

**Bayesian Inference of Manning's n coefficient of a
Storm Surge Model: an Ensemble Kalman filter
vs. a polynomial chaos-based MCMC**

Thesis by
Adil Siripatana

In Partial Fulfillment of the Requirements

For the Degree of

Masters of Science

King Abdullah University of Science and Technology, Thuwal,
Kingdom of Saudi Arabia

Insert Date (August, 2014)

Bayesian Inference of Manning's n coefficient of a Storm Surge Model: an Ensemble Kalman filter vs. a polynomial chaos-based MCMC

Committee Chairperson: Ibrahim Hoteit

Committee Member: Omar M. Knio

Committee Member: Shuyu Sun

Copyright ©2014

Adil Siripatana

All Rights Reserved

ABSTRACT

Bayesian Inference of Manning's n coefficient in a Storm Surge
Model Framework: comparison between Kalman filter and
polynomial based method

Adil Siripatana

Conventional coastal ocean models solve the shallow water equations, which describe the conservation of mass and momentum when the horizontal length scale is much greater than the vertical length scale. In this case vertical pressure gradients in the momentum equations are nearly hydrostatic. The outputs of coastal ocean models are thus sensitive to the bottom stress terms defined through the formulation of Manning's n coefficients. this thesis considers the Bayesian inference problem of the Manning's n coefficient in the context of storm surge based on the coastal ocean ADCIRC model.

In the first part if the thesis, we apply an ensemble-based Kalman filter, the singular evolutive interpolated Kalman (SEIK) filter to estimate both a constant Manning's n coefficient and a 2-D parameterized Manning's coefficient on one ideal and one of more realistic domain using observation system simulation experiments (OSSEs). We study the sensitivity of the system to the ensemble size. we also access the benefits from using an inflation factor on the filter performance.

To study the limitation of the Guassian restricted assumption on the SEIK filter,

we also implemented in the second part of this thesis a Markov Chain Monte Carlo (MCMC) method based on a Generalized Polynomial chaos (gPc) approach for the estimation of the 1-D and 2-D Manning's n coefficient. The gPc is used to build a surrogate model that imitate the ADCIRC model in order to make the computational cost of implementing the MCMC with the ADCIRC model reasonable.

We evaluate the performance of the MCMC-gPc approach and study its robustness to different OSSEs scenario. we also compare its estimates with those resulting from SEIK in term of parameter estimates and full distributions. we present a full analysis of the solution of these two methods, of the contexts of their algorithms, and make recommendation for fully realistic application.

ACKNOWLEDGEMENTS

I would like to declare the highest gratitude toward my advisor, Prof. Ibrahim Hoteit, for your guidance throughout the process of the thesis study, for providing me abundant amount of resources and connections needed for me to finish the work. I also would like to give a big thanks to Dr. Talea Mayo and Dr. Ihab Sraj for your knowledge and assistance, for teaching me the researching procedures in details. It is also my pleasure to thank Prof. Omar M. Knio, Prof. Shuyu Sun and Prof. Olivier Le Maitre, as a very strong committee, giving me such a deep and broad perspectives in the field of my on-going research. Finally, I would like to thank my family who always supported me.

TABLE OF CONTENTS

Examination Committee Approval	2
Copyright	3
Abstract	4
Acknowledgements	6
List of Figures	9
List of Tables	13
1 Introduction	14
1.1 Problem descriptions	16
1.2 Bayesian viewpoint on parameter estimation problems	18
1.2.1 The dynamic models and measurements	18
1.2.2 Bayesian formulation	20
1.2.3 Recursive formulation of Bayes' rule	21
1.3 Related studies	25
1.4 Thesis objectives and outline	26
2 ADCIRC model	28
2.1 The Shallow Water Equations	29
2.2 ADCIRC Features and Capabilities	32
2.2.1 Model Inputs	32
2.2.2 Model Outputs	33
2.2.3 Parallel implementation	33
2.3 Experimental setups	34
2.3.1 Observation Simulation System Experiments	34
2.3.2 Computational Domains	35
2.3.3 Parameterizing a field of Manning's n Coefficient	37

2.3.4	Parameter estimation methodology	37
3	Manning's coefficient estimation with SEIK	41
3.1	Kalman filtering applications in shallow water model	42
3.2	Kalman filtering and its variants	43
3.3	The SEIK filter for state estimation	47
3.4	The SEIK filter for parameter estimation	49
3.5	Numerical results and discussion	51
3.5.1	Constant Manning's n coefficient	52
3.5.2	A field of piecewise constant Manning's n coefficients	53
3.6	The effect of increasing ensembles size	55
3.7	The effect of using inflation factor	56
3.8	Discussions	57
4	PC-MCMC based for Manning's n coefficient estimation	74
4.1	MCMC theory and Algorithm	76
4.2	Polynomial Chaos, Theory and Methodology	79
4.2.1	Linear systems	81
4.2.2	Nonlinear systems	84
4.2.3	Polynomial chaos quadrature	85
4.2.4	Non-intrusive spectral projection (NISP)	88
4.3	ADCIRC uncertainties recast as stochastic variables	92
4.3.1	PC representation of ADCIRC model for the idealized ebb shoal	93
4.4	Manning's n coefficient estimation using PC-based ADCIRC model and MCMC	94
4.5	Conclusion	95
5	Concluding Remarks	109
	References	113

LIST OF FIGURES

2.1	idealized inlet with ebb shoal domain	35
2.2	Galveston Bay domain	36
3.1	point estimation of various constant Manning's n fields parameterized for the idealized inlet with ebb shoal case.	59
3.2	Estimates of various Manning's n fields parameterized by a single value for the idealized inlet with ebb shoal case using 10 ensemble members. The width of the pdf from the mean is equal to 3th standard deviation.	60
3.3	point estimation of various constant Manning's n fields parameterized for Galveston Bay.	61
3.4	Estimates of various Manning's n fields parameterized by a single value for Galveston Bay. The width of the pdf from the mean is equal to 3th standard deviation.	62
3.5	point estimates of α (left) and β (right) from various initial guesses for the idealized inlet with ebb shoal case.	63
3.6	pdf estimates of α (left) and β (right) from various initial guesses for the idealized inlet with ebb shoal case using 10 ensembles.	63
3.7	point estimates of β from various initial guesses with the same α for Galveston bay case.	63
3.8	point estimates of β from various initial guesses with the same α for Galveston bay case using 10 ensembles.	64
3.9	Comparison of the posterior distribution resulting from SEIK for con- stant Manning's n coefficient with three different number of ensembles and initial guess = 0.015. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	65
3.10	Comparison of the posterior distribution resulting from SEIK for con- stant Manning's n coefficient with three different number of ensembles and initial guess = 0.06. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	66

3.11	Comparison of the posterior distribution result from SEIK for piecewise Manning's n coefficient in the open ocean with different number of ensembles from several initial guesses. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	67
3.12	Comparison of the posterior distribution result from SEIK for piecewise Manning's n coefficient in the landlocked area with different number of ensembles from several initial guesses. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	68
3.13	Estimates of various true Manning's n fields parameterized by a single value for the idealized inlet with ebb shoal case using 10 ensemble members with inflation factor of 0.5. The width of the pdf from the mean is equal to 3th standard deviation.	69
3.14	Comparison of the posterior distribution result from SEIK for constant Manning's n coefficient between SEIK and SEIK with inflation factor, initial guess = 0.015. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	70
3.15	Comparison of the posterior distribution result from SEIK for constant Manning's n coefficient between SEIK and SEIK with inflation factor, initial guess = 0.06. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	71
3.16	Comparison of the posterior distribution result from SEIK for 2D piecewise Manning's n coefficient between SEIK and SEIK with inflation factor in the open ocean of idealized inlet case. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	72
3.17	Comparison of the posterior distribution result from SEIK for 2D piecewise Manning's n coefficient between SEIK and SEIK with inflation factor in the open ocean of idealized inlet case. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.	73

4.1	(Left) The distributions of the ADCIRC surrogate model output(water elevation) of idealized Ebb Shoal case with several polynomial orders at the selected points and times. The distributions is created by a million samples. (Right) Polynomial chaos coefficients at the corresponding points and times.	97
4.2	(Left) The distributions of the ADCIRC surrogate model output(water elevation) of idealized Ebb Shoal case with several polynomial orders at the selected points and times. The distributions is created by a million samples. (Right) Polynomial chaos coefficients at the corresponding points and times.	98
4.3	(Left) The distributions of the ADCIRC surrogate model output(water elevation) of idealized Ebb Shoal case with several polynomial orders at the selected points and times. The distributions is created by a million samples. (Right) Polynomial chaos coefficients at the corresponding points and times.	99
4.4	(Right) Water elevations simulated in time from the true model and the surrogate model at several quadrature points (different Manning's n coefficients). (Left) The residuals.	100
4.5	The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter(after 108 analysis iterations)and MCMC method.	101
4.6	The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter(after 108 analysis iterations)and MCMC method.	102
4.7	The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter(after 108 analysis iterations)and MCMC method.	103
4.8	The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter with inflation(after 108 analysis iterations)and MCMC method.	104
4.9	The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter with inflation(after 108 analysis iterations)and MCMC method.	105
4.10	The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter with inflation(after 108 analysis iterations)and MCMC method.	106

4.11	The comparisons between the posterior distributions of 2D Manning's n coefficient between SEIK filter(after 108th analysis iterations)and MCMC method, (Right) the distributions of α , (Left) the distribution of β	107
4.12	The comparisons between the posterior distributions of 2D Manning's n coefficient between SEIK filter(after 108th analysis iterations)and MCMC method, (Right) the distributions of α , (Left) the distribution of β	108

LIST OF TABLES

2.1	Values of the Manning's n coefficient for various surfaces. the left columns in the tables describe the land characteristics and the right columns are the associated empirically defined Manning's n coefficients.	31
4.1	The type of generalized polynomial chaos and their corresponding random variables	80
4.2	The random input parameters for ADCIRC	92
4.3	The random input parameters for RMSE between the true model and the surrogate model (water elevations)	94

Chapter 1

Introduction

Coastal and estuarine systems are home to more than a half of the human population [1, 2]. Understanding and forecasting the dynamics of coastal and estuarine systems is critical for human 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 moderate conditions, accurately forecasting the ocean states (e.g. tides and coastal flows) is crucial for human's marine and related activities. Coastal ocean modeling has been extensively utilized to simulate the main features of the ocean circulation such as water surface elevation and currents, for the purposes of conservation, contaminant transport, development of coastal structures (e.g. bridges, dams and breakwaters) and emergency and economic planing.

Tidal and coastal flows are very complex phenomena and cannot be easily predicted with even the most comprehensive mathematical models and computational facilities available today [1, 2]. Instead of employing complex computational fluid dynamics, which require prohibitive computational cost, most researchers resort on model simplifications to develop computationally less intensive techniques. The goal is to obtain good enough models while balancing the usability and the resources requirements.

State-of-the-art coastal ocean models are based on the shallow water equations

(SWEs), which are derived from the depth-integrating Navier–Stokes equations, assuming hydrostatic pressure and horizontal length scales that are large in comparison to the vertical length scales. The result of this derivation is a first–order hyperbolic continuity equation for water elevation coupled with the momentum equation for horizontal depth-averaged velocities. The shallow water equations describe the flow below a pressure surface in a fluid. Even at this level of simplification, the model contains many parameters, carrying uncertainties on which the accuracy of model prediction depends [1]. Moreover, many of the parameters cannot be directly measured or the relevant data to estimate these parameters are difficult to collect [61]. Thus the values of these parameters must be inferred and estimated using the data from different, but simpler to observe variables, which are not always be very accurate [3]. Quantifying and reducing the uncertainties in the model outputs associated with the uncertainties in the parameter is essential for more reliable and robust storm surge predictions [4]. Of particular importance in tides and storm surge prediction is the Manning’s n coefficient of roughness, introduced in the SWEs through the bottom stress as components in the momentum equation [5, 6].

The traditional approach to estimate model parameters in coastal models is the variational approach. This deterministic approach consists of looking for the parameter that best fit the model to the collected data. The disadvantage of variational methods is that they can be computational intensive and are difficult to implement because they normally require an adjoint model [67]. In addition, these methods are not formulated for real-time parameter estimation, in the sense they are not suitable for applications with highly transient and/or abrupt flow regime changes, eg. as a hurricane approaches landfall. Consequently recently developed statistical data assimilation methods became more popular because they are non-intrusive and could be implemented with reasonable computational cost. They also provide a framework to quantify the uncertainty associated with the predicted model state. These methods

were first developed for the estimation of the model state [4]. They have been recently reformulated to include the estimation of the model parameter as well, though it is not yet widely used in coastal modeling. Moreover, statistical data assimilation methods are not based on physical principles, and may violate mass and momentum conservation principles [73].

1.1 Problem descriptions

Statistical data assimilation methods are based on the Bayes rule, which states that the probability distribution function (pdf) of the unknown parameters given available data [7, 8], can be written as

$$\pi(\theta | y) = \frac{p(y | \theta)\pi_{pr}(\theta)}{\pi_Y(y)} \quad (1.1)$$

with $\pi(\theta | y)$ is the pdf of the unknown parameter θ given the observation \mathbf{y} , $p(y | \theta)$ is the likelihood function of obtaining data \mathbf{y} if we have parameter value θ , $\pi_{pr}(\theta)$ is the prior distribution defines the prior knowledge of the parameters and finally, $\pi_Y(y)$ is the scaling factor, which is the density function of all possible measurements.

One popular approximation framework used to compute $p(y | \theta)$ is the well-known Kalman filter (KF). The KF provides the best estimation of a quantity of interests (QoIs) given the observations arrive sequentially in time for a linear system with Gaussian noise assumption [9].

The parameter estimation problems is often dealing with nonlinear system. One popular approach that has been developed in oceanography to cope with the non-linearity of the ocean modeling problem is the Ensemble Kalman filter (EnKF) and its variants [10, 11, 12, 13]. The EnKF avoid the linearity of the system by using an ensemble of realizations of the unknown parameter, which is used to estimate the

first two moments of the quantity estimated assuming a Gaussian distribution. The parameter correction step can then be applied every time new observations become available. EnKF methods can be divided into stochastic EnKF which perturb the observations in each assimilation cycle and deterministic EnKF which do not require perturbing the observations [14, 15].

In the last decade, Bayesian framework was established as a preferred method for uncertainty quantification and calibration of computer simulation. Since then the interest in the application of Bayesian approach in calibration of parameters in stochastic models have grown exponentially [16, 17, 18, 19, 20].

One of the appealing features of Bayesian approach is the ability to provide the complete posterior estimation and statistics. However, this approach requires statistical sampling techniques to obtain the solutions of the Bayesian calibration and uncertainty propagation. For this purpose, Markov Chain Monte Carlo (MCMC) method is one of the most popular sampling technique [16, 21, 22]. However, MCMC requires a large number of samples for satisfactory approximation (often in the range $10^3 - 10^6$), rendering Bayesian framework computational prohibitive and the development of an efficient solution generation is essential. To this end, Polynomial Chaos (PC) based spectral projection method has been investigated extensively. PC methods, particularly the generalized polynomial chaos (gPC), have become one of standard approaches for solving stochastic problems to propagate and quantify uncertainties in various disciplines including both physical [23, 24, 25] and chemical systems [26, 27, 28, 29]. These methods approximate the model variables and parameters in terms of spectral expansion in an orthogonal polynomial basis according to their probabilistic distributions. The resulting surrogate models then can be used to efficiently produce the solutions and the statistical properties of quantities of interests (QOIs). The current popular trend is to incorporate Bayesian inference, gPC, Markov chain Monte Carlo (MCMC) and field measurement in data assimilation

process which continuously updates the state variables, parameters. Although earlier attempts at using PC for turbulent fluid flow modeling were not successful [30, 31], it was found satisfactory to obtain efficient surrogate model for stochastic finite element and stochastic fluid flow problems [32, 33]. In recent development, Xiu and his coworkers [34, 35, 36] have developed a generalized version of polynomial chaos called generalized polynomial chaos (gPC). Since then gPC has been applied for uncertainty propagation through computational simulators in many field of engineering.

1.2 Bayesian viewpoint on parameter estimation problems

The traditional approaches in parameter estimation are concerned with minimizing cost function, which measured the distance between the model solution and the measurements in addition to the deviation between the estimated prior and its prior, with respect to the parameters [66]. Alternatively, parameter estimation problems can be viewed as an attempt to formulate the joint distribution between the model state and the parameters, given a set measurement and a dynamical model with known uncertainties.

1.2.1 The dynamic models and measurements

Consider the following equations define the model associates with initial and boundary conditions on the spatial domain D , boundary ∂D and a set of measurements,

$$\frac{\partial \psi(\mathbf{x}, t)}{\partial t} = \mathbf{G}(\psi(\mathbf{x}, t), \boldsymbol{\alpha}(\mathbf{x})) + \mathbf{q}(\mathbf{x}, t), \quad (1.2)$$

$$\boldsymbol{\psi}(\boldsymbol{x}_0, t) = \Psi_0(\boldsymbol{x}) + \boldsymbol{a}(\boldsymbol{x}), \quad (1.3)$$

$$\boldsymbol{\psi}(\boldsymbol{x}, t)|_{\partial D} = \Psi_b(\boldsymbol{\xi}, t) + \boldsymbol{a}(\boldsymbol{\xi}, t), \quad (1.4)$$

$$\boldsymbol{\alpha}(\boldsymbol{x}) = \boldsymbol{\alpha}_0(\boldsymbol{x}) + \boldsymbol{\alpha}'(\boldsymbol{x}), \quad (1.5)$$

$$\boldsymbol{M}[\boldsymbol{\psi}, \boldsymbol{\alpha}] = \boldsymbol{d} + \boldsymbol{\epsilon} \quad (1.6)$$

Here $\boldsymbol{\psi}(\boldsymbol{x}, t) \in \mathfrak{R}^{n_\psi}$ is the model state vector consists of spatially-time varying variables (i.g. water elevation computed in ADCIRC model). The dynamic parts of the system is first defined by the nonlinear operator $\boldsymbol{G}(\boldsymbol{\psi}(\boldsymbol{x}, t) \in \mathfrak{R}^{n_\psi}$. The model state is initialized and evolve from the initial condition $\Psi_0(\boldsymbol{x}) \in \mathfrak{R}^{n_\psi}$ define in (1.3). The boundary condition is defined in (1.4) where $\boldsymbol{\xi}$ is the coordinate running over the domain's boundary. In addition, the poorly known model parameters is defined in (1.5). This parameter can either be a vector of spatial fields or a vector of scalar with is a stationary process i.e. these parameters do not change in time. (1.5) specify the measurements as a function of the model state and the parameters, this can be a direct point measurements or a more complex form of function related to the model state. Finally, the system is closed by the additional terms at the end of each and every equations (1.2 - 1.6). Without these terms, the system over-determined with no solution. However, simply adding these errors term will not solve the system. The statistical assumptions about these errors must be incorporated in order for the system to be solved, i.g. in many applications these errors is assumed to be Gaussian which means equal to zero and subjected to specific error covariances.

1.2.2 Bayesian formulation

Following the problem formulation from equations (1.2 - 1.6), the model variables, the indisposed parameters, the initial boundary and initial conditions and the measurements are now treated as random variables. The solution of the system, which is subjected to uncertainties, cannot be described merely by point estimations but instead can be represented by pdfs.

Suppose the joint pdf of the model state $f(\boldsymbol{\psi}, \boldsymbol{\alpha})$ and the likelihood function comprise of the measurements $f(\mathbf{d}|\boldsymbol{\psi}, \boldsymbol{\alpha})$ are known, Bayes' theorem for the parameter estimation dictates

$$f(\boldsymbol{\psi}, \boldsymbol{\alpha}|\mathbf{d}) \propto f(\boldsymbol{\psi}, \boldsymbol{\alpha})f(\mathbf{d}|\boldsymbol{\psi}, \boldsymbol{\alpha}) \quad (1.7)$$

In general, the normalized function is used as a denominator on the right-hand-side of (1.7) to equalized this expression.

We have previously defined the boundary condition ,initial condition and the parameter prior as random variables $f(\boldsymbol{\psi}_0)$, $f(\boldsymbol{\psi}_b)$ and $f(\boldsymbol{\alpha})$. The likelihood Bayes' theorem is now expressed as

$$f(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b) = f(\boldsymbol{\psi}, \boldsymbol{\alpha}|\boldsymbol{\psi}_0, \boldsymbol{\psi}_b)f(\boldsymbol{\psi}_0)f(\boldsymbol{\psi}_b) \quad (1.8)$$

$$= f(\boldsymbol{\psi}|\boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b)f(\boldsymbol{\psi}_0)f(\boldsymbol{\psi}_b)f(\boldsymbol{\alpha}) \quad (1.9)$$

(1.7) is then written as

$$f(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b|\mathbf{d}) \propto f(\boldsymbol{\psi}|\boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b)f(\boldsymbol{\psi}_0)f(\boldsymbol{\psi}_b)f(\boldsymbol{\alpha})f(\mathbf{d}|\boldsymbol{\psi}, \boldsymbol{\alpha}) \quad (1.10)$$

Here $f(\boldsymbol{\psi}|\boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b)$ is the prior density of the model solution given the parameter, initial and boundary conditions.

