. remote sensing, mathematical modeling and in-situ data for improving coastal management supporting tools.http://www.sciencedirect.com/science/article/pii/S0924796311002715.
Tamura, H., Bacopoulos, P., Wang, D., Hagen, S.C., Kubatko, E.J., 2014. State estimation of tidal hydrodynamics using ensemble kalman filter. Adv. Water Resour. 63, 45–56. http://dx.doi.org/10.1016/j.advwatres.2013.11.002.. URLhttp://www.sciencedirect.com/science/article/pii/S0309170813002030.
Altaf, M., Raboudi, N., Gharamti, M., Dawson, C., McCabe, M., Hoteit, I., 2014. Hybrid vs adaptive ensemble kalman filtering for storm surge forecasting. AGU Fall Meeting Abstracts 1, 3352.
Gómez-Hernández, J.J., Journel, A.G., 1993. Joint sequential simulation of multigaussian fields, 85–94. doi: http://dx.doi.org/10.1007/978-94-011-1739-5.
Pham, D., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. Mon. Weather Rev. 12 (5), 1194–1207.
Butler, T., Altaf, M., Dawson, C., Hoteit, I., Luo, X., Mayo, T., 2012. Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. Mon. Weather Rev. 140, 2215–2231.
Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129 (1), 123–137 arXiv:doi: 10.1175/1520-0493(2001) 1290123:ASEKFF2.0.CO;2, doi:10.1175/1520-0493(2001) 1290123:ASEKFF2.0.CO;2.
Anderson, J.L., 2003. A local least squares framework for ensemble filtering. Mon. Weather Rev. 131 (4), 634–642 arXiv:doi: 10.1175/1520-0493(2003) 1310634:ALLSFF2.0.CO;2, doi:10.1175/1520-0493(2003) 1310634:ALLSFF2.0.CO;2. https://doi.org/10.1175/1520-0493(2003) 1310634:ALLSFF2.0.CO;2.
Hoteit, I., Köhl, A., 2006. Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. J. Oceanogr. 62 (4), 539–550. http://dx.doi.org/10.1007/s10872-006-0074-2.
Chang, H., Zhang, D., 2014. History matching of statistically anisotropic fields using the karhunen-loeve expansion-based global parameterization technique. Computat. Geosci. 18 (2), 265–282. http://dx.doi.org/10.1007/s10596-014-9409-z.
Luo, X., Hoteit, I., 2014. Ensemble kalman filtering with residual nudging. Tellus A 64. URLhttp://www.tellusa.net/index.php/tellusa/article/view/17130.
Gharamti, M., Ait-El-Fquih, B., Hoteit, I., 2015. An iterative ensemble kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. J. Hydrol. 527, 442–457. http://dx.doi.org/10.1016/j.jhydrol.2015.05.004.. URLhttp://www.sciencedirect.com/science/article/pii/S002216941500339X.
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., 2008. A basin-to channel-scale unstructured grid hurricane storm surge model applied to southern louisiana. Mon. Weather Rev. 136 (3), 833–864.
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., 2011. Origin of the hurricane Ike forerunner surge. Geophys. Res. Lett. 38 (8).
Johnson, M.E., 2011. Multivariate Statistical Simulation. Springer Berlin Heidelberg, Berlin, Heidelberg. http://dx.doi.org/10.1007/978-3-642-04898-2.
Anderson, T.W., 1984. Multivariate Statistical Analysis. Wiley.
Loeve, M., 1947. Fonctions aleatoires de second order, Processus Stochastiques et Movement Brownien. Hermann, Paris.
Karhunen, K., 1947. Ueber lineare Methoden in der Wahrscheinlichkeitsrechnung, Annales Academiae scientiarum Fennicae. Series A. 1, Mathematica-physica.https://books.google.co.in/books?id=Dn3gSAAACAAJ.
Xiu, D., 2010. Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press, Princeton, NJ, USA.
Katterbauer, K., Arango, S., Sun, S., Hoteit, I., 2015. Multi-data reservoir history matching for enhanced reservoir forecasting and uncertainty quantification. J. Petrol. Sci. Eng. 128, 160–176. http://dx.doi.org/10.1016/j.petrol.2015.02.016.. URLhttp://www.sciencedirect.com/science/article/pii/S0920410515000686.
Yang, X., Delsole, T., 2009. Using the ensemble kalman filter to estimate multiplicative model parameters. Tellus A 61 (5), 601–609. http://dx.doi.org/10.1111/j.1600-0870.2009.00407.x.. arXiv: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0870.2009.00407.x.
Koyama, H., Watanabe, M., 2010. Reducing forecast errors due to model imperfections using ensemble kalman filtering. Mon. Weather Rev. 138 (8), 3316–3332. http://dx.doi.org/10.1175/2010MWR3067.1.
DelSole, T., Yang, X., 2010. State and parameter estimation in stochastic dynamical models. Physica D 239 (18), 1781–1788. http://dx.doi.org/10.1016/j.physd.2010.06.001.
Aksoy, A., Zhang, F., Nielsen-Gammon, J., 2006. Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. Mon. Weather Rev. 134 (10), 2951–2970.
Franssen, H.H., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44 (9).
Mayo, T., 2013. Data assimilation for parameter estimation in coastal ocean hydrodynamics modeling.
Kivman, G.A., 2003. Sequential parameter estimation for stochastic systems. Nonlinear Process. Geophys. 10 (3), 253–259.
Chen, Y., Oliver, D.S., Zhang, D., 2009. Data assimilation for nonlinear problems by ensemble kalman filter with reparameterization. J. Petrol. Sci. Eng. 66 (1), 1–14. http://dx.doi.org/10.1016/j.petrol.2008.12.002.. URLhttp://www.sciencedirect.com/science/article/pii/S0920410508001186.
Subramanian, A.C., Hoteit, I., Cornuelle, B., Miller, A.J., Song, H., 2012. Linear versus nonlinear filtering with scale-selective corrections for balanced dynamics in a simple atmospheric model. J. Atmos. Sci. 69 (11), 3405–3419. http://dx.doi.org/10.1175/JAS-D-11-0332.1.
Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble kalman filter. Adv. Water Resour. 28 (2), 135–147. http://dx.doi.org/10.1016/j.advwatres.2004.09.002.. URLhttp://www.sciencedirect.com/science/article/pii/S0309170804001605.
Wang, X., Bishop, C.H., Julier, S.J., better, Which is, 2004. an ensemble of positive-negative pairs or a centered spherical simplex ensemble? Mon. Weather Rev. 132 (7),