1.2.3 Recursive formulation of Bayes' rule

In order to see how the sequential data assimilation is derived from Bayes' rule, it is convenient to assume the model as the discretization in time. The model state is represented at the fixed time interval $\boldsymbol{\psi}_i(\mathbf{x}) = \boldsymbol{\psi}(\mathbf{x}, t_i)$, $i = 0, 1, \dots, k$.

Assuming the model is a first order Markov process, the pdf for the model integration from time t_{i-1} to t_i is defined as $f(\boldsymbol{\psi}_i|\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}, \boldsymbol{\psi}_b(t_i))$. The joint pdf of the model state and the parameters can now be written as

$$f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b) \propto f(\boldsymbol{\psi}_0)f(\boldsymbol{\psi}_b)f(\boldsymbol{\alpha}) \prod_{i=1}^k f(\boldsymbol{\psi}_i|\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}, \boldsymbol{\psi}_b) \quad (1.11)$$

Now we can assume that the measurements are divided into subsets of measurement vector $\mathbf{d}_j \in \mathbb{R}^{m_j}$, acquired at time $t_{i(j)}$, with $j = 1, \dots, J$ and $0 < i(1) < i(2) < \dots < i(J)$. Together with the uncorrelated measurement errors assumption in time, the likelihood function is rewritten as

$$f(\mathbf{d}|\boldsymbol{\psi}, \boldsymbol{\alpha}) = \prod_{j=1}^J f(\mathbf{d}_j|\boldsymbol{\psi}_{i(j)}, \boldsymbol{\alpha}) \quad (1.12)$$

Combining the results from (1.11) and (1.12), Bayes' theorem can now be written as

$$f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b|\mathbf{d}) \propto f(\boldsymbol{\psi}_0)f(\boldsymbol{\psi}_b)f(\boldsymbol{\alpha}) \prod_{i=1}^k f(\boldsymbol{\psi}_i|\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) \prod_{j=1}^J f(\mathbf{d}_j|\boldsymbol{\psi}_{i(j)}, \boldsymbol{\alpha}) \quad (1.13)$$

Or the more general form where the model is not assumed to be the first order Markov process,

$$f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \mathbf{d}) \propto f(\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_b) f(\boldsymbol{\alpha}) \prod_{i=1}^k f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{l \neq i}, \boldsymbol{\alpha})) \prod_{j=1}^J f(\mathbf{d}_j | \boldsymbol{\psi}_{i(j)}, \boldsymbol{\alpha}) \quad (1.14)$$

With the assumption of correlated model errors, equation (1.14) establishes the most general formulation of the state and parameter estimation problem.

In the application of ocean and atmospheric modeling, it is not over-simplifying to assume the dynamical process as a first order Markov process [67]. A general smoother and filter could be derived from the Bayesian formulation given in (1.13) as follows:

$$\begin{aligned} f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \mathbf{d}) &\propto f(\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_b) f(\boldsymbol{\alpha}) \\ &\prod_{i=1}^{i(1)} f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})) f(\mathbf{d}_1 | \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}) \\ &\vdots \\ &\prod_{i=i(J-1)+1}^{i(J)} f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})) f(\mathbf{d}_J | \boldsymbol{\psi}_{i(J)}, \boldsymbol{\alpha}) \\ &\prod_{i=i(J)+1}^k f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})) \end{aligned} \quad (1.15)$$

It is simple to break apart the this expression as the sequential evaluation in time as follows:

$$\begin{aligned}
& f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \boldsymbol{d}_1) \propto \\
& f(\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_b) f(\boldsymbol{\alpha}) \\
& \prod_{i=1}^{i(1)} f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})) f(\boldsymbol{d}_1 | \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}),
\end{aligned} \tag{1.16}$$

$$\begin{aligned}
& f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{i(2)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \boldsymbol{d}_1, \boldsymbol{d}_2) \propto \\
& f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \boldsymbol{d}_1) \\
& \prod_{i=i(1)+1}^{i(2)} f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})) f(\boldsymbol{d}_2 | \boldsymbol{\psi}_{i(2)}, \boldsymbol{\alpha}),
\end{aligned} \tag{1.17}$$

\vdots

$$\begin{aligned}
& f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{i(J)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \boldsymbol{d}_1, \dots, \boldsymbol{d}_J) \propto \\
& f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{i(J-1)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \boldsymbol{d}_1, \dots, \boldsymbol{d}_{J-1}) \\
& \prod_{i=i(J-1)+1}^{i(J)} f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})) f(\boldsymbol{d}_J | \boldsymbol{\psi}_{i(J)}, \boldsymbol{\alpha}),
\end{aligned} \tag{1.18}$$

$$\begin{aligned}
& f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \boldsymbol{d}_1, \dots, \boldsymbol{d}_J) \propto \\
& f(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{i(J)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_b | \boldsymbol{d}_1, \dots, \boldsymbol{d}_J) \\
& \prod_{i=i(J)+1}^k f(\boldsymbol{\psi}_i | (\boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})).
\end{aligned} \tag{1.19}$$

The expressions (1.16-1.19) demonstrate the implementation of the propagation

rule in discrete form on Bayes' theorem, on the condition that the dynamical model in first order Markov process with independent measurement errors in time. In this case the measurements can be assimilated sequentially at any data-available time. Overall, these expressions represent filtering approach in data assimilation problem.

(1.16) start off by computing the joint conditional pdf from the product of the priors $f(\boldsymbol{\alpha})$, $f(\boldsymbol{\psi}_0)$ and $f(\boldsymbol{\psi}_b)$ with the likelihood function containing measurement \mathbf{d}_1 . In Kalman filter scheme, this is equivalent to evaluating both forecasting and analysis steps between $[t_1, t_{i-1}]$ interval. The sequence is then proceeded in (1.16) with the introduction of the new observation \mathbf{d}_2 , but the priors is replaced by the joint conditional pdf from the previous step and so on. These process is then continue until the last assimilation cycle where the measurements are available. Then the model forecasting continue toward the finishing time of the model run from (1.18).

It is worth emphasize that although we have arrived at (1.16-1.19) where the full inversion scheme is formulated sequentially. The sequential-in-time algorithm, while intuitive, is not the fundamental the the update. It is convenient in the case where the model is assume to be the first order Markov process where the posterior pdf of the present step can be generated if the model state vector form the previous time step is known so the sequential scheme is well-suited to apply to the online model forecasting where the data is made available from sensor networks in real-time.

In general the calculation of the posterior distribution can be done by processing all the measurements simultaneously as is normally done in variational formulation. In the next section, the approach to sampling the joint conditional pdf from full inverse formulation: MCMC, will be introduced. The posterior distribution of the estimated parameters from MCMC is used to compare the pdf from SEIK approach.

The detail derivations of the Ensemble Kalman Smoother (EnKS) and the Ensemble Kalman Filter (EnKF) from Bayes' theorem can be found in [66] for further reading.

1.3 Related studies

Uncertainties in spatially variable parameter can be represented as a particular realization of a random field [37, 38]. We aim to estimate low-dimensional representations of the Manning's n field. Since only small number of realizations are needed to estimate second-order statistics, this representation is thus suitable for low-rank non-intrusive data assimilation methodologies such as the singular evolutive interpolated Kalman (SEIK) filter [39, 40, 41]. Furthermore, since modeling the uncertainty propagation of the uncertain the Manning's n coefficient through the model by SEIK requires a large number of samplings which in turn affects the time of the simulations, Polynomial chaos projection is implemented to construct surrogate versions of the highly complicated original model. The PCs implementation greatly reduces the time of simulations and at the same time neglect the Gaussian assumption enforced by Kalman filter yielding the full distribution of state vectors at forecasting step in statistical data assimilation scheme.

Among the recent work in the field, the work of Mayo *et al.* [4, 2] and Alexanderian *et al.* [42] are closely related to this work and need special mention here. Mayo *et al.* reformulated statistical data assimilation to estimate the bottom stress terms in a 2-D coastal ocean model. A square root Kalman filter was implemented to minimize the amount of noise due to the data assimilation. Specifically, the authors inverted synthetic water elevation data by implementing the singular interpolated evolutive Kalman (SEIK) filter to the Advanced Circulation (ADCIRC) coastal ocean model. In all cases they studied, true field of Manning's n coefficients were accurately recovered. With the methodology, they improved the accuracy of forecasted water elevations in all experiments. When estimating the values of the Mannings n coefficients directly, they found smaller values could be recovered from nearly any initial guesses. In estimating larger values, however, either a relatively accurate initial guess or many data assimilation cycles (updates using data) were required for the method to converge

to the true parameter value.

Alexanderian and his co-workers [42] used the polynomial chaos (PC) expansion to propagate parametric uncertainties in ocean global circulation model (OGCM), focusing on short-time, high resolution simulations of the Gulf of Mexico (GOM), using the hybrid coordinate ocean model, with wind stresses corresponding to hurricane Ivan. A sparse spectral projection approach was implemented to propagate and quantify parametric uncertainty in an OGCM. A non-intrusive spectral projection scheme, based on a Smolyak sparse quadrature grid, was used to derive the PC representation of the stochastic response of selected QoIs. The quality of the PC representation was examined, refined and utilized to compute distributions of quantities of interest (QoIs) and to analyze the local and global sensitivity of these QoIs to the uncertain parameters. Their work provides an essential tool (gPC) and illustrates the advantage gPC to create a surrogate model used to generate the response in very a efficient way, facilitating the calculation of various statistics. In this thesis similar method the surrogate model for Markov chain Monte Carlo (MCMC) is derived to obtain probability distribution function (pdf) of the responses. The development gPC in this thesis is described in Chapter 4.

1.4 Thesis objectives and outline

In this work, we exploit a specific coastal ocean model to solve the SWEs and implement the SEIK filter to estimate the Manning's n coefficient. By conducting observation system simulation experiments on coastal ocean model, we attempt to recover the true Manning's n coefficient from sets of initial guesses and available data. The OSSEs is divided into two parts. We first conduct the OSSES using SEIK filter, which well-suited to the parameter estimation problem with large number of uncertain parameters and for real-time online applications, however, has a limitation on

the correctness of the posterior distribution of the parameter estimated due to its Gaussianity. The solution to avoid the Gaussian restriction of Kalman based methods is to use MCMC sampling based method, which produces full distribution of the posterior. As the implementation of MCMC becomes too costly for highly sophisticated model, PC is used to reduce the size of the model in order to apply MCMC efficiently. Our final objective is to compare the posterior distributions produced by OSSEs from both Kalman based method and PC-MCMC based.

In the remainder of this thesis, we will formally describe all ingredients used in this research. In Chapter 2 provides the description of shallow water equations (SWEs) as implemented in advanced circulation model (ADCIRC) and used in this work. Chapter 3 focuses on the description of Manning's n coefficients and popular methods the estimate this important parameters. Since some part of the thesis is to follow up the work of Mayo [4, 2], it is appropriate to the discuss on the principle of data assimilation, Kalman filters and its variation, particularly singular evolutive Kalman filter (SEIK). Similar layout is used for Chapter 4 but their main focuses are on the development of surrogate model using generalized polynomial chaos (gPC) and the uncertainty characterization using Markov chain Monte Carlo (MCMC) respectively. The experimental setups and the results for each topic are embedded in the corresponding Chapters as described. Finally, Chapter 5 is designed to wrap up the findings, discussions and the concluding remarks.

Chapter 2

ADCIRC model

The sophisticated coastal ocean model namely ADvanced CIRCulation (ADCIRC) [43] was selected for this study. The model has been used in many coastal, estuaries water elevation and currents analysis in many studies [4, 2, 44]. ADCIRC solves the shallow–water equations on an unstructured, triangular elements discretized in the physical domain. Instead of solving directly 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 [45]. The coupling between the reformulation of continuity equation and the momentum equation represents the governing equations of ADCIRC which solves for the quantities of interest (i.e. water elevation and velocity). This system of equations are discretized spatially by a first–order continuous Galerkin finite element scheme. The time derivatives are approximated using centered finite differences in GWCE and forward differences in the momentum equations.

Variety of physical domain scales and complex bathymetry structures can be incorporated by ADCIRC, including the scales range from the deep ocean through basins and continental shelves to coastal in land [13]. The minimum requirements for ADCIRC are the description of the finite element mesh and tidal forcing parameter [46]. ADCIRC has been implemented and validated intensively in many studies using hindcast studies. The hindcast studies have been conducted on several storms,

including Hurricanes Betsy (1965), Ivan (2004), Dennis (2004), Katrina (2005), Rita (2005) [47, 48, 49], Gustav (2008) [3] and Ike (2008) [5]. The accuracy of the hindcasts depends significantly upon the the accuracy of model inputs and data. In this study we focus on the characterization of the bottom friction and we use ADCIRC input files for uncertainty propagation of the Manning's n coefficient.

2.1 The Shallow Water Equations

The shallow water model is derived from the Navier-Stokes equations assuming the incompressibility and hydrostatic pressure approximation. The SWEs is well-suited to model the flow with large horizontal scales relative to the vertical scales. ADCIRC in particular implements the depth–integration where the water elevation and depth-averaged velocities are solved for each node. The SWEs in ADCIRC framework is given by

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(Q_x) + \frac{\partial}{\partial y}(Q_y) = 0 \quad (2.1)$$

$$\begin{aligned} \frac{\partial Q_x}{\partial t} + \frac{\partial U Q_x}{\partial x} + \frac{\partial V Q_x}{\partial y} - f Q_y &= -gH \frac{\partial[\zeta + P_s/g\rho_0 - \alpha\eta]}{\partial x} \\ &\quad + \frac{\tau_{sx}}{\rho_0} - \frac{\tau_{bx}}{\rho_0} + M_x - D_x - B_x \\ \frac{\partial Q_y}{\partial t} + \frac{\partial U Q_y}{\partial x} + \frac{\partial V Q_y}{\partial y} - f Q_x &= -gH \frac{\partial[\zeta + P_s/g\rho_0 - \alpha\eta]}{\partial y} \\ &\quad + \frac{\tau_{sy}}{\rho_0} - \frac{\tau_{by}}{\rho_0} + M_y - D_y - B_y. \end{aligned} \quad (2.2)$$

Here, ζ is free-surface elevation relative to geoid, h is the bathymetry depth relative to geoid, $H = \zeta + h$ is the water depth, U_{x_i} are the depth-avaraged horizontal velocity components, $Q_{x_i} = U_{x_i}H$ is the flux per unit width in the x_i direction, f

is the Coriolis parameter, 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_i} are the applied free surface stresses, τ_{bx_i} are the bottom friction components, M_{x_i} is the vertically-integrated lateral stress gradient, D_{x_i} is the momentum dispersion, and B_{x_i} is the vertically-integrated baroclinic pressure gradient.

To replace the continuity equation with GWCE, a multiple, $\tau_0 \geq 0$, of the continuity equation is added to the time derivative of the continuity equation. Then by assumming the constant bathymetric depth, $\frac{\partial H}{\partial t} = \frac{\partial \zeta}{\partial t}$. Finally, substituting the momentum equations into this equation completes the derivation:

$$\begin{aligned}
& \frac{\partial^2 \zeta}{\partial t^2} + \tau_0 \frac{\partial \zeta}{\partial t} \\
& + \frac{\partial}{\partial x} \left(-\frac{\partial U Q_x}{\partial x} - \frac{\partial V Q_x}{\partial y} + f Q_y - g H \frac{\partial [\zeta + P_s / g \rho_0 - \alpha \eta]}{\partial x} + \frac{\tau_{sx}}{\rho_0} - \frac{\tau_{bx}}{\rho_0} + M_x - D_x - B_x + \tau_0 Q_x \right) \\
& + \frac{\partial}{\partial y} \left(-\frac{\partial U Q_y}{\partial x} - \frac{\partial V Q_y}{\partial y} + f Q_x - g H \frac{\partial [\zeta + P_s / g \rho_0 - \alpha \eta]}{\partial y} + \frac{\tau_{sy}}{\rho_0} - \frac{\tau_{by}}{\rho_0} + M_y - D_y - B_y + \tau_0 Q_y \right) \\
& - U H \frac{\partial \tau_0}{\partial x} - V H \frac{\partial \tau_0}{\partial y} = 0.
\end{aligned} \tag{2.3}$$

The bottom stress components in the momentum equation are defined through the coefficient $K_{slip} = c_f |\mathbf{u}|$ as stated in a quadratic drag law. Then c_f is determined using Manning's n formulation,

$$c_f = \frac{gn^2}{H^{1/3}} \tag{2.4}$$

where n is the Manning's n coefficient of roughness, which vary spatially depending on the surface characteristics of the sea floor. The Manning's n coefficient values

are obtained from tables of well-defined empirical estimates for land classifications since there is no method to directly measure it. the Manning’s n coefficient can be roughly defined at the nodes of discretized physical domain (e.g. at the seabed) which in turn construct the piecewise linear representation of bottom friction. Since these values are often determined from the data sources that are incomplete or with several missing time periods, the outputs from coastal ocean model is always subjected to large amounts of uncertainty due to the uncertainty in the Manning’s n coefficient (Table 2.1).

Open water	0.020	Mixed forest	0.170
Ice/snow	0.022	Shrub land	0.070
Pasture	0.033	Grassland	0.035
Commercial	0.050	Low residential	0.120
Bare rock/sand	0.040	High residential	0.121
Gravel pit water	0.060	Row crops	0.040
Fallow	0.032	Small grains	0.035
Transitional	0.100	Recreational grass	0.030
Deciduous forest	0.160	Woody wetland	0.140
Evergreen forest	0.180	Herbaceous wetland	0.035

Table 2.1: Values of the Manning’s n coefficient for various surfaces. the left columns in the tables describe the land characteristics and the right columns are the associated empirically defined Manning’s n coefficients.

Multiple formulations of the bottom stress are accommodated by ADCIRC model. A linear, quadratic or the hybrid law (i.e. a constant friction coefficient in deep water and a quadratic friction law elsewhere) can be implemented. A Chezy or Manning’s n formulation of the bottom stress is supported. ADCIRC allow for the point-wise parameters specification on each discretized vortex. This can be assigned via the input file described in the next section. In this work, we adopt the quadratic friction coefficient law using the Manning’s n formulation.

The SEIK filter is implemented to the ADCIRC in matlab in order to run the model and to update the Manning’s n coefficients at each node at every assimilation step. The ensemble is initialized by sampling on the random distribution around the

expected value. Then the ensembles are used to create the multiple instance of the input files. These files then run simultaneously by ADCIRC. When the model run reach the iteration where the data are available, the matlab script is used to read in the model state variables. The discrepancy between the model state and the synthetic data, which considering the truth is then employed to update the parameters expected value with SEIK filter. New ensembles is then created from this update, and the of model run/filtering cycle is repeated until the end of the parameter estimation.

2.2 ADCIRC Features and Capabilities

The depth-integration GWCE mode in ADCIRC is designed to solve for water elevation and velocity. The model represents the domain by using finite elements to generate unstructures triangular finite-element mesh. This is a standard and efficient approach for resolving complex bathymetry, coastline and island boundaries. The size of unstructured elements can vary as desired.

2.2.1 Model Inputs

The ADCIRC code is written in Fortran. The minimum requirement to run the model are two input files. The first is fort.14: this file describes the structure of the can be obtained from several geophysical database. Then forming and refining are required to generate the desired bathymetry mesh file for ADCIRC. After we acquired fort.14, the file remains the same from the beginning through the subsequence simulation until the end. The second mandatory input file is fort.15: this file contains the wide range of adjustable parameters such as the mode of the run (2D or 3D), the type of coordinate used, model timestep and the period of model running etc.

Supplementary input files can be incorporated depending on the problems the user has to solve. The following bullets summarize the main forcing used in ADCIRC:

- Tidal potential due to the gravity pull of both the Earth and the moon. The phase of the tidal constituent being forced can be defined in fort.15
- Meteorological conditions, mostly used for storm surge prediction. Meteorological forcings are represented through fort.22 file.
- Freshwater inflows, contains in fort.20 file.

Of particular importance for our experimental setup is fort.13 file, which contains nodal attribute on each of the element mesh. The Manning's n coefficient of the bottom surface can be customized nodally in this file. Nodal Manning's n coefficient in fort.13 will be constantly changed during the simulation when data assimilation for parameter estimation is applied.

2.2.2 Model Outputs

The type of output data is specified in fort.15. There exists many options for output format. For instance, users can customize the model outputs in fort.61 and fort.62 to obtain the selected points water elevations and velocity respectively. On the other hand, the water elevations and velocities at all points are provided by fort.63 and fort.64, respectively. This thesis focus solely on the surface water elevations output.

2.2.3 Parallel implementation

Serial and parallel versions of the ADCIRC are available in researchers and developer communities. The parallelized version of ADCIRC can run on interconnected processors. This specific feature gives the advantage of decreasing the run times. It is worth emphasizing that parallelized ADCIRC mode is the most appropriate to the problem encountered in our research. The data assimilations scheme we adopt in this research requires a large amount of parameter sample and therefore model runs. the parallel code may significantly reduces the amount of running time for each sample.

2.3 Experimental setups

In this chapter, we implement the SEIK to estimate the Manning's n coefficients defined in the bottom stress terms within the shallow water equations of the ADCIRC model. We conduct observation simulation system experiments (OSSEs), where the synthetic water elevation data generated by ADCIRC in the moderate conditions are assimilated into a perturbed model, in order to recover the true parameters defining various values of Manning's n coefficient.

In this thesis, we adopt the same experiment setup as in [4], i.e. with the similar domains and forcing. we estimate both constant fields and parametrically defined 2-D field of Manning's n coefficients. We are interested in estimating the parameters in two different domain configurations, the first domain is the idealized inlet with an ebb shoal, and the more realistic domain representation of the Galveston Bay as the second domain. Here only the parameters are estimated with the SEIK, while the state is assumed known.

2.3.1 Observation Simulation System Experiments

Observation simulation system experiments (OSSEs) is conducted in this study. The synthetic data obtained from running ADCIRC model with specific initial condition are considered as the true observations. The Manning's n coefficients, which generate these data are considered the truth and we attempt to recover it based on the data and ADCIRC. SEIK uses the synthetic data to update the parameters at each assimilation cycle. We initiate the model simulation with an incorrect estimated field of Manning's n coefficients, then apply the data assimilation scheme for the parameters to converge to the truths in later iterations. Performing the OSSEs enable the effect of the data assimilation to be quantified.

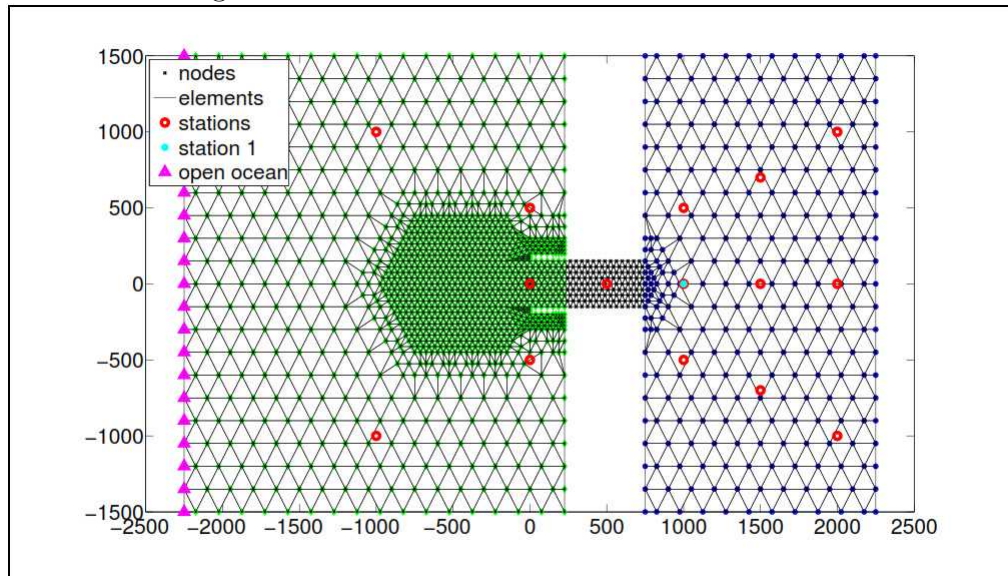
2.3.2 Computational Domains

Idealized Inlet with Ebb Shoal

The first computational domain in our OSSEs is the idealized inlet with ebb shoal. 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. An ebb shoal is located right to the west of the inlet. This domain structure is the simplified version of the real ebb shoal, which is a common natural occurrence at coastal inlets. An ebb shoal is formed by the deceleration of the water while exiting the inlet, which in turn withdraw the large amount of sediment from the bay [50].

This computational domain contains 1,518 grid nodes and 2,828 elements, covers the area of 4500 m wide and 3000 m long (Figure 3.1). The bathymetry of the domain is measured downward from the geoid toward the ocean floor. The depth of the ocean floor is increase 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 (Figure 2.1).

Figure 2.1: idealized inlet with ebb shoal domain

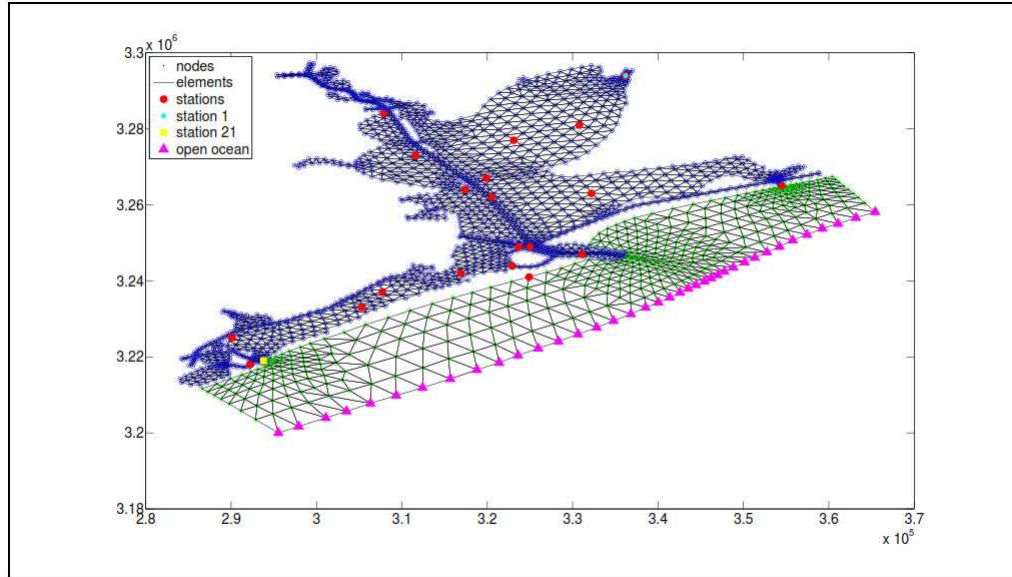


Galveston Bay

A more realistic domain is the coarse representation of Galveston Bay, located on the upper Texas coast in the Gulf of Mexico. Galveston Bay is economically active area, with Port of Houston, one of the busiest port in the world. The location of the Bay make it vulnerable to the effect of hurricane storm surge, which cause a lot of damages to the U.S. economy and the security of its habitats. There is a record of the third costliest hurricane in U.S. history, Hurricane Ike, which destroyed many facilities and households in Galveston Bay area in 2008 [51].

The computational domain representing Galveston Bay contains 2,113 nodes and 3,397 elements. The open ocean boundary located on the southeast side of the domain and the landlocked area on the remaining sides. There are also island boundaries surrounding 17 small islands inside the bay area. The bathymetry of the domain ranges from 0.345 inside the bay to 17.244 m in the Gulf of Mexico (Figure 2.2).

Figure 2.2: Galveston Bay domain



2.3.3 Parameterizing a field of Manning's n Coefficient

Instead of estimating Manning's n coefficient on each and every node on the domains, we first model the field with low-dimensional parameterizations. Then the parameter estimation problem reduces to the estimation of small number of parameters. In this thesis, we primarily focus on the parameter estimation on a field of piecewise constant Manning's n Coefficients. In the first case, we assign the same constant value of Manning's n coefficient at every node of the domain. This field represents an invariant field in space, which is often implemented in most of the coastal ocean models..

The second aim is an attempt to estimate two model parameters, α and β , which parametrically define 2-D field of Manning's n coefficients. This is simply the parameterization of the field with two piecewise constants. We let α be the Manning's n coefficient in the open ocean, and β be the Manning's n coefficient within the inlet and the bay. This field is denoted by $n_{\alpha,\beta}$. This configuration is enforced to both computational domain mentioned before.

2.3.4 Parameter estimation methodology

In general, the parameter estimation problem is about solving for the parameters that produce the model output that best match the observations. However, this problem can often be ill-posed due to the it's complexity, since the different sets of parameters may produce very similar model outputs. The study in [4] shows that a small perturbation of Manning's n coefficient only induce small changes in the model water elevation. So it is sufficient in our OSSE to attempt to recover a value that approximately produces close enough water elevation to the observations.

In order to evaluate the performance in estimating the Manning's n coefficient, we divide the range of Manning's n coefficients into classes based on the water elevation data they produce. Then we generate the synthetic water elevation data from

the middle of each class, and finally try to recover these values from initial guesses initiated at the middle of the other classes.

Classifying Manning’s n coefficients

We divide the Manning’s n coefficients ranging from 0.005 to 0.2 (defined in Table 1.2), incremented value of 0.005, into five classes. These classes are defined by the level of the water elevation data produced by ADCIRC model. It is worth mentioning that the smaller Manning’s n coefficients produce the larger water elevations. ADCIRC is forced using the principal lunar semi-diurnal (M_2) tidal constituent, with the amplitude of 0.25 m for the idealized inlet case and 0.1 m for Galveston Bay relative to the geoid. The simulation is run for five 5 days including 12 hours ramp up period, with a 2 s time step. The largest mean amplitude of the tides at several locations throughout the whole domain is computed, and the Manning’s n coefficients that generate tides with mean amplitudes less than 20%, 40%, 60%, 80%, and 100% of this value are divided in five classes. The more details of this classification framework can be found in [4]

It is worth emphasizing that for both domains, the tides are more sensitive to the variation in smaller Manning’s n coefficients than that of the larger values, as the classes become larger. Thus, the larger classes contain more Mannin’s n coefficients in comparison to the smaller classes. This will affect the recovering of the parameters using SEIK in the larger classes from the smaller classes.

Simulating synthetic water elevation data

The result from the previous section is the classification of Manning’s n coefficients into 5 classes for both idealized inlet and Galveston bay domains. We simulate tides using the same forcing and setups as prescribed in the previous sections, and set the constant Manning’s n coefficient to a value from the middle of each prescribed

classes. For the idealized inlet, the values are 0.015, 0.035, 0.06, 0.105 and 0.17. For Galveston Bay, the constant values we attempt to estimate are 0.01, 0.02, 0.03, 0.06 and 0.14. We generate the data considering as "truth" from these values and trying to recover them from several initial guesses.

Recovering the true class of Manning's n coefficients

For each constant Manning's n coefficients, we first generate the synthetic water elevation data at the predefined location representing the observation stations. There are 15 stations for the inlet case (represented by red circles in Figure 2.1). We run our simulation using the values from four remaining classes as the initial guesses. After the 12 hour ramp up period, the assimilation cycle begins. We assimilate the data every 5 hour over 5 day. We vary the ensemble size for both domain with 10, 20 and 100 ensembles, with $\sigma = 0.01$, where σ^2 is the standard deviation in the initial estimate of the parameter. For the case of Galveston Bay, we use 21 observation stations (Figure 2.2) with the same ensemble sizes and again we let $\sigma = 0.01$.

For the two parameter case, $n_{\alpha,\beta}$, we perform OSSEs in the same manner as described in 1D constant value case. We generate synthetic water elevation data from prescribed setting of Manning's n coefficients and consider these data the truth. Then we try to recover these values from several initial guesses. However in this case, we cannot divided the Manning's n coefficients by class as was done in the 1D case without substantial amount of computation. Thus, we instead define a single true field $n_{0.005,0.1}$, where the coefficient is set to 0.005 in the deep water of the open ocean (highlighted in green in Figure 2.1 and 2.2), and let 0.1 in the shallow bay area (highlighted in blue). The value of the coefficient increases linearly within the inlet for the idealized inlet case (highlighted in purple) and the initial guesses selected are $n_{0.005,0.005}, n_{0.1,0.1}, n_{0.06,0.06}$ and $n_{0.1,0.1005}$. Thus we evaluate the results from incorrect initial guesses of either α , β or α and β . The forcing used in 2D case is similar to

that in the 1D case. We however use a total simulation time of 30 days for Galveston Bay.

Chapter 3

Manning's coefficient estimation with SEIK

In general, there are two directions to reduce the model uncertainties due to uncertainties in the parameters used in the model simulation. The first approach is parameter estimation attempts to reduce the uncertainty of the parameters themselves. The second approach called, state estimation, attempts to compensate for errors in the parameters by adjusting the model input against the observed data.

Traditionally, parameter estimation is done through numerical methods based on optimal control theory where a cost functional penalizing model-data discrepancies is minimized over the estimated parameters. This is called the variational approach. Many parameter estimation methods have been used to estimate the bottom stress parameters in coastal ocean models [52, 53, 54, 55, 56, 57, 58, 59] with varying complexity and success. Most of the successful cases are limited to a few numbers of parameters on very simple domains. More complex cases are still an active area of research although implementing variational methods for these cases is computationally quite expensive. Moreover, variational methods may suffer from instabilities, non-uniqueness, failing to converge to the optimum parameter values [4, 60].

3.1 Kalman filtering applications in shallow water model

In recent years state estimation, particularly statistical data estimation, has been widely applied to reduce the uncertainties in coastal ocean models due to uncertainty in various parameters and inputs.

Budgell [61] was the first to use state estimation in storm surge modeling. He applied the Kalman filter to a linearized 1-D SWEs to reduce the uncertainty in the boundary conditions and the momentum balances. He was able to recover the true state of systems with the accuracy of the numerical approximation of the governing equations. In a subsequent work, Budgell used the extended Kalman filter to estimate the true state of nonlinear SWEs. He found that the filter performance did not degrade even when the model and observation error were non-Gaussian. The filter also improved the estimates of water elevations as well as velocities. Heemink and Kloosterhuis [12, 62] applied the extended Kalman filter to a 2-D shallow water model with a continuity equation assumed to be error free in the momentum equation to predict tides during normal weather and during storm surge. The state estimation problem was solved on a large and realistic domain with a reduced rank approximation of the model error covariance in the Kalman filter. In a more complex domain, Sorensen and Madsen [10] estimated the state variables of a 3-D model of surface flow in the North Sea and Baltic Sea [63]. A four-week hindcast data including a storm surge event were used to test the state estimation, concluding that the methods were suitable for on-line applications. Later [10] assimilated tidal gauge data into a 3-D hydrodynamic model using several variations of the ensemble Kalman filters. They found that the filters provide robust performance in the experimental range of Gaussian noise, the rank of the model error covariance and the standard deviation of the measurement error.

Other examples of successful applications of statistical data assimilation, particularly the variants of Kalman filters to hydrodynamic models, can be found in a few other works [64, 13, 11, 65]. Thus it can be concluded that statistical data assimilation is a promising approach for reducing the uncertainties in applications of shallow water models with a reasonable computational cost. Moreover the methods also produce not only the best estimate of the modeled state, but also an estimate of the corresponding error covariance, which could be used as a measure of uncertainty in the filter solution.

3.2 Kalman filtering and its variants

The Kalman filter was formulated in 1960 by Rudolf Kalman [66], and has been successfully used to assimilate data for linear models. In this method, a model forecast, generally represented as an n -dimensional state vector, \mathbf{x}_k^f at time t_k , is updated by observed data prior to and at the time, t_k , of the forecast. The data at a certain time, \mathbf{y}_k^o , form a p -dimensional vector and the model forecast is updated using a weighted residual between \mathbf{x}_k^f and \mathbf{y}_k^o to form the analyzed state, \mathbf{x}_k^a . The weight for updating, called Kalman gain, is chosen such that the error covariance of the analyzed state is minimized under certain assumptions.

It is customary for all statistical data assimilation methods to begin with an initial estimate of true model state, and a numerical forecast model, \mathbf{M}_k , which predicts the state at one time step ahead, t_k :

$$\mathbf{x}_k^t = \mathbf{M}_k \mathbf{x}_{k-1}^t + \eta_k$$

The model error, η_k , is unknown but we assume that it has Gaussian pdf with zero mean and a given error covariance, \mathbf{Q}_k .

The current model state, \mathbf{x}_k^t , can be projected into the observation space through

an observation operator, \mathbf{H}_k :

$$\mathbf{y}_o^t = \mathbf{H}_k \mathbf{x}_k^t + \epsilon_k$$

Here \mathbf{y}_o^t is the observation of the true state, where the error in the observation operator, ϵ_k is assumed Gaussian with zero-mean and a given error covariance, \mathbf{R}_k . \mathbf{R}_k is often assumed diagonal matrix $\sigma_k^2 \mathbf{I}$, where \mathbf{I} is the identity matrix having dimension of the observation space and σ_k^2 denotes the error variance. These errors are from data points which are assumed to be independent and identically distributed.

The Kalman filter applies both operators in its two main steps, the forecast step and the analysis step. The forecast step computes the system state at one future step ahead from the product of \mathbf{M}_k and the current analyzed state, \mathbf{x}_{k-1}^a :

$$\mathbf{x}_k^f = \mathbf{M}_k \mathbf{x}_{k-1}^a$$

The error covariance of the initial state, \mathbf{P}_{k-1}^a can be then directly computed as follows:

$$\begin{aligned} \mathbf{P}_k^f &= \overline{(\mathbf{x}_k^t - \mathbf{x}_k^f)(\mathbf{x}_k^t - \mathbf{x}_k^f)^T} \\ &= \mathbf{M} \overline{(\mathbf{x}_{k-1}^t - \mathbf{x}_{k-1}^a)(\mathbf{x}_{k-1}^t - \mathbf{x}_{k-1}^a)^T} \mathbf{M}^T + \overline{\eta_k^2} \\ &+ 2\mathbf{M} \overline{(\mathbf{x}_{k-1}^t - \mathbf{x}_{k-1}^a) \eta_k} \\ &= \mathbf{M} \mathbf{P}_{k-1}^a \mathbf{M}^T + \mathbf{Q}_k. \end{aligned} \tag{3.1}$$

Using the assumption of independent model and observation noise, in the analysis step, the observation is used to update the forecast model state and the associated error covariance. Firstly, the forecast state, \mathbf{x}_k^f , is projected into the observation space, providing a direct comparison with the current measured data, \mathbf{y}_k^o . Then the Kalman gain matrix \mathbf{K}_k is used to weight the residual and updating \mathbf{x}_k^f to obtain the

analysis state as

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f). \quad (3.2)$$

Let \mathbf{P}_k^a be the error covariance of \mathbf{x}_k^a which can be computed by expressing \mathbf{x}_k^f , \mathbf{x}_k^a , and \mathbf{y}_k^o in terms of their unknown errors, as follows.

$$\begin{aligned} \mathbf{x}_k^f &= \mathbf{x}_k^t + \epsilon_k^f \\ \mathbf{y}_k^o &= \mathbf{H}_k \mathbf{x}_k^t + \epsilon_k^y \\ \mathbf{x}_k^a &= \mathbf{x}_k^t + \epsilon_k^a \end{aligned}$$

From these relations, we can define the error in the analyzed state, \mathbf{x}_k^a , by the following expression:

$$\epsilon_k^a = \epsilon_k^f + \mathbf{K}_k(\epsilon_k^y - \mathbf{H}_k \epsilon_k^f).$$

One can show that the optimal gain that minimize the mean of the error variance is given by

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T \mathbf{R}_k]^{-1},$$

the analysis error covariance is then

$$\mathbf{P}_k^a = \overline{\epsilon_k^a \epsilon_k^{aT}} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \quad (3.3)$$

The matrix \mathbf{K}_k is the weight that minimizes the error covariance, \mathbf{P}_k^a , of \mathbf{x}_k^a and is

called the Kalman gain matrix. For a linear perfect model, the equation for the analyzed state (3.2) is equivalent to the result of minimizing the cost function in the variational formulation of the data assimilation problem [67].

There are several assumptions used in the formulation of the Kalman filter, but the most restrictive one is that the model is linear. Most practical problems are however non-linear, making the computation of the forecast error covariance not straightforward. Several methods have been developed to tackle this problem including Extended Kalman Filtering (EKF) [68] and the ensemble Kalman filters (EnKF) [39, 40, 41].

In practice, the extended Kalman filter requires linearization the system equation and may be computational prohibitive when implemented with large dimensional systems [4]. Reduced rank Kalman filters were then formulated to mitigate these problems. With a good selection of the initial error covariance, the computational and storage saving of a low rank approximation can be achieved while maintaining acceptable filter performance. The SEEK is developed based on a similar approach in the assimilation process with a modified analysis step. However, in the case of highly non-linear models, EKF based filters have difficulties in accurately computing the gradient of the model.

Evensen proposed the Ensemble Kalman Filter (EnKF) [69] as a Monte Carlo approximation of the EKF, reducing its computational requirements and avoiding the need for linearization. In addition the EnKF is designed to deal with nonlinear models, allowing the forecast error covariance to be more accurately estimated than it is in the EKF. One of the problem with EnKF is that it requires perturbing the measurements used to create the analysis ensemble of and this may introduce noise in the system [13]. Square root EnKFs have been then developed to avoid this noise. Among the most popular EnKFs we cite are the ETKF, the EAKF and the SEIK. All these filters differ in the way the analysis ensemble is generated after every analysis

step, and are statistically equivalent, SEIK will be discussed more thoroughly in the next section.