1590–1605 arXiv:doi: 10.1175/1520-0493(2004) 1321590:WIBAEO2.0.CO;2, doi:10.1175/1520-0493(2004) 1321590:WIBAEO2.0.CO;2.

Nerger, L., Janji, T., Schrter, J., Hiller, W., 2012. A unification of ensemble square root kalman filters. Mon. Weather Rev. 140 (7), 2335–2345. http://dx.doi.org/10.1175/MWR-D-11-00102.1.

Hoteit, I., Triantafyllou, G., Korres, G., 2007. Using low-rank ensemble kalman filters for data assimilation with high dimensional imperfect models. JNAIAM 2 (1–2), 67–78.

Jardak, M., Navon, I.M., Zupanski, M., 2010. Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation. Int. J. Numer. Meth. Fluids 62 (4), 374–402. http://dx.doi.org/10.1002/fld.2020.

Lorentzen, R.J., Naevdal, G., 2011. An iterative ensemble kalman filter. IEEE Trans. Autom. Control 56 (8), 1990–1995. http://dx.doi.org/10.1109/TAC.2011.2154430.

Engl, H., Hanke, M., Neubauer, A., 2000. Regularization of Inverse Problems. Springer.

Song, X., Shi, L., Ye, M., Yang, J., Navon, I.M., 2013. Numerical comparison of iterative ensemble kalman filters for unsaturated flow inverse modeling. Vadose Zone J. 13 (2). http://dx.doi.org/10.2136/vzj2013.05.0083.. arXiv: http:890//vzj.geoscienceworld.org/content/13/2/vzj2013.05.0083.full.pdf.

Luo, X., Hoteit, I., 2014. Ensemble kalman filtering with residual nudging: an extension to state estimation problems with nonlinear observation operators. Mon. Weather Rev. 142 (10), 3696–3712. http://dx.doi.org/10.1175/MWR-D-13-00328.1.

Nerger, L., Danilov, S., Hiller, W., Schröter, J., 2006. Using sea-level data to constrain a finite-element primitive-equation ocean model with a local seik filter. Ocean Dyn. 56 (5), 634–649. http://dx.doi.org/10.1007/s10236-006-0083-0.

Triantafyllou, G., Hoteit, I., Petihakis, G., 2003. A singular evolutive interpolated kalman filter for efficient data assimilation in a 3-d complex physical-biogeochemical model of the cretan sea. J. Mar. Syst. 40, 213–231.

Sraj, I., Le Mâtre, O., Knio, O., Hoteit, I., 2016. Coordinate transformation and polynomial chaos for the bayesian inference of a gaussian process with parametrized prior covariance function. Comput. Methods Appl. Mech. Eng. 298, 205–228.