3.3 The SEIK filter for state estimation

The singular evolutive interpolated Kalman (SEIK) filter is a variation of EnKF [40]. It is among the class of square-root ensemble Kalman filters [70], which update the mean of the ensemble of forecasted state, $\overline{\mathbf{x}}_k^t$, along with a square root matrix of the respective error covariance [71]. This update of the mean eliminates the addition of Gaussian noise to the observations required in the EnKF. The SEIK filter is generally implemented in three step: (1) the sampling step, (2) the forecast step, and (3) the analysis step. The sampling step uses second-order sampling so that the sample mean and sample covariance computed from the ensemble are exactly the analyzed state and error covariance \mathbf{x}_{k-1}^a and \mathbf{P}_{k-1}^a . The forecast step uses the model to evolve the ensemble members forward in time. The analysis step updates the forecasted mean and covariance.

In the SEIK filter, the initial error covariance is represented using an ensemble of states. However, SEIK has a minimal size of the ensemble because of the reduced rank of this matrix. In this case, the ensemble contained only $r + 1$ members. Moreover, an ensemble formulation is used to represent the matrix $\mathbf{H}_k \mathbf{L}_k$ used in (??), avoiding the linearization of \mathbf{H}_k if it is nonlinear.

The forecast and analysis steps of the SEIK filter differ from those of the SEEK filter only slightly. Firstly, we determine an ensemble of initial state vector, $\mathbf{x}_{i,k-1}^a$, from an initial estimate of the model state, \mathbf{x}_{k-1}^a and the factored form of its corresponding error covariance, $\mathbf{P}_{k-1}^a = \mathbf{L}_{k-1} \mathbf{U}_{k-1} \mathbf{L}_{k-1}^T$. Here \mathbf{L}_{k-1} is $n \times r$, and \mathbf{U}_{k-1} is an $r \times r$ positive definite matrix, where $r \ll n$ is the rank of \mathbf{P}_{k-1}^a . Then we compute a Cholesky factor of \mathbf{U}_{k-1}^{-1} , \mathbf{C}_{k-1} , is computed and multiply it by an $(r + 1) \times r$

random matrix, $\mathbf{\Omega}_{k-1}$ with orthonormal columns and zero column sum. Thus \mathbf{P}_{k-1}^a can be written as

$$\mathbf{P}_{k-1}^a = \mathbf{L}_{k-1}(\mathbf{C}_{k-1}^{-1})^T \mathbf{\Omega}_{k-1}^T \mathbf{\Omega}_{k-1} \mathbf{C}_{k-1}^{-1} \mathbf{L}_{k-1}^T.$$

The introduction of the random matrix $\mathbf{\Omega}_{k-1}$ results in a stochastic sampling of the initial state:

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

Then the nonlinear model is applied to this ensemble of initial state vector, creating an ensemble of forecasted state vectors, $\mathbf{x}_{i,k}^f$. Taking the everage of these, $\overline{\mathbf{x}_{i,k}^f}$, as the forecasted state, and the sample covariance as the forecast error covariance, which can be written as

$$\mathbf{P}_k^f = \mathbf{L}_k \mathbf{U}_{k-1} \mathbf{L}_k^T + \mathbf{Q}_k, \quad (3.4)$$

where

$$\begin{aligned} \mathbf{L}_k &= \begin{bmatrix} \mathbf{x}_{1,k}^f - \overline{\mathbf{x}_{i,k}^f} & \dots & \mathbf{x}_{r+1,k}^f - \overline{\mathbf{x}_{i,k}^f} \end{bmatrix} \mathbf{T} \\ \mathbf{L}_k &= \begin{bmatrix} \mathbf{x}_{1,k}^f & \dots & \mathbf{x}_{r+1,k}^f \end{bmatrix} \mathbf{T}, \end{aligned} \quad (3.5)$$

and

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

Here, \mathbf{T} is an $(r+1) \times r$ full rank matrix with zero column sums and (3.4) is the sample error covariance of the $\mathbf{x}_{i,k}^f$. Just like the SEEK filter, \mathbf{U}_{k-1} is updated through its

inverse in analysis step. However, \mathbf{U}_{k-1}^{-1} has new definition and is no longer recursive:

$$\mathbf{U}_k^{-1} = \mathbf{U}_{k-1}^{-1}(\mathcal{H}_k \mathbf{L}_k)^T \mathbf{R}_k^{-1}(\mathcal{H}_k \mathbf{L}_k).$$

Here we apply the observation operator \mathcal{H}_k to the columns of \mathbf{L}_k in (3.5) to compute $\mathcal{H}_k \mathbf{L}_k$. Then \mathbf{U}_k and $\mathcal{H}_k \mathbf{L}_k$ are used to compute the Kalman gain matrix,

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

and the current data are used to update the forecasted state:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k (\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f). \quad (3.6)$$

Finally, the analysis error covariance can be written as

$$\mathbf{P}_k^a = \mathbf{L}_k \mathbf{U}_k \mathbf{L}_k^T. \quad (3.7)$$

A few works successfully utilized the SEIK filter for state estimation in the studying storm surge forecasting [13, 64, 4]. It was found that, in SEIK filter, an ensemble is used to accurately capture the error statistics of the system without introducing additional observation noise to the system, the size of the ensemble is minimized without losing the filter's effectiveness.

3.4 The SEIK filter for parameter estimation

Anderson [71] suggested that an EnKF could be used for parameter estimation of nonlinear models through joint estimation [72], also commonly referred to as state augmentation. In this method model parameters are considered to be part of the model state and are then appended to the state vector. Thus whenever the state

variables are forecasted and analyzed, they are also updated.

In this process, an initial guess model parameters, \mathbf{w}_{k-1}^a , is appended to the initial state vector, \mathbf{x}_{k-1}^a , to form the joint state-parameter vector, $\tilde{\mathbf{x}}_{k-1}^a$:

$$\tilde{\mathbf{x}}_{k-1}^a = \begin{bmatrix} \mathbf{x}_{k-1}^a \\ \mathbf{w}_{k-1}^a \end{bmatrix}.$$

In the ensemble methods, the initial ensemble of state-parameter vectors can be created by adding the Gaussian noise to $\tilde{\mathbf{x}}_{k-1}^a$. It is important that the chosen variance of the Gaussian noise reflects the uncertainty in the model parameters. A procedure for initializing the parameters is described in [73]

In the process of estimating the true parameter values, it is assumed that the evolution of the models parameters is a stationary process, and so the (state) numerical forecast model, \mathcal{M}_k , can be modified so that the parameters remain constant after the model is applied:

$$\tilde{\mathbf{x}}_k^f = \begin{bmatrix} \mathbf{x}_k^f \\ \mathbf{w}_k^f \end{bmatrix} = \begin{bmatrix} \tilde{\mathcal{M}}_k \mathbf{x}_{k-1}^a \\ \mathbf{w}_{k-1}^a \end{bmatrix} = \tilde{\mathcal{M}}_k \tilde{\mathbf{x}}_{k-1}^a,$$

and so $\tilde{\mathcal{M}} := [\mathcal{M}_k \ \mathbf{I}_k]$. The model parameter are not observed, so that the augmented observation operator can be written as

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

Finally, the initial error covariance matrix is formulated as a cross-covariance matrix, $\tilde{\mathbf{P}}_{k-1}^a$, which represents the relationship between the error in \mathbf{x}_{k-1}^a and \mathbf{w}_{k-1}^a . After this modification, the forecast and analysis steps proceed as in any state estimation

method discussed previously, so one can estimate both model state and unknown parameters, and the associated error statistics.

Joint estimation has a few desirable characteristics. Its implementation is straightforward and the computational cost is modest. In addition, it also provides the information about the uncertainty of parameter estimates. However, in practice we should consider the following issues in its implementation. The most implemented one is when the scales of state variables and parameters differ vastly, the cross-covariance matrix can be extremely ill-conditioned. There is also the issue related to the stationary assumption which may cause the filter to diverge in certain situations. To prevent this problem, the common practice is to use inflation factor. In this approach, the posterior standard deviation is inflated by some factor to maintain the variance in the ensemble members. However, Aksoy *et al.* [73] found that inflation may not be always effective, depending on the characteristics of the parameter being estimated.

Another inherent issue in the joint estimation is that the parameters are only observed through the model state. Thus, in most case the model state variables tend to converge to the true state much more rapidly than the estimated parameters. This means that more assimilation cycles are required to obtain satisfactory estimates of the parameters. Furthermore, the effectiveness of joint estimation depends on the size of the ensembles and observation locations [73]. Regardless of this issues, the joint-state estimation approach has been successfully implemented and used in a wide range of problems [72, 74, 40, 75, 73].

3.5 Numerical results and discussion

In this section, we discuss the results of the OSSEs presented in the previous sections. Since our experiment setup is similar to that conducted in [4], We only briefly discuss the results of the same experiment in this thesis. However, specifically to this work,

we are also concerned with the distribution of the parameter estimation with SEIK filter, while the work in [4] only evaluated the point estimations of the Manning's n coefficient. The pdf of the parameter posterior resulting from the SEIK filter shown in this chapter will be used for the comparison of the results with those of the MCMC method will be discussed in the later chapters.

3.5.1 Constant Manning's n coefficient

For the idealized inlet case, the OSSEs show that we are able to accurately recover the Manning's n coefficient to the correct class in the case where the truth lie in of three smallest classes, A, B, and C, regardless of from which classes the initial guesses were initialized. The final estimate lies in the same class as the truth. In the case where we try to estimate the truth in class D (i.e. constant field equal to 0.105), final estimates belong to class D were obtained from all but the smallest initial guess from class A. Finally, for the estimation of the truth in class E, SEIK is able to recover this value accurately with the initial guess from one class below (i.e. class D). The difficulty in recovering the large value of the "true" Manning's n coefficients from a small initial guess arises from the fact that the ADCIRC model is more sensitive to small variations in the smaller coefficients than that in the larger coefficients. Thus, in estimating the large coefficients from small guesses, resulting smaller model/data discrepancies, and thus smaller filter updates in each iteration. It is then expected to experience more difficulty as the gap between a larger truth and a smaller initial guess increases. The parameter estimation requires more time to converge to the truth. Aside from the estimation of the parameters, [4] also mentions the improvement in the water elevation prediction after the data are assimilated. However, we will not focus on this topic since our focus remains on the study of the pdf of the estimated parameters. The point estimations of Manning's n coefficient for inlet case are shown in Figure 3.1.

In SEIK, the initial ensemble is generated by adding Gaussian noise to the initial state vector. Then the ensemble is updated after every forecast step and analyzed step. By construction, the filter ensemble represents, approximately and based on Gaussian assumption, the pdf of the parameter conditional on the available observation. We can then plot the evolution of the 95% percentile of Manning's n coefficient along with their filter estimates, take as the mean of the ensemble member (Figure 3.2). It is clear from this figure that the uncertainties in the parameters decrease as the assimilation advance.

For Galvenson Bay, SEIK again accurately estimates almost all of the Manning's n coefficients from any initial estimates. Similarly to the inlet case, the only exception occurs in the estimation of the largest class, where the truth is 0.14. In this case, final estimates belonging to this class only obtained from all classes but the two smallest class with the initial guesses 0.1 and 0.2. The underline cause of this phenomena could be explained in the same manner as the inlet case (i.e. the difference in the sensitivity of ADCIRC upon the different in Manning's n coefficient). The point estimations of Manning's n coefficient for inlet case are shown in Figure 3.3.

For Galveston Bay, we again plot the point estimations with their 95% percentile around the estimated parameters (Figure 3.4). The pdf produced by the SEIK filter will be compared with the posterior pdf computed by the MCMC in chapter 4.

3.5.2 A field of piecewise constant Manning's n coefficients

In the second set of OSSES we estimate a 2-D parameterized field of Manning's n coefficients. For the idealized inlet case, we obtain the accurate estimates of the parameter for all initial guesses. For each initial estimate, the initial errors with respect to the truth are reduced by at least 80%. We have found that the estimation of $\alpha = 0.005$ is slightly overestimated and $\beta = 0.1$ is slightly under estimated. The inaccuracy of estimating β is greater than α by nearly 33%. However, in the 2-D

parameterized case, the choice of the initial guess hardly affect the final estimate. We find no specific correlation between errors in initial guess and errors in the final estimate. The point estimation of the 2-D parameterized Manning's n coefficients plots can be seen in Figure 3.5 and their extended version including uncertainties is shown in Figure 3.6.

In estimating the 2-D parameterized Manning's n coefficients field for Galveston Bay, we observe that the variations of β significantly affect the water elevations computed by ADCIRC, while varying the value of α barely cause any noticeable variation in the water elevations, meaning that ADCIRC is not sensitive to variations in the value of Manning's n coefficient in the deep water for this computational domain. Thus, the experiment is designed in such a way that α is fixed to 0.005 we only estimate β . This is sufficient for our purpose of our study.

We begin with the initial guesses $n_{0.005,0.01}$, $n_{0.005,0.05}$, $n_{0.005,0.15}$ and $n_{0.005,0.2}$. We assimilate data from 20 stations located in Galveston Bay. We run the model for 30 days including a 12 hours ramp up period. The simulation time for Galveston Bay has to be longer in comparison to the case of Inlet (only 5 days) since it takes longer for the sequence to converge to the true value.

As final result, the final estimates from all initial guesses converge to 0.1. All initial estimates accurately recover the true β except that from the smallest initial guess 0.01. This is to be expected with the same explanation regarding the sensitivity of the ADCIRC model toward the variations in small Manning's n coefficients. The plots of the point parameters estimation for the Galveston Bay is shown in Figure 3.7 and its extended version including uncertainties is presented in Figure 3.8.

3.6 The effect of increasing ensembles size

In general on EnKF, as the SEIK, benefit from increasing the ensemble size. However, increasing the number of ensemble members means increased computational cost, doubling ensemble would double the model runs, which in the case of ADCIRC may significantly increase the computational load. This section studies the SEIK filter performance with respect to ensemble size with the idealized inlet Ebb Shoal.

For the constant Manning's n coefficient case, we run SEIK with the ensemble sizes of 10, 20 and 100, and compared the parameter estimates in time focusing on both the means and the 95% percentile of the parameter. The posterior distribution at the final assimilation time are compared between all ensemble size. We conducted this experiment in two cases where the initial guess did and did not recovered the truth, using the initial guesses equal to 0.06 (Figure 3.10) and 0.015 (Figure 3.9) respectively.

Using 100 ensemble members, the results show significant improvement in comparison to using only 10 ensemble members, especially for the case where we estimate the Manning's n coefficient 0.17 from the smallest initial guess 0.015. In the 10 ensemble case, the recovering truth after the end of the assimilation is 0.07. When we used 100 members, the estimation increased to 0.14, which is considered to be within the same class as the truth. However, it's clear that a hundred ensemble members is still not sufficient to accurately recover the class E exactly for small initial guesses. As for the case where we start the guess from class C (Figure 3.10), no significant improvement results from increasing the ensemble size as the estimates was accurate enough to begin with. Nevertheless, the recovering of the truth from class E from class C is more accurate with increased ensemble size. one can also observe faster convergence toward the truth with increased ensemble sizes.

In the case of estimating a 2-D parameterized Manning's n field, we compare the results obtained from 10 ensemble and 100 ensemble members for both α (Manning's

n coefficient in the open ocean) and β (Manning's n coefficient in the landlocked area). As expected, we found the improvement in estimating the parameters, especially β when using 100 ensembles, in comparison to using 10 only ensembles, which underestimated the truth for all initial guesses.

3.7 The effect of using inflation factor

The previous section was shown that increasing ensemble size is effective, however large number of members are needed to accurately recovering the parameters, and this means increased computational cost. Another way to improve the robustness of an EnKF is to inflate the forecast error covariance by multiplying by an inflation factor > 1 [76]. The use of an inflation factor is described thoroughly in Pham [39]. Covariance inflation is becoming important assimilation technique in the successful implementation.

We perform the same OSSEs as in the previous section for the idealized inlet Ebb Shoal, with an inflation of 2. Figure 3.13 shows the performance of the SEIK filter in recovering the true Manning's n coefficient from several initial guesses for the 1D constant Manning's n value. Significant improvement was observed and any true value is now accurately recovered from any starting point. This is due to the inflation of the background error covariance at every forecast step, as illustrated by the opaque color background around the estimated means.

Similarly to the previous section, we compared the SEIK implementation with and without inflation for the initial guess equal to 0.015 and 0.06, which represent the classes with the resulting least and most accurate estimates respectively (Figure 3.14 and Figure 3.14). The recovering of a constant Manning's n coefficient is now accurate regardless of the classes we started from, with an additional widening of the distribution around the mean.

The results estimating of 2D parameterized Manning's n field in Figure 3.16 and Figure 3.17 show somewhat similar improvements as for the 1D cases.

3.8 Discussions

In this chapter, we discuss the use of the SEIK for parameter estimation to recover the true value of the Manning's n coefficient. We conducted the OSSEs in two distinct domains. The effort was spent to recover both constant fields and 2D parameterized piecewise constant field. The result shows that in the case where we apply SEIK with small number of ensembles, large true values can hardly be recovered from small initial guesses with short simulation windows, and better estimates required longer simulation time will be needed. Increasing the ensemble size significantly improve the filter solutions, however this also requires much more computation. The most suitable solution to tackle this underestimation issue is to inflate the background forecast error covariance, which in turn speeds up the convergence of the estimation to the truth, with no significant increase in computational resources. In addition to the work of Talea *et al.*, which mainly focus only on the point estimation of Manning's n field, we also look at the uncertainty propagation of this parameter by analyzing at its marginal distribution. The distribution of the parameters estimated by the SEIK are Gaussian, with their statistics can be completely determined by the first to moment statistics. We observed that while increasing the ensemble size significantly improves the estimates, the variance in respect to ensemble size are approximately the same. The use of an inflation factor widen the estimated Manning's n coefficient distributions depends on the size of inflation factor. While the EnKF-based approach for the parameter estimation provide good estimations of the mean, the correctness of its marginal distribution is still needed to be addressed. In the next chapter, the PC-MCMC based parameter estimation will be implemented to compute the full

distribution of the estimated Manning's n coefficient, and then compared with the results from EnKF based estimation approach.

Figure 3.1: point estimation of various constant Manning's n fields parameterized for the idealized inlet with ebb shoal case.

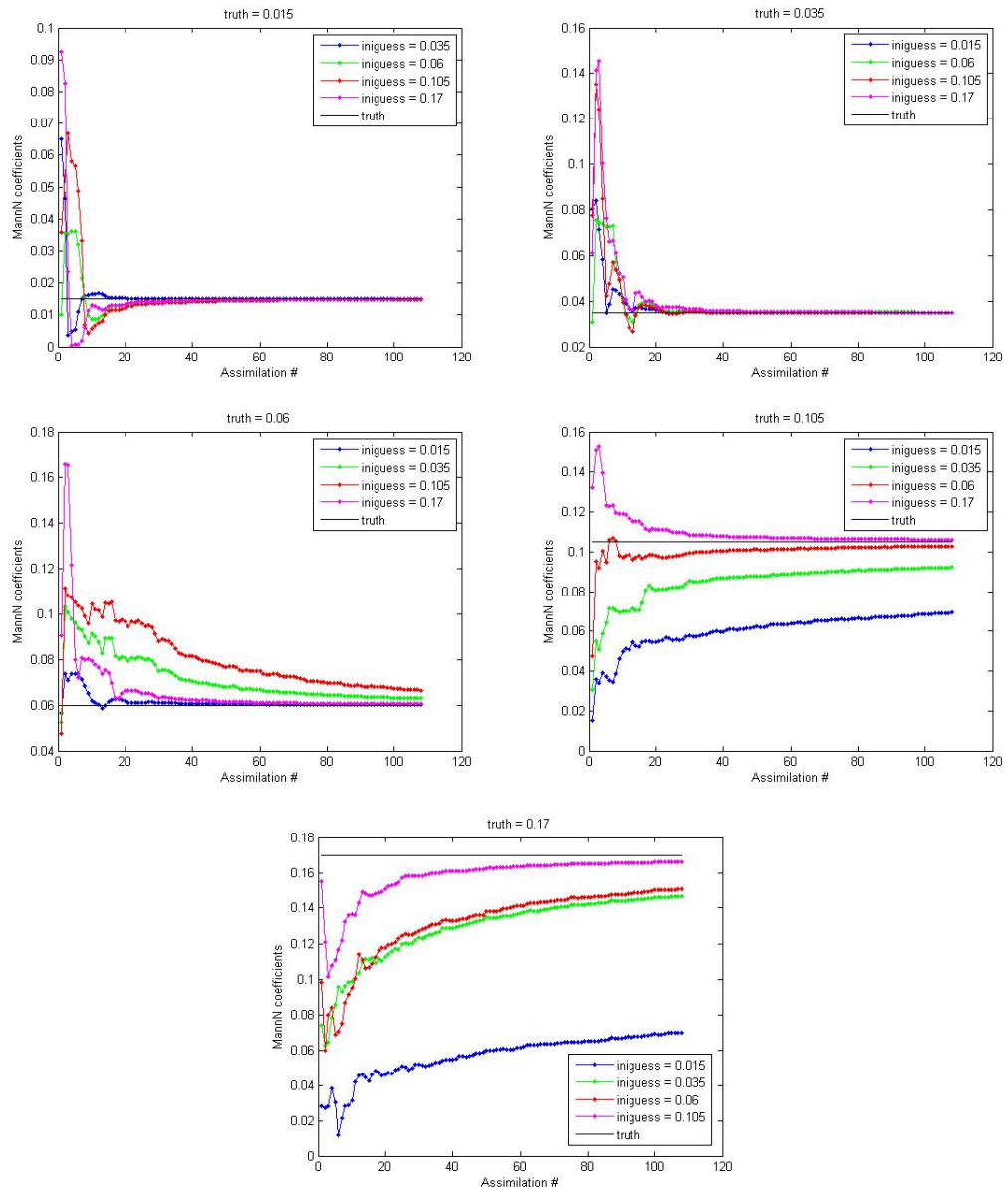


Figure 3.2: Estimates of various Manning's n fields parameterized by a single value for the idealized inlet with ebb shoal case using 10 ensemble members. The width of the pdf from the mean is equal to 3th standard deviation.

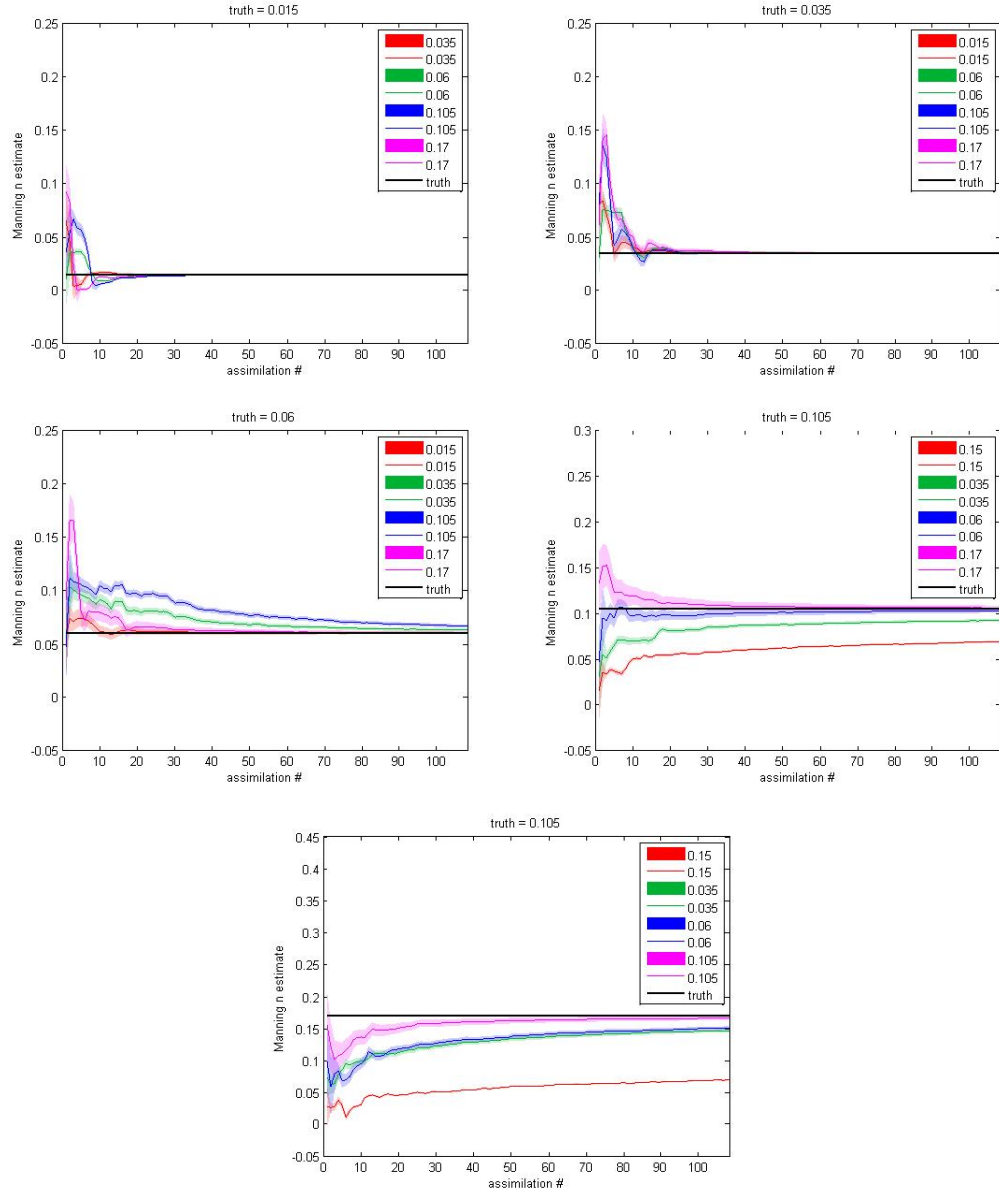


Figure 3.3: point estimation of various constant Manning's n fields parameterized for Galveston Bay.

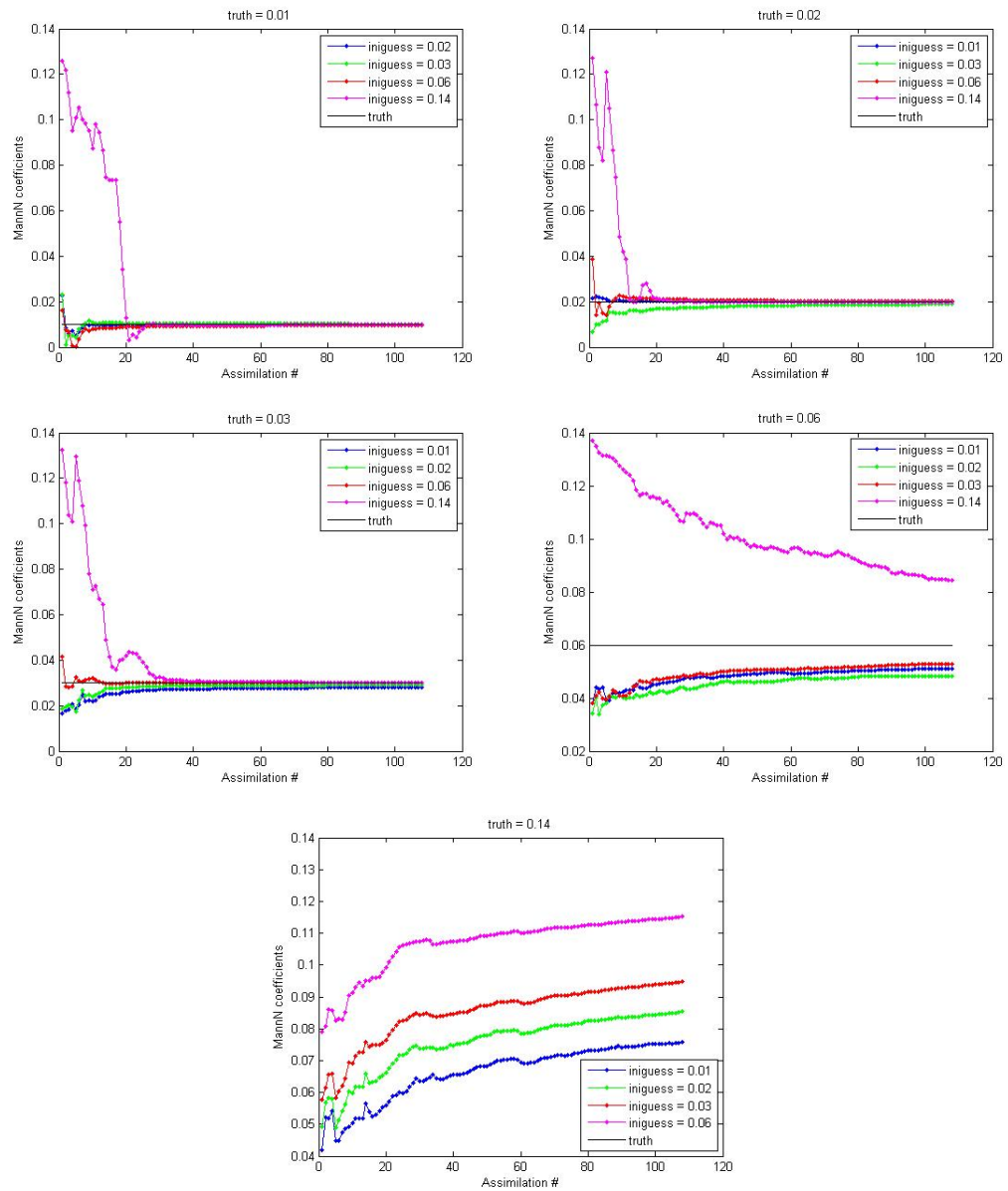


Figure 3.4: Estimates of various Manning's n fields parameterized by a single value for Galveston Bay. The width of the pdf from the mean is equal to 3th standard deviation.

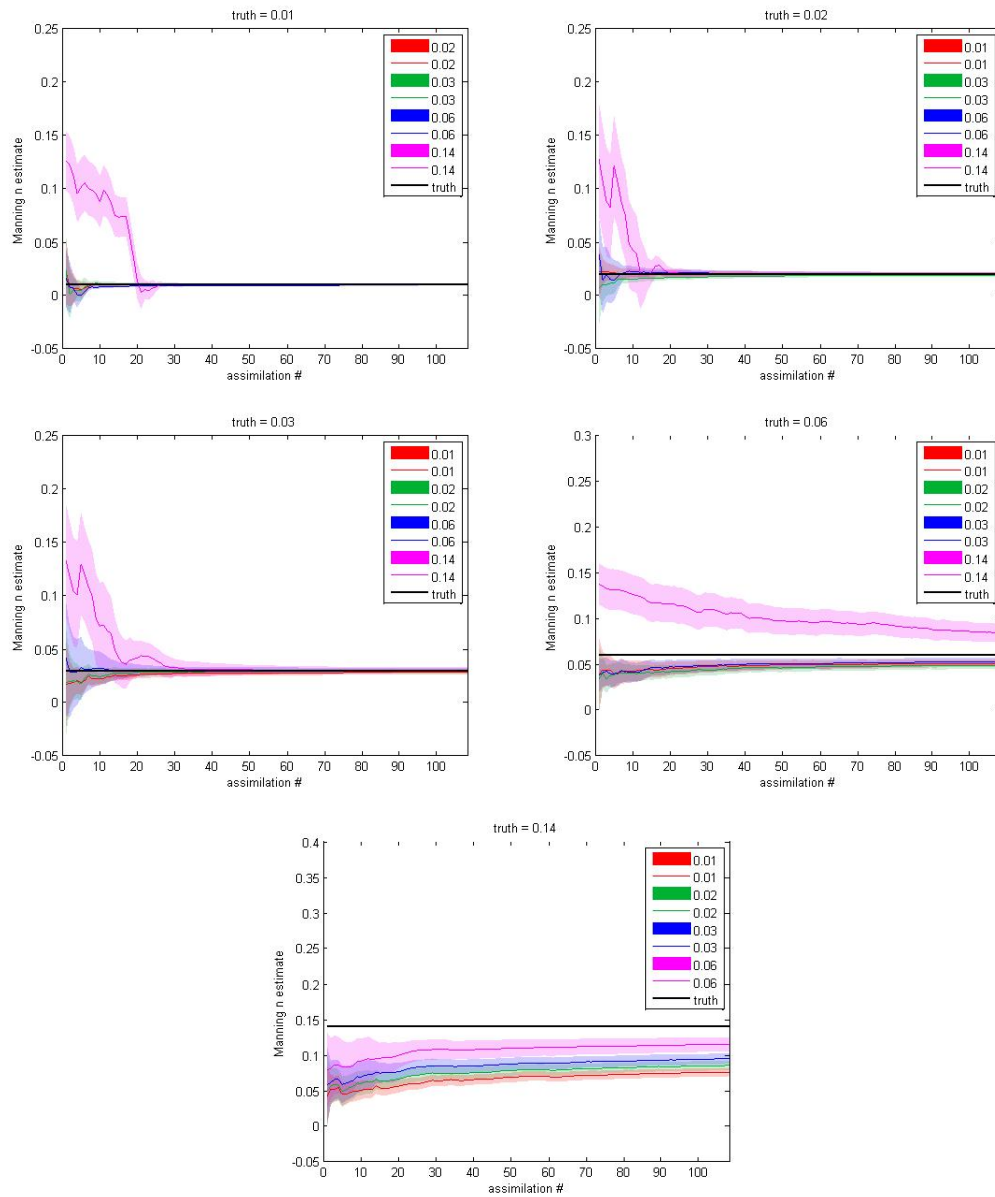


Figure 3.5: point estimates of α (left) and β (right) from various initial guesses for the idealized inlet with ebb shoal case.

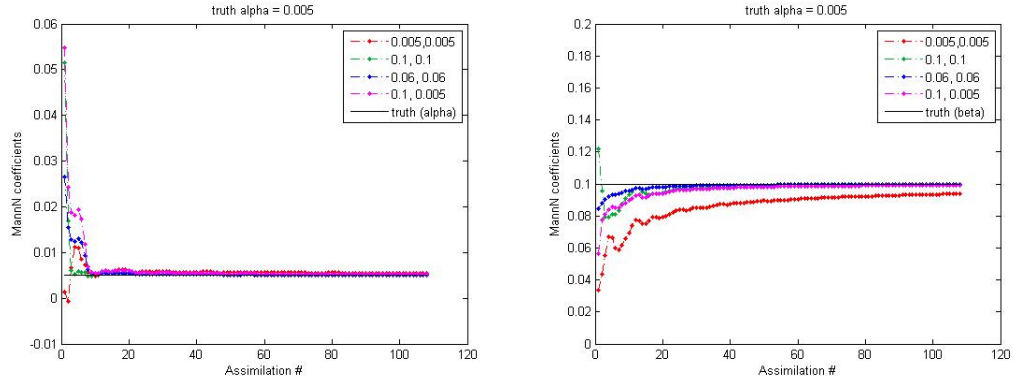


Figure 3.6: pdf estimates of α (left) and β (right) from various initial guesses for the idealized inlet with ebb shoal case using 10 ensembles.

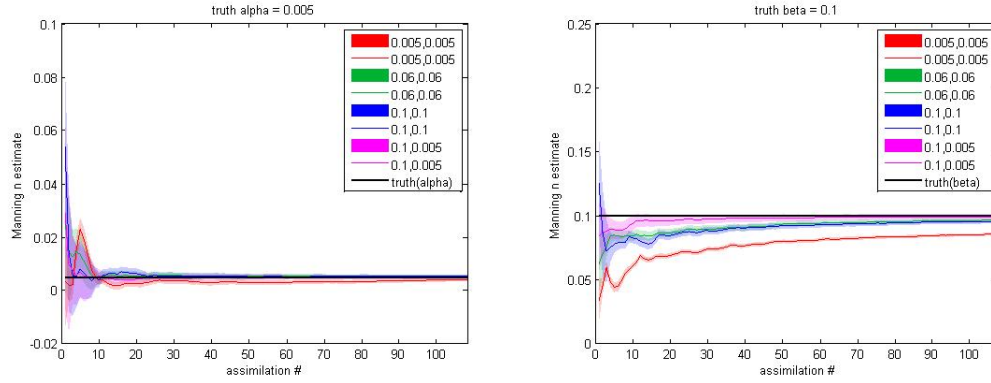


Figure 3.7: point estimates of β from various initial guesses with the same α for Galveston bay case.

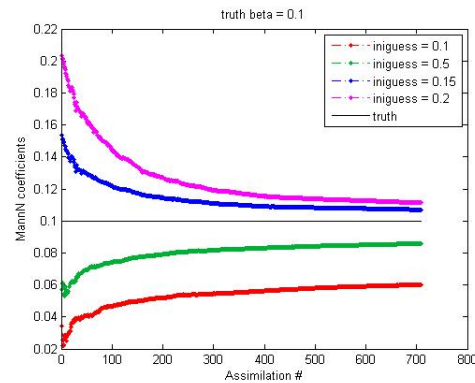


Figure 3.8: point estimates of β from various initial guesses with the same α for Galveston bay case using 10 ensembles.

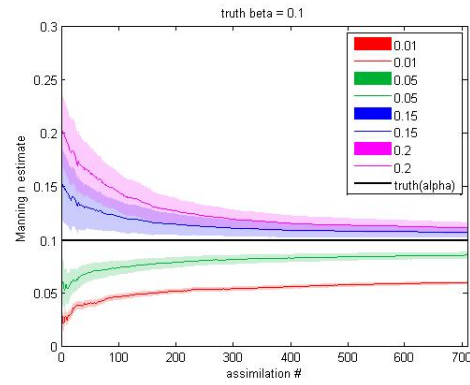


Figure 3.9: Comparison of the posterior distribution resulting from SEIK for constant Manning's n coefficient with three different number of ensembles and initial guess = 0.015. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.

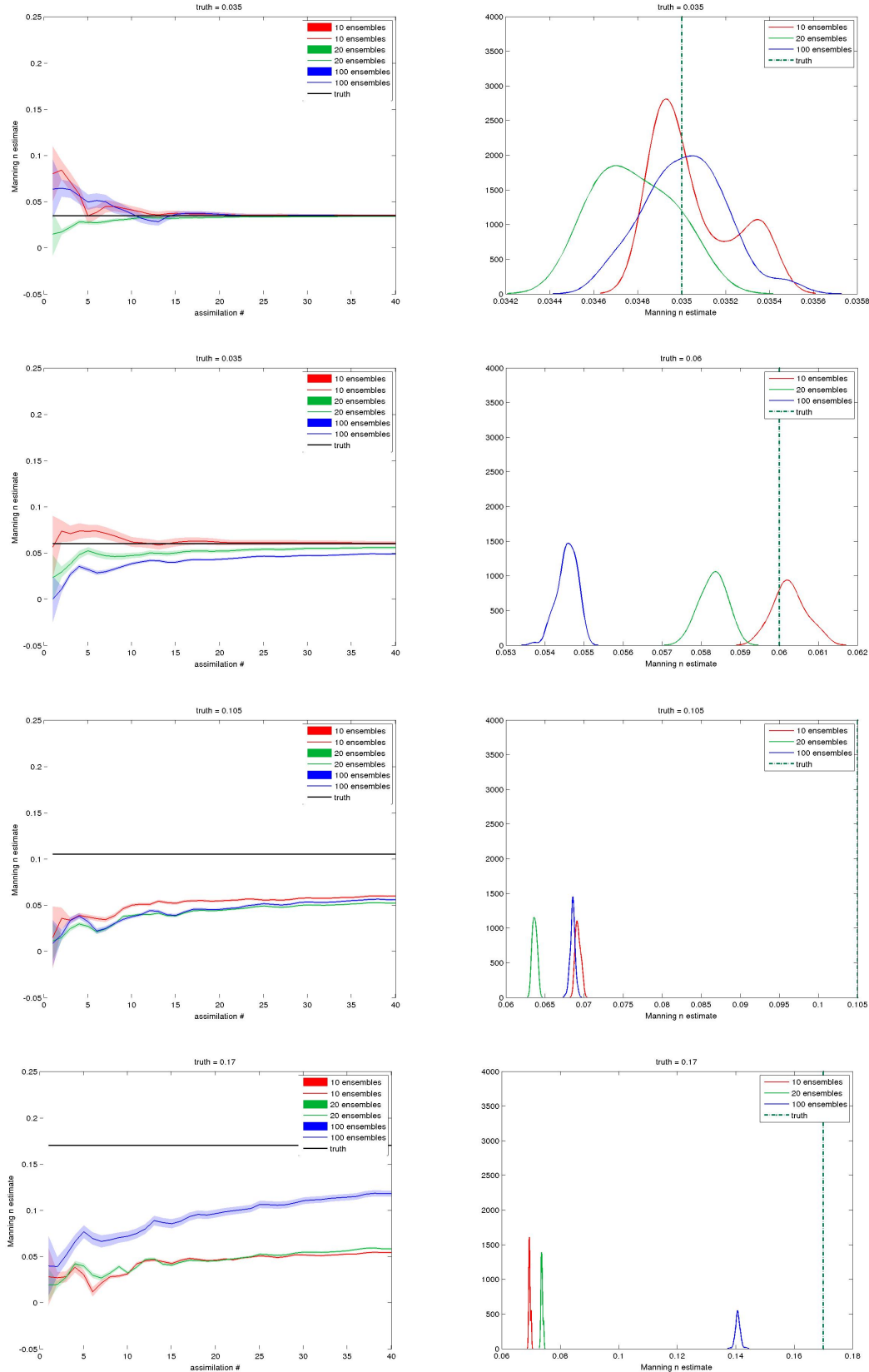


Figure 3.10: Comparison of the posterior distribution resulting from SEIK for constant Manning's n coefficient with three different number of ensembles and initial guess = 0.06. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.

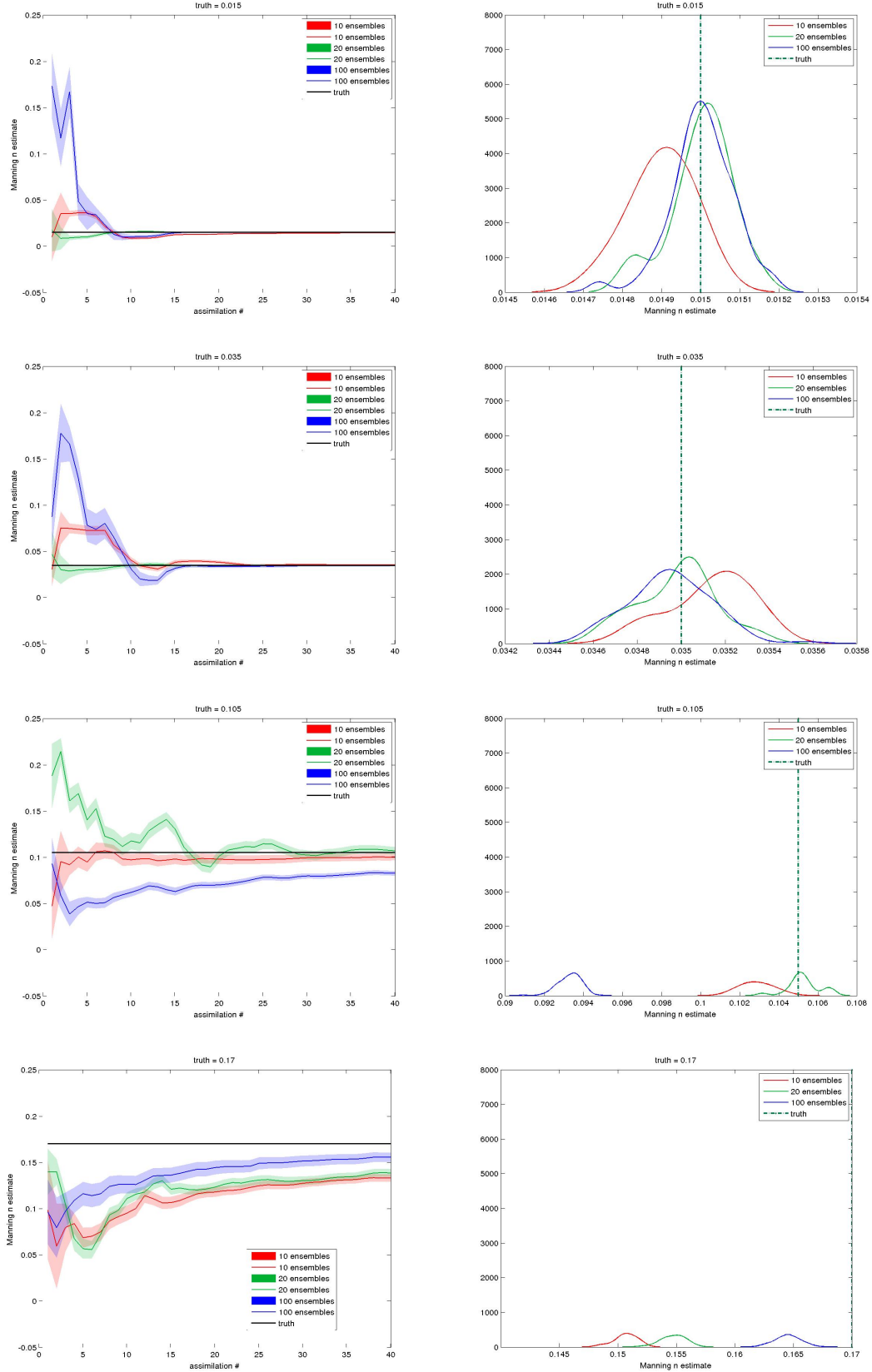


Figure 3.11: Comparison of the posterior distribution result from SEIK for piecewise Manning's n coefficient in the open ocean with different number of ensembles from several initial guesses. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.

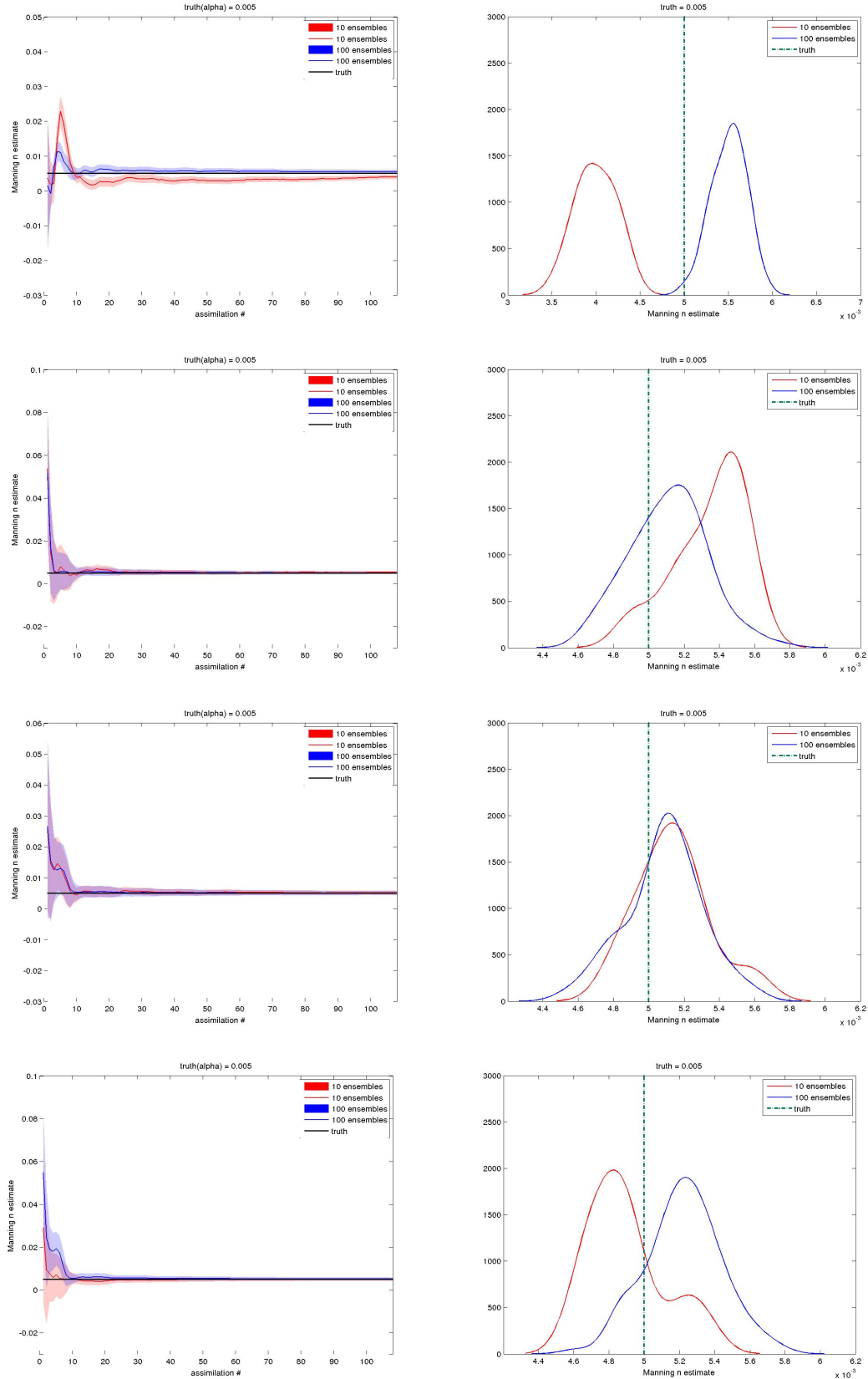


Figure 3.12: Comparison of the posterior distribution result from SEIK for piecewise Manning's n coefficient in the landlocked area with different number of ensembles from several initial guesses. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.

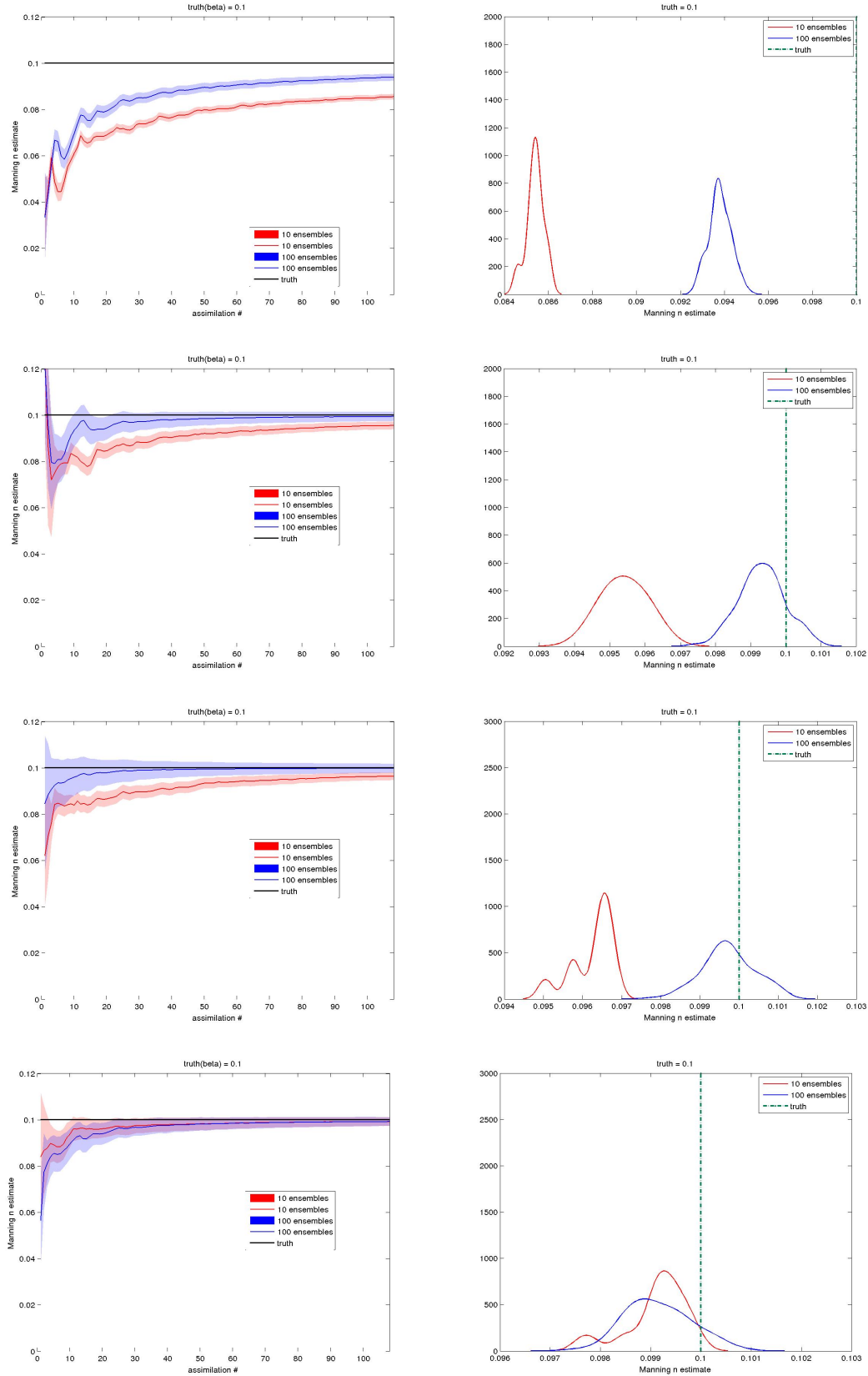


Figure 3.13: Estimates of various true Manning's n fields parameterized by a single value for the idealized inlet with ebb shoal case using 10 ensemble members with inflation factor of 0.5. The width of the pdf from the mean is equal to 3th standard deviation.

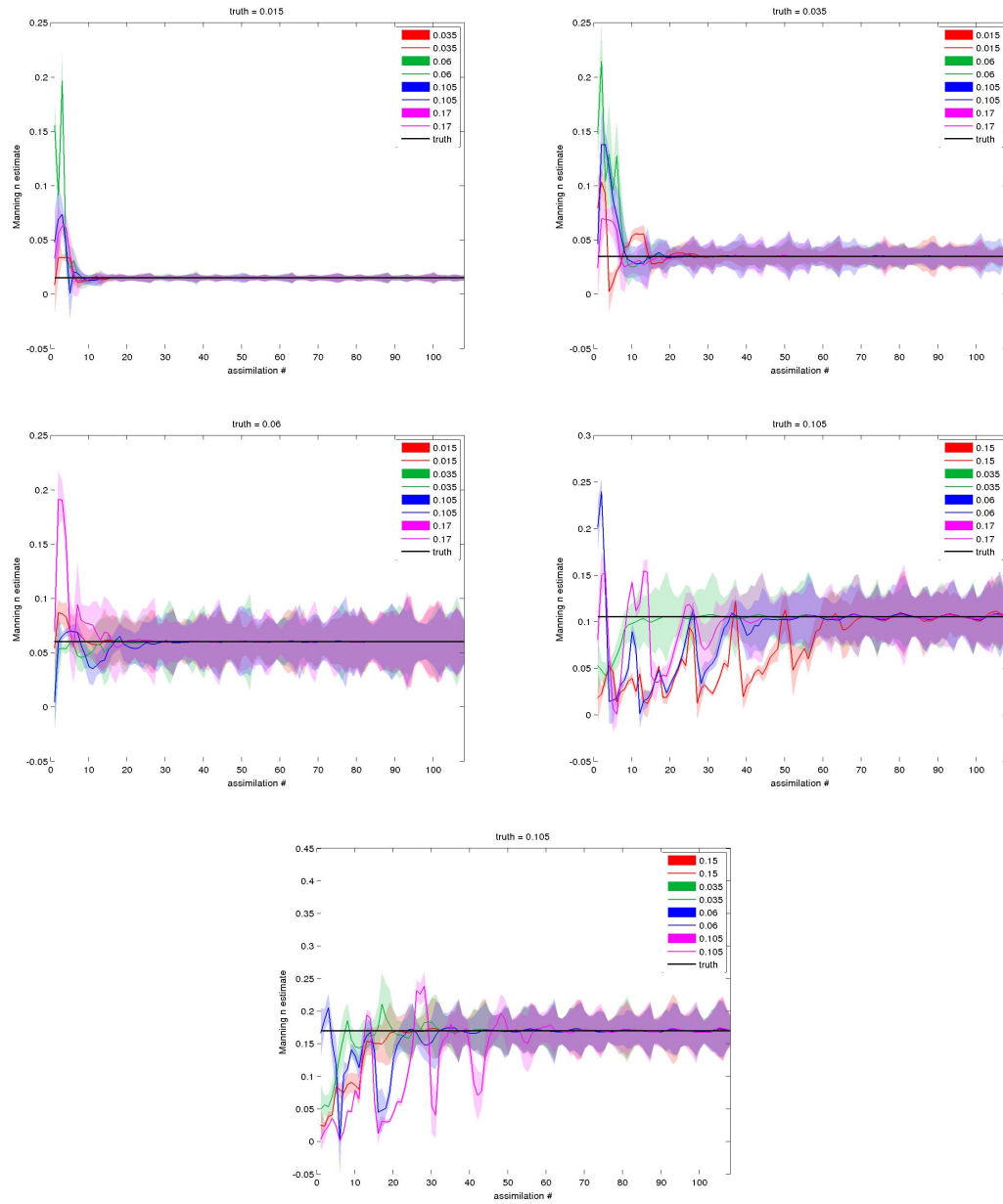


Figure 3.14: Comparison of the posterior distribution result from SEIK for constant Manning's n coefficient between SEIK and SEIK with inflation factor, initial guess = 0.015. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.

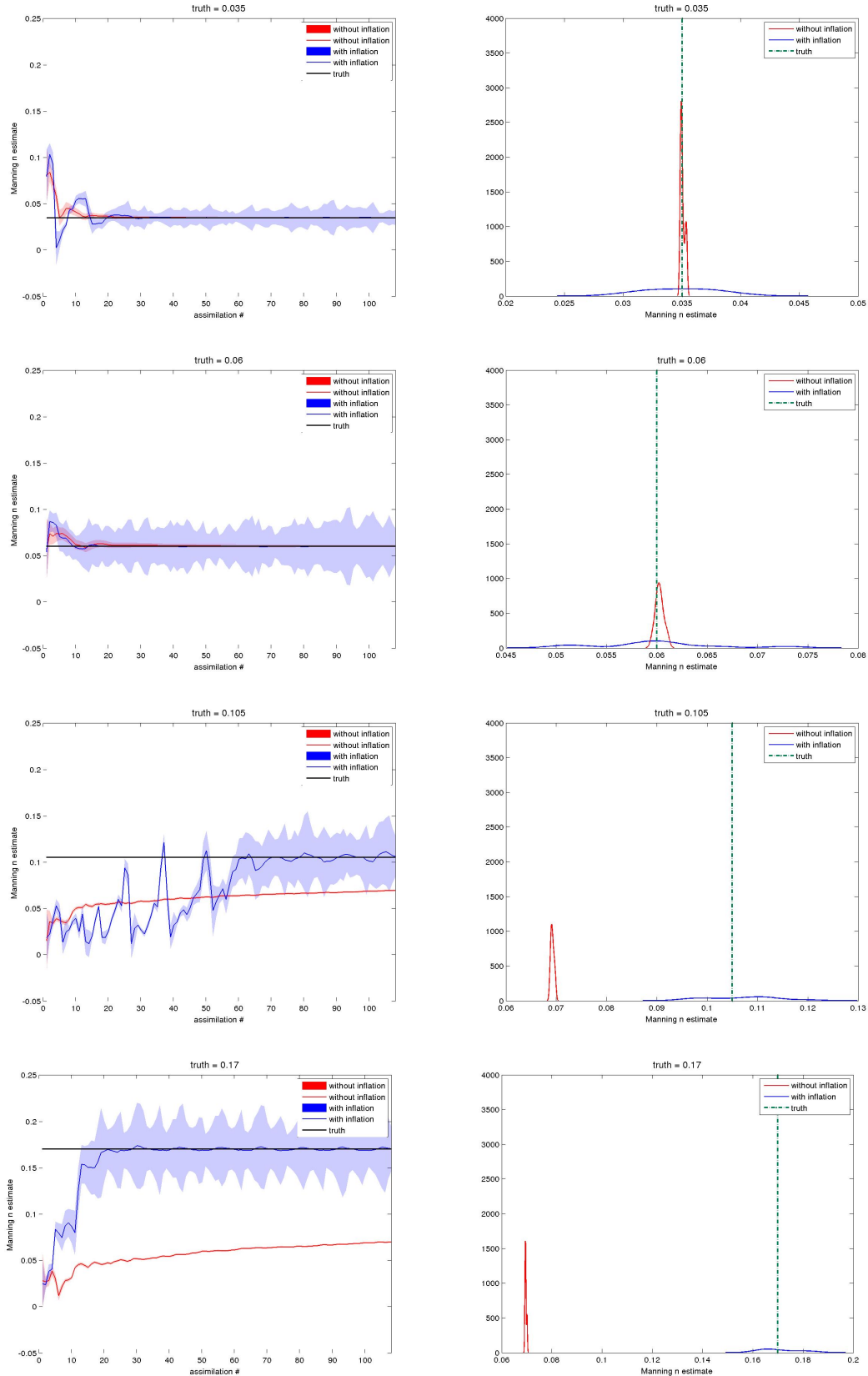


Figure 3.15: Comparison of the posterior distribution result from SEIK for constant Manning's n coefficient between SEIK and SEIK with inflation factor, initial guess = 0.06. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.

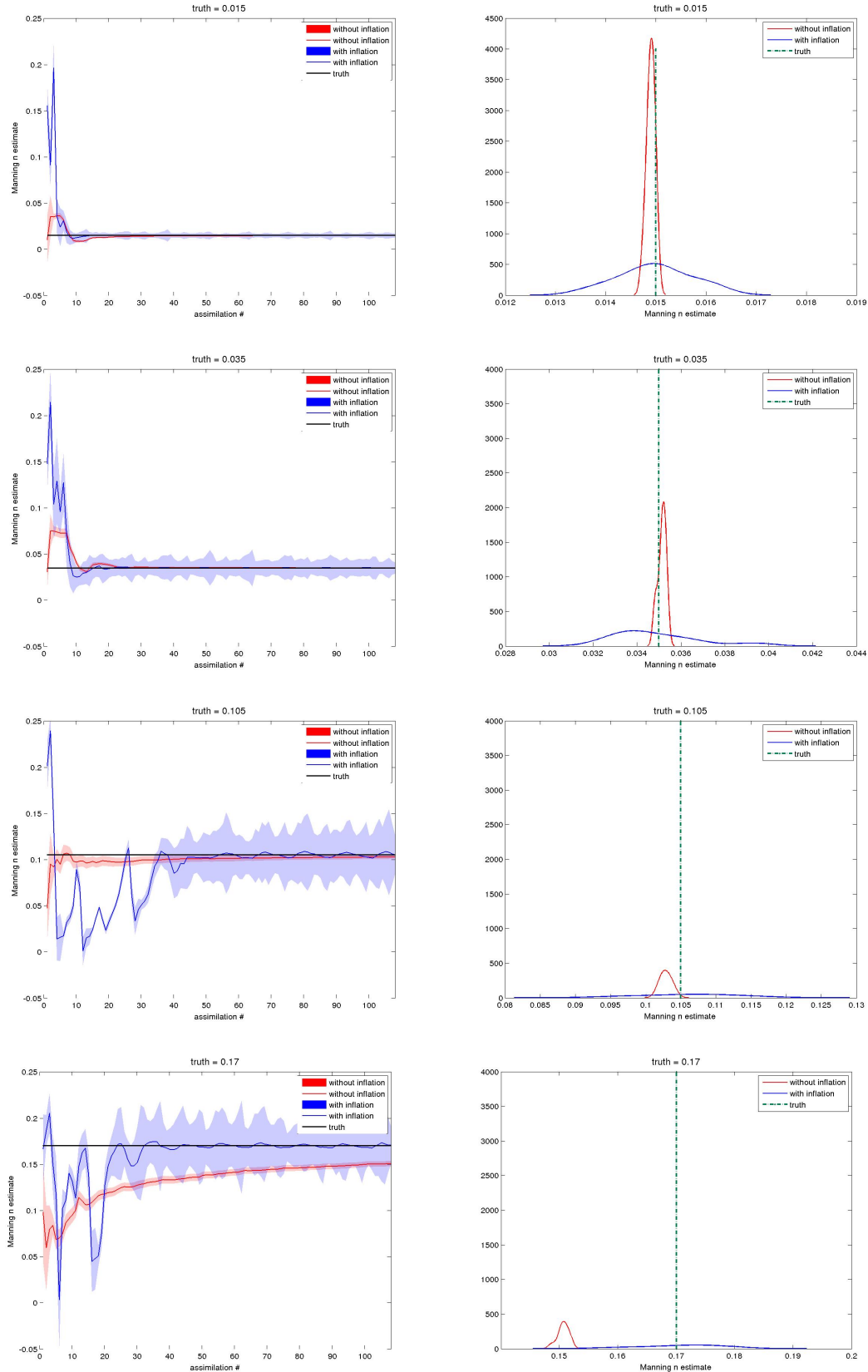


Figure 3.16: Comparison of the posterior distribution result from SEIK for 2D piecewise Manning's n coefficient between SEIK and SEIK with inflation factor in the open ocean of idealized inlet case. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.

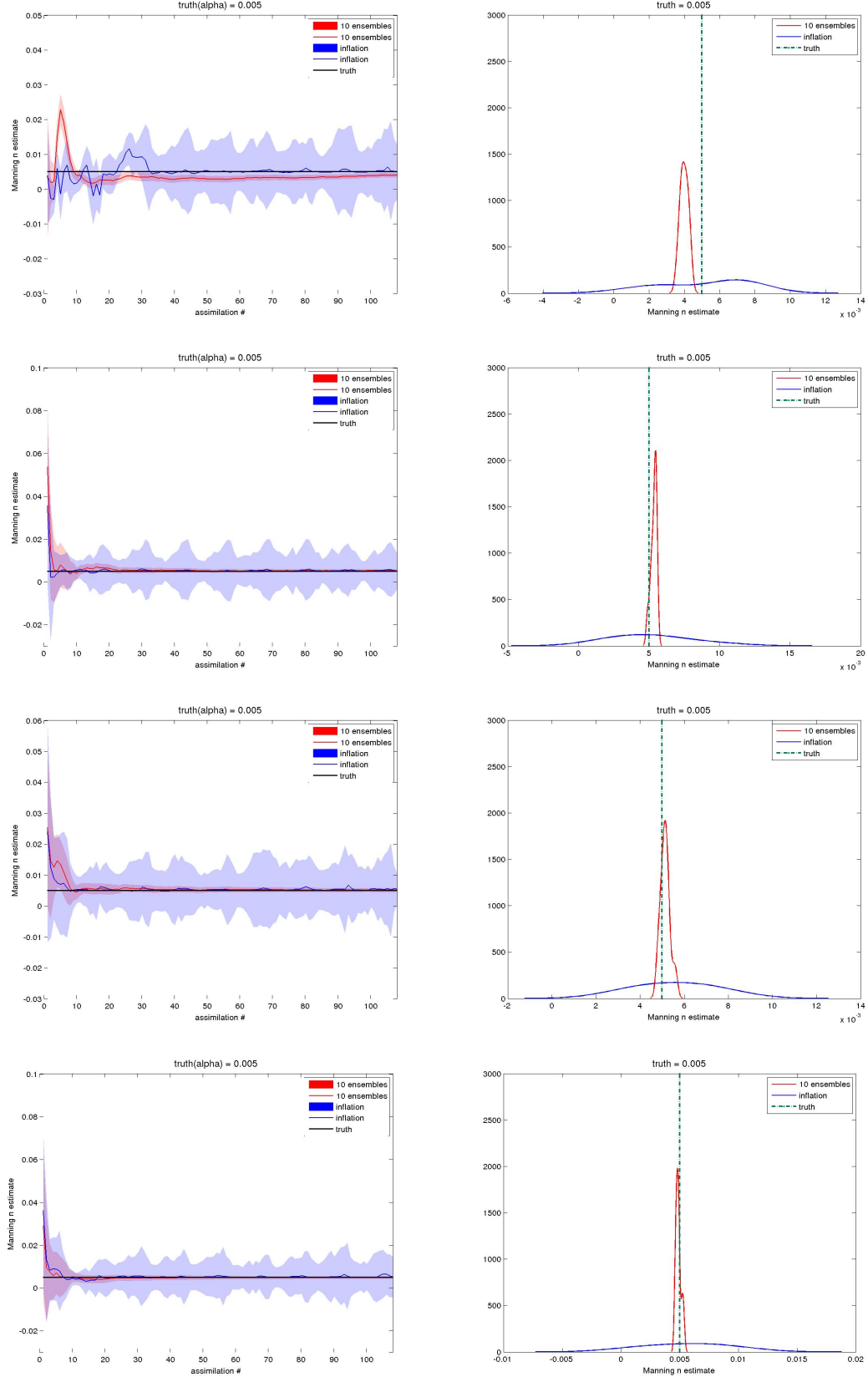
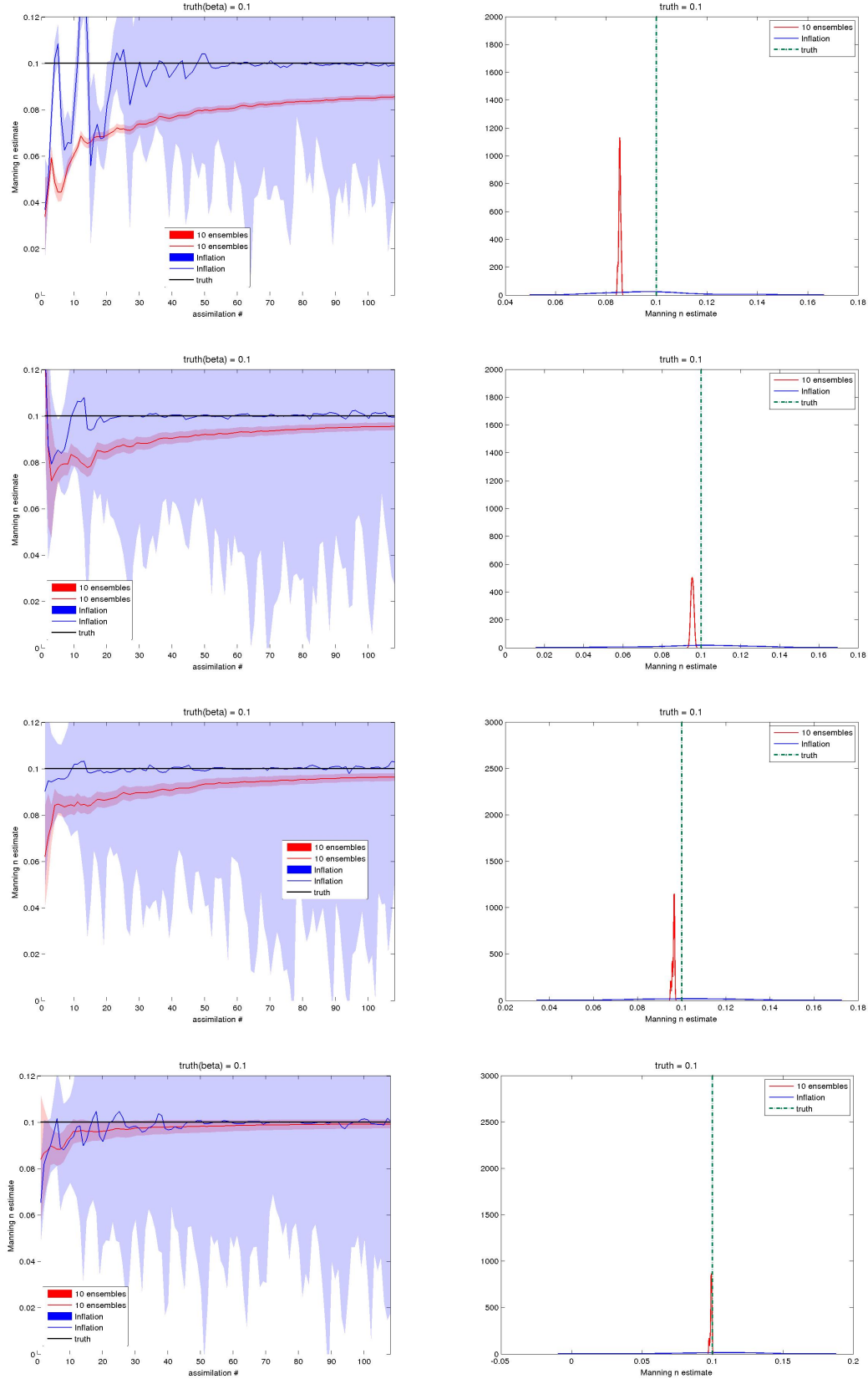


Figure 3.17: Comparison of the posterior distribution result from SEIK for 2D piecewise Manning's n coefficient between SEIK and SEIK with inflation factor in the open ocean of idealized inlet case. (Left column) the 95 percentile over time, (Right column) the pdf of the posterior at the end of the simulations.



Chapter 4

PC-MCMC based for Manning's n coefficient estimation

Chapter 3 presented and discussed the use of a Bayesian filtering approach based on the ensemble Kalman filter, the SEIK filter, for estimating the Manning's n coefficient of the ADCIRC model. In this chapter we use the MCMC method to directly sampling the pdf of Manning's n coefficient from Bayes's rule. To alleviate the huge computational requirements of MCMC we also resort to a PC-based approach to build a surrogate model that allows to sample the likelihood with reasonable computational requirements.

In the following section it will be shown that the EnKF and SEIK filter can be derived from the Bayes' theorem under specific assumptions they rely on. The notion that all data assimilation methods can be formulated from the Bayes' theorem will lead us to the resemblance between filtering problem and Markov Chain Monte Carlo (MCMC) methods for parameter estimation since MCMC also provides the method for sampling the distribution from the posterior provided by Bayes' theorem with less restriction than Kalman filter.

In general circumstances, since the large number of sampling iterations in MCMC is required for the distribution to converge to the posterior, it is too costly and time-consuming to apply MCMC to the sophisticated model such that of the ADCIRC

model since for each model run is already costly. However, since we have reduced the size of the model in the form of the surrogate model presented in the previous chapter (By applying polynomial chaos expansion to the model), the evaluation of MCMC with huge sampling become cheap and applicable.

In fact, KF and its derivatives based on the filtering theory are considered to be well-established, robust and flexible method in data assimilation for real-world applications. Their sequential nature has made them suitable for real-time ocean and atmospheric forecasting. They also capable of updating the huge state vector which is not the case for MCMC if the model evaluated by MCMC is slow. Furthermore, the reduced version of the model using PCs required that the number of parameters we want to estimate must be small. It cannot account for the big vector of the initial condition (i.g. The water elevation at each and every node on the domain) using PCs. In this thesis, it is only reasonable to apply the PCs to the model because the maximum number of the parameters we want to approximate is two, while the forcing, boundary conditions and other initial conditions are remain unchanged.

Since the filtering approaches in data assimilation are more practical and appropriately utilized in storm surge forecasting, however, with the restricted Gaussian statistics assumption, the question that needs to be addressed here is, how well the posterior distribution obtained from SEIK approximates the true posterior distribution?. In this chapter, we represent the posterior distribution of the Manning's n coefficients with the PDF generated by MCMC methods. The comparison of both approaches is the main objective of this chapter and also serve as the conclusion to the whole work devoted in this thesis.

4.1 MCMC theory and Algorithm

In the context of parameter estimation, it is useful to look at the data assimilation problem in the view of Bayesian inference. Especially since MCMC algorithm is constructed on the calculation of the prior and the likelihood function derived from Bayes' theorem, we will start our framework by explaining the Bayes' theorem.

Suppose that we have a set of data $(\{d_i\}_{i=1}^N)$ and we assume a certain model to describe the data. Let \mathbf{H} be the set of parameters (i.e. our hypotheses) defining our model. Bayes' theorem is given by,

$$p(\mathbf{H} | \{d_i\}_{i=1}^N) \propto p(\{d_i\}_{i=1}^N | \mathbf{H})p(\mathbf{H}) \quad (4.1)$$

where,

- $p(\mathbf{H})$ is the prior of \mathbf{H} .
- $p(\{d_i\}_{i=1}^N | \mathbf{H})$ is the likelihood function.
- $p(\mathbf{H} | \{d_i\}_{i=1}^N)$ is the posterior probability.

The likelihood function $p(\{d_i\}_{i=1}^N | \mathbf{H})$ represents the probability of obtaining the data given the hypotheses \mathbf{H} . The prior $p(\mathbf{H})$ represents the information that we have about the unknown parameters before the observations are taken into consideration. The choice of a prior is a key step in the inference process and should be based, when possible, on some a priori knowledge about the parameters. In general, we distinguish between informative (e.g. a Gaussian with known mean and variance), and non-informative priors (e.g. a uniform distribution where we only need the upper and lower bounds). Bayes' theorem describes a process of continuously updating the

current state of knowledge in view of new observation.

Let's now consider our model,

$$\mathbf{h}(\xi, x, t) = \sum_{k=0}^P c_k(x, t) \Psi_k(\xi, x, t) \quad (4.2)$$

where $\mathbf{h}(\xi, x, t)$ represents the solution of shallow water equation is the water elevations as a function of $\xi = (\alpha, \beta)^T$, position on x-y coordinate and time. It follows that our hypothesis is $H = \{\alpha, \beta\}$.

The purpose is to recover the original model given the true data from the observation which means recovering the parameters. In our case, we run the original ADCIRC model with a set of Manning's n coefficients α and β , recording the elevation at the chosen points and times then we consider those data the truth (or the observation vector).

To build the likelihood function we assume the following relationship:

$$d(x, t) = \mathbf{h}(\xi, x, t) + \epsilon \quad (4.3)$$

where ϵ is a random variable which represents the discrepancy between our model, $\mathbf{h}(\xi, x, t)$, and the observations. Assuming N independent realizations and to be normally distributed, the likelihood function can be written as

$$L \equiv p(\{d_i\}_{i=1}^N | \mathbf{H}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_{i,j} - h(x_i, t_j))^2}{2\sigma^2}\right) \quad (4.4)$$

By assuming that ϵ follows a Gaussian distribution with mean zero and variance σ^2 , i.e. $\epsilon \simeq N(0, \sigma^2)$. The objective here is to jointly infer σ^2 , α and β .

The prior of α and β is defined by a uniform distribution,

$$p(\xi_i) = \begin{cases} 1/2 & \text{for } -1 \leq \xi_i \leq 1, \\ 0 & \text{Otherwise ,} \end{cases}$$

We know that σ^2 cannot be negative. This is what we defined a priori knowledge about a parameter. We assume a Jeffreys prior:

$$p(\sigma^2) = \begin{cases} 1/\sigma^2 & \text{for } \sigma^2 > 0 \\ 0 & \text{Otherwise ,} \end{cases}$$

Now the final form of the joint posterior is,

$$p(\{\xi_k\}_{k=0}^p | \{d_i\}_{i=1}^N) \propto [\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{(d_{i,j} - h(x_i, t_j))^2}{2\sigma^2})] p(\sigma^2) \prod_{j=0}^p p(\xi_j) \quad (4.5)$$

The next step after we formulate the posterior is to simulate it. This can be done by using MCMC sampling algorithm.

Markov Chain Monte Carlo (MCMC) methods generates a Markov chain in which at a certain time t , the each state x_t depends only on the previous one x_{t-1} . Suppose the current value of the chain is x_t . We draw a proposal state x from a Gaussian function centered at the current state: $Q(x_t) \sim N(x_t, \beta^2 I)$. Then we Calculate the ratio of the likelihood evaluated for the proposed sample x and the previous sample x_t :

$$r = \frac{L(x')}{L(x_t)} \quad (4.6)$$

The new state x_{t+1} is chosen according to the following rule:

$$x_{t+1} = \begin{cases} x' & \text{if } \alpha < r, \text{ ACCEPTED,} \\ x_t & \text{if Otherwise, REJECTED.} \end{cases}$$

Here $\alpha \sim U(0, 1)$ and β^2 is the variance in proposal drawing step. β^2 need to be well-tuned in order to produce a well-mixed chain. Unfortunately there is no standard procedure to know the best β^2 . This value has to be figured out by trials-and-errors method. In general, the objective is to have an average acceptance ratio between 0.40 and 0.50.

4.2 Polynomial Chaos, Theory and Methodology

Uncertainties in the input variables, parameters and state variables can be modeled as stochastic variables. The system states can be represented by a random vector, $x(t)$ and its time evolution is given by the following differential equation:

$$\dot{\mathbf{x}}(t, \Theta) = \mathbf{f}(t, \Theta, \mathbf{x}, \mathbf{u}), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (4.7)$$

Here, Θ represents uncertain parameters and \mathbf{u} is deterministic forcing terms. The initial state estimate \mathbf{x}_0 may be also uncertain. Let $p(t_k, x_k, \Theta)$ is the probability distribution function (pdf) of the state vector \mathbf{t}_k . Our aim is to compute full pdf, characterizing the space-time evolution of state vector x_k , monitoring uncertainty in

terms of multi-modal distribution, exploiting Bayesian inference to update estimated parameters.

Various techniques have been developed to propagate uncertainties in the inputs and the parameters forward in time [77]. The most popular one being Monte Carlo (MC) [78], Gaussian closure [79], equivalent linearization [80], and stochastic averaging [81, 82]. All of these methods, except the MC methods are suitable only for linear or moderately nonlinear systems. The MC method requires however extensive computational resources and can be prohibitive for high-dimensional dynamic systems.

Recently new uncertainties propagation methods have been developed based on the polynomial chaos theory [33]. Generalized polynomial chaos (gPC) is an extension of the homogeneous chaos idea of Wiener. The idea is to separate the random variables from deterministic one while solving the stochastic differential equation. Then the random variables are expanded using a suitable polynomial expansion. Suitable type of basis polynomial for random variables of different probability distribution as shown in the following table.

Table 4.1: The type of generalized polynomial chaos and their corresponding random variables

	Distribution of Z	gPC basis polynomials	Support
Continuous	Gaussian	Hermite	$(-\infty, \infty)$
	Gamma	Laguerre	$[0, \infty)$
	Beta	Jacobi	$[a, b]$
	Uniform	Legendre	$[a, b]$
Discrete	Poisson	Charlier	$\{0, 1, 2, \dots\}$
	Binomial	Krawtchouk	$0, 1, \dots, N$
	Negative binomial	Meixner	$\{0, 1, 2, \dots\}$
	Hypergeometric	Hahn	$0, 1, \dots, N$

4.2.1 Linear systems

A linear first-order stochastic equation can be written in the following generic form:

$$\dot{\mathbf{x}}(t, \boldsymbol{\Theta}) = \mathbf{A}(\boldsymbol{\Theta}) + \mathbf{B}(\boldsymbol{\Theta})\mathbf{u}(t) \quad (4.8)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$. $\mathbf{u} \in \mathbb{R}^{p \times 1}$ is vector of input signals and $\boldsymbol{\Theta} \in \mathbb{R}^r$ is a vector of uncertain parameters which is a function of the random variable $\boldsymbol{\xi}$ with a known probability distribution $\mathbf{p}(\boldsymbol{\xi})$. We assume that the uncertain state vector $\mathbf{x}(t, \boldsymbol{\Theta})$ and system parameter A_{ij} and B_{ij} can be written as a linear combination of basis function, $\phi_k(\boldsymbol{\xi})$, which span the stochastic space of random variable $\boldsymbol{\xi}$:

$$x_i(t, \boldsymbol{\xi}) = \sum_{k=0}^N x_{i_k}(t) \phi_k(\boldsymbol{\xi}) = x_i^T(t) \boldsymbol{\Phi}(\boldsymbol{\xi}) \quad (4.9)$$

$$A_{ij} = \sum_{k=0}^N a_{ij_k} \phi_k(\boldsymbol{\xi}) = \mathbf{x}_{ij}^T \boldsymbol{\Phi}(\boldsymbol{\xi}) \quad (4.10)$$

$$B_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^N b_{ij_k} \phi_k(\boldsymbol{\xi}) = \mathbf{b}_{ij}^T \boldsymbol{\Phi}(\boldsymbol{\xi}) \quad (4.11)$$

where $\boldsymbol{\Phi}(\cdot) \in \mathbb{R}^N$ is a vector of polynomial basis function orthogonal to $p(\boldsymbol{\xi})$ constructed using the Gram-Schmidt Orthogonalization process. The choice of polynomial basis functions and the corresponding pdf of $p(\boldsymbol{\xi})$ is summarized in 4.1 [36].

The coefficient a_{ij_k} and b_{ij_k} are calculated from the following *normal equations*:

$$a_{ij_k} = \frac{\langle \mathbf{A}_{ij}(\boldsymbol{\Theta}(\boldsymbol{\xi})), \phi_k(\boldsymbol{\xi}) \rangle}{\langle \phi_k(\boldsymbol{\xi}), \phi_k(\boldsymbol{\xi}) \rangle} \quad (4.12)$$

$$b_{ij_k} = \frac{\langle \mathbf{B}_{ij}(\boldsymbol{\Theta}(\boldsymbol{\xi})), \phi_k(\boldsymbol{\xi}) \rangle}{\langle \phi_k(\boldsymbol{\xi}), \phi_k(\boldsymbol{\xi}) \rangle} \quad (4.13)$$

where $\langle u(\boldsymbol{\xi}), v(\boldsymbol{\xi}) \rangle \int_{\mathbb{R}^r} u(\boldsymbol{\xi})v(\boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi}$ is the inner product with $p(\boldsymbol{\xi})$ as a weighing function.

The total number of terms in the gPC expansion (N) depends on the selected highest order of basis polynomial $\phi_k(\boldsymbol{\xi})$, denoted by l . The dimension of the vector of uncertain parameters $\boldsymbol{\Theta}$ is then

$$N = \binom{l+m}{m} = \frac{(l+m)!}{m!l!} \quad (4.14)$$

Now, substituting Eq. 4.9, 4.10 and Eq. 4.11 in Eq. 4.8, we obtain:

$$\begin{aligned} \mathbf{e}_i(\boldsymbol{\xi}) = & \sum_{k=0}^N \dot{x}_{i_k}(t) \phi_k(\boldsymbol{\xi}) - \sum_{j=1}^n \left(\sum_{k=0}^N a_{ij_k} \phi_k(\boldsymbol{\xi}) \right) \left(\sum_{k=0}^N x_{i_k}(t) \phi_k(\boldsymbol{\xi}) \right) \\ & - \sum_{j=1}^p \left(\sum_{k=0}^N b_{ij_k} \phi_k(\boldsymbol{\xi}) \right) u_j, \quad i = 1, 2, \dots, n \end{aligned} \quad (4.15)$$

Eq.(4.15) gives the error of gPC solution of Eq.(4.8) which contains $n(N+1)$ time-varying unknown coefficients $x_{i_k}(t)$. This unknown coefficients can be obtained by projecting the error of Eq. 4.8 onto the space of basis function $\phi_k(\boldsymbol{\xi})$ (Galerkin process).

$$\langle e_i(\mathbf{C}, \boldsymbol{\xi}), \phi_k(\boldsymbol{\xi}) \rangle = 0, \quad i = 1, 2, \dots, n, \quad k = 1, 2, \dots, N \quad (4.16)$$

Thus we obtain the following set of $n(N+1)$ *deterministic differtial equations*:

$$\dot{\mathbf{x}}_{pc}(t) = \mathcal{A} \mathbf{x}_{pc}(t) + \mathcal{B} \mathbf{u}(t) \quad (4.17)$$

Here $\mathbf{x}_{pc}(t) = \{\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, N\}$ is a vector of $n(N+1)$ unknown coefficients, $\mathcal{A} \in \mathbb{R}^{n(N+1) \times n(N+1)}$ and $\mathcal{B} \in \mathbb{R}^{n(N+1) \times p}$. Let P and T_k , for $k = 0, 1, 2, \dots, N$ denote the

matrices of the inner product between the orthogonal polynomials defined as follows:

$$P_{ij} = \langle \phi_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi}) \rangle, \quad i, j = 0, 1, 2, \dots, N \quad (4.18)$$

$$T_{k_{ij}} = \langle \phi_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi}), \phi_k(\boldsymbol{\xi}) \rangle, \quad i, j = 0, 1, 2, \dots, N \quad (4.19)$$

Then we can write \mathcal{A} and \mathcal{B} as an $n(N+1) \times n(N+1)$ block-diagonal matrix, each of them being an $(N+1) \times (N+1)$ matrix. The matrix \mathcal{A} is composed of blocks $\mathcal{A}_{ij} \in \mathbb{R}^{(N+1) \times (N+1)}$ defined as:

$$\mathcal{A}_{ij} = A_{ij}P, \quad i, j = 1, 2, \dots, n \quad (4.20)$$

which can be written as

$$\mathcal{A}_{ij}(k, :) = \mathbf{a}_{ij}^T T_k, \quad i, j = 1, 2, \dots, n \quad (4.21)$$

The matrix \mathcal{B} consists of blocks $\mathcal{B}_{ij} \in \mathbb{R}^{(N+1 \times 1)}$, which is defined as:

$$\mathcal{B}_{ij} = P\mathbf{b}_{ij} \quad i = 1, 2, \dots, n, \quad i = 1, 2, \dots, p \quad (4.22)$$

Eq. (4.9) together with Eq. (4.17) define the uncertain state vector $\mathbf{x}(t, \boldsymbol{\xi})$ as a function of the random variable $\boldsymbol{\xi}$. Following this formulation any order moment of uncertain state variables can then be computed. For example, the first two moments for state vector $\mathbf{x}(t)$ can be computed as follow:

$$\mathcal{E}[x_i(t)] = \sum_{i=1}^n x_{i_1}(t), \quad i = 1, \dots, n \quad (4.23)$$

$$\mathcal{E}[x_i(t)x_j(t)] = \sum_{k=0}^N x_{i_k}(t)x_{j_k}(t), \quad i, j = 1, \dots, n \quad (4.24)$$

4.2.2 Nonlinear systems

To extend the gPC method to propagating the state and parameter uncertainties with a nonlinear system, we start from the following generic equation:

$$\dot{\mathbf{x}}(t, \boldsymbol{\Theta}) = \mathbf{f}(t, \boldsymbol{\Theta}, \mathbf{x}, \mathbf{u}), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (4.25)$$

where $\mathbf{u}(t)$ is the input vector at time t , $\mathbf{x}(t, \boldsymbol{\Theta}) = [x_2(t, \boldsymbol{\Theta}), x_1(t, \boldsymbol{\Theta}), \dots, x_n(t, \boldsymbol{\Theta})]^T \in \mathbb{R}^n$ is the state vector of stochastic system. Here the uncertain parameter vector $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_m]^T \in \mathbb{R}^m$ is assumed to be constant with time and a function of a random vector $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_m]^T \in \mathbb{R}^m$ defined by $p(\boldsymbol{\xi})$ over the support Ω . In general $\mathbf{f}(t, \boldsymbol{\Theta}, \mathbf{x}, \mathbf{u})$ can be a nonlinear function.

In this case the gPC expansion for the state vector \mathbf{x} and the uncertain parameter $\boldsymbol{\Theta}$ is given by:

$$x_i(t, \boldsymbol{\Theta}) = \sum_{k=0}^N x_{i_k}(t) \phi_k(\boldsymbol{\xi}) = \mathbf{x}_i^T(t) \boldsymbol{\Phi}(\boldsymbol{\xi}) \Rightarrow \mathbf{x}(t, \boldsymbol{\xi}) = \mathbf{X}_{pc}(t) \boldsymbol{\Phi}(\boldsymbol{\xi}) \quad (4.26)$$

$$\theta_i(\boldsymbol{\xi}) = \sum_{k=0}^N \theta_{i_k} \phi_k(\boldsymbol{\xi}) = \boldsymbol{\theta}_i^T \boldsymbol{\Phi}(\boldsymbol{\xi}) \Rightarrow \boldsymbol{\Theta}(t, \boldsymbol{\xi}) = \boldsymbol{\Theta}_{pc} \boldsymbol{\Phi}(\boldsymbol{\xi}) \quad (4.27)$$

where, \mathbf{X}_{pc} and $\boldsymbol{\Theta}_{pc}$ are matrices of coefficients of gPC expansion for state \mathbf{x} and parameter $\boldsymbol{\Theta}$ respectively. The coefficients θ_{i_k} are obtained through the *normal equations*, as in the linear case:

$$\theta_{i_k} = \frac{\langle \theta_i(\boldsymbol{\xi}), \phi_k(\boldsymbol{\xi}) \rangle}{\langle \phi_k(\boldsymbol{\xi}), \phi_k(\boldsymbol{\xi}) \rangle} \quad (4.28)$$

Similar to linear case, we substitute Eq. (4.26) and (4.27) into (4.25) and obtain:

$$\mathbf{e}_i(\mathbf{X}_{pc}, \boldsymbol{\xi}) = \sum_{k=0}^N \dot{x}_{i_k}(t) \phi_k(\boldsymbol{\xi}) - \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \boldsymbol{\Phi}(\boldsymbol{\xi}), \boldsymbol{\Theta}_{pc} \boldsymbol{\Phi}(\boldsymbol{\xi}), \mathbf{u}), \quad i = 1, 2, \dots, n \quad (4.29)$$

Applying the Galerkin process to Eq. (4.16) we obtain $n(N + 1)$ time-varying coefficients x_{i_k} . However, for nonlinear stochastic differential equations, Galerkin process will lead to a set of $n(N + 1)$ *nonlinear deterministic differential equations* and difficulties may arise during the computation of projection integrals of Eq. (4.16). This problem can be overcome by using the *Polynomial Chaos Quadrature (PCQ) technique*, which is discussed in the next section.

4.2.3 Polynomial chaos quadrature

Dalbey et al. [83] proposed a technique called Polynomial Chaos Quadrature (PCQ), for the implementation of polynomial chaos integration with non linear model. The idea is to replace the projection step of the gPC by a PCQ. After substituting Eq. (4.26) and Eq. (4.27) into (4.25) one obtains:

$$\sum_{k=0}^N \dot{x}_{i_k}(t) \phi_k(\boldsymbol{\xi}) - \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}), \mathbf{u}) = 0, \quad i = 1, \dots, n \quad (4.30)$$

Upon the projection step of PC yields:

$$\sum_{k=0}^N \langle \phi_k(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi}) \rangle \dot{x}_{i_k} - \langle \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}), \mathbf{u}), \phi_j(\boldsymbol{\xi}) \rangle = 0 \quad (4.31)$$

$$i = 1, \dots, n, \quad j = 0, \dots, N$$

When \mathbf{f} is linear it is possible to evaluate the projection integral in Eq. (4.31) analytically. When \mathbf{f} is nonlinear, the exact integration is replaced by a numerical integration, in this case the Gauss quadrature method. Then one obtains:

$$\langle \phi_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi}) \rangle = \int \phi_i(\boldsymbol{\xi}) \phi_j(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \simeq \sum_{q=1}^M w_q \phi_i(\boldsymbol{\xi}_q) \phi_j(\boldsymbol{\xi}_q) \quad (4.32)$$

$$\begin{aligned}
\langle \phi_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi}) \phi_k(\boldsymbol{\xi}) \rangle &= \int \phi_i(\boldsymbol{\xi}) \phi_j(\boldsymbol{\xi}) \phi_k(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&\simeq \sum_{q=1}^M w_q \phi_i(\boldsymbol{\xi}_q) \phi_j(\boldsymbol{\xi}_q) \phi_k(\boldsymbol{\xi}_q)
\end{aligned} \tag{4.33}$$

$$\begin{aligned}
\langle \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}), \mathbf{u}), \phi_j(\boldsymbol{\xi}) &= \int \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}), \mathbf{u}) \\
\phi_j(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} &\simeq \sum_{q=1}^M w_q \langle \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}_q), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}_q), \mathbf{u}), \phi_j(\boldsymbol{\xi}_q)
\end{aligned} \tag{4.34}$$

where M is the number of quadrature used. Substituting Eq. (4.32), (4.33) and (4.34) into Eq. (4.31) and interchanging summation and differentiation we obtain:

$$\frac{d}{dt} \sum_{q=1}^M \sum_{k=0}^N w_q \phi_j(\boldsymbol{\xi}_q) \phi_k(\boldsymbol{\xi}_q) x_{i_k} - \sum_{q=1}^M w_q \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}_q), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}_q), \mathbf{u}), \phi_j(\boldsymbol{\xi}_q) = 0, \tag{4.35}$$

which becomes after simplification:

$$\frac{d}{dt} \sum_{q=1}^M \phi_j(\boldsymbol{\xi}_q) x_i(t, \boldsymbol{\xi}_q) w_q - \sum_{q=1}^M w_q \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}_q), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}_q), \mathbf{u}), \phi_j(\boldsymbol{\xi}_q) = 0. \tag{4.36}$$

Performing integration with respect to time t yields:

$$\begin{aligned}
&\sum_{q=1}^M (x_i(t, \boldsymbol{\xi}_q) - x_i(t_0, \boldsymbol{\xi}_q)) \phi_j(\boldsymbol{\xi}_q) w_q \\
&- \int_{t_0}^t \sum_{q=1}^M w_q \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}_q), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}_q), \mathbf{u}), \phi_j(\boldsymbol{\xi}_q) dt = 0.
\end{aligned} \tag{4.37}$$

Interchanging the order of integration and summation leads to the following equation.

$$\sum_{q=1}^M \left\{ (x_i(t, \boldsymbol{\xi}_q) - x_i(t_0, \boldsymbol{\xi}_q)) - \int_{t_0}^t \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}_q), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}_q), \mathbf{u}), \phi_j(\boldsymbol{\xi}_q) dt \right\} \phi_j(\boldsymbol{\xi}_q) w_q = 0$$

$$i = 1, \dots, n \quad (4.38)$$

The integral term in Eq. (4.38) can be evaluated by integrating the model equation with a specific instance of the random variable $\boldsymbol{\xi}$. In this way the statistics of the system responses can be computed by sampling the random input with a specific pdf. Finally, the coefficients of the gPC representation can be computed as follows:

$$x_{i_k} = 1/d_k^2 \sum_{q=1}^M \mathcal{X}_i(t_0, t, \boldsymbol{\xi}_q, \mathbf{u}) \phi_k(\boldsymbol{\xi}_q) w_q, \quad k, j = 0, 1, \dots, n \quad (4.39)$$

where

$$\mathcal{X}_i(i_t, \boldsymbol{\xi}_q, \mathbf{u}) = x_i(t_0, \boldsymbol{\xi}_q) + \int_{t_0}^t \mathbf{f}_i(t, \mathbf{X}_{pc}(t) \Phi(\boldsymbol{\xi}_q), \boldsymbol{\Theta}_{pc} \Phi(\boldsymbol{\xi}_q), \mathbf{u}) \quad (4.40)$$

$$d_k^2 = \int_{\Omega} \phi_k(\boldsymbol{\xi}) \phi_k(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \quad (4.41)$$

The moments of the system state and parameters can be estimated by the following relation:

$$\begin{aligned} \mathcal{E} [x_i(t)^N] &= \int_{\Omega} \left(\int_{t_0}^t \dot{x}_i dt \right) dp(\boldsymbol{\xi}) \\ &= \int_{\Omega} \left(x_i(t_0, \mathbf{x} \mathbf{i}) + \int_{t_0}^t \mathbf{f}_i(t, \mathbf{x}, \boldsymbol{\Theta}, \mathbf{u}) dt \right)^N dp(\boldsymbol{\xi}). \quad i = 1, \dots, n \end{aligned} \quad (4.42)$$

For a fixed $\boldsymbol{\Theta} = \boldsymbol{\Theta}_q$, the time integration is simply deterministic and can be computed with standard numerical methods. Applying PCQ over the certain inputs to estimate

the state pdf yields the following moment relation.

$$\mathcal{E} [x_i(t)^N] = \sum_q w_q [\mathcal{X}_i(t_0, t, \boldsymbol{\xi}_q, \mathbf{u})]^N \quad i = 1, 2, \dots, n \quad (4.43)$$

Thus the moments of state variables are approximated as a weighted sum of the outputs of simulation run at the quadrature points of the uncertain input parameters. In the standard Gaussian quadrature applied in this work, we have to exactly integrate polynomials up to degree $2N + 1$ with $N + 1$ quadrature points. Consequently the number of quadrature points increase exponentially as the number of input parameters increases. To mitigate this error of PCQ under sampling, an adaptive or nested quadrature scheme which successively refines the accuracy of the solutions by increasing the number of sample points, such as Clenshaw-Curtis quadrature method [84], [85] was developed. Some examples of gPC formulation and simulation results for linear and nonlinear systems are given in [36] and [77].

4.2.4 Non-intrusive spectral projection (NISP)

In the process of formulating gPC, one can compute the coefficients of PC expansion following two main approaches: (1) intrusive methods, and (2) the non-intrusive methods. The most straightforward way to build a gPC model is to reformulate the original stochastic model through Galerkin projection onto PC basis and then solve for the time/space evolution of PC coefficients [23], [24]. This is exactly what we call, an intrusive approach. This reformulation may however be difficult to code and requires considerable skills for rather complex models like ADCIRC. In the non-intrusive methods, the requisite spectral representation is computed through direct application of existing deterministic solvers. This approach reduces the complexity of gPC formulation while still providing satisfied gPC representation of the original model as has been shown in many application [42, 34, 16].

In the non-intrusive approach we may use different methods to compute the PC expansion coefficients including non-intrusive spectral projection (NISP), the collocation methods (CM), and the regression-like methods. In the regression-like methods [86], the PC coefficients can be estimated by minimizing the distance between the PC expansion and experimental observations/simulations. Using interpolation, the CM methods computes the PC coefficients by using the PC basis as a set of interpolants [87, 88, 28, 89]. The NISP method, which we chose to apply in this work, computes the PC coefficients through L^2 -projection of variables onto PC basis.

However NISP is not without drawbacks but in general all non-intrusive methods suffer from the so-called curse of dimensionality. In the NISP method, a rapid increase in the number of deterministic realizations is required to compute the PC coefficient as the order of expansion and number of stochastic dimensions increase. One way to mitigate this drawback of NISP is to use sparse quadrature techniques where the quality of the PC representation can be monitored and the sparse grid is successively refined to meet the accuracy required.

The following discussion summarize the use of Non-intrusive spectral projection in the PC method. We denote a probability space by $(\Omega, \mathcal{F}, \mu)$, where Ω is the sample space, \mathcal{F} is an appropriate σ -algebra on Ω , and μ is a probability measure. Let ξ is a random variable on Ω having a uniform pdf $\xi \sim \mathcal{U}(a, b)$ on the interval $[a, b]$. The cumulative distribution function of random variable ξ on Ω is given by $F_\xi(x) = \mu(\xi \leq x)$ for $x \in \mathbb{R}$

Let us consider a model with finite uncertain parameters. These parameters are parameterized by a finite collection of real-valued, independently and identical distributed (iid) random variables ξ_1, \dots, ξ_d on Ω where d refers to the number of random variables or the dimension of the stochastic problem. Let F_ξ denote the joint distribution function of the random vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)^T$. Since all events ξ_j are independent, $F_\xi(\mathbf{x}) = \prod_{j=1}^d F(x_j)$ for $\mathbf{x} \in \mathbb{R}^d$, where F is the common distribution

function for ξ_1, \dots, ξ_d .

In computation we normally work in the image probability space $(\Omega^*, \mathcal{B}(\Omega^*), F_\xi)$ for convenience. Here $\Omega^* \subseteq \mathbb{R}^d$ is the image of Ω under ξ , and $\mathcal{B}(\Omega^*)$ is the Borel σ -algebra on Ω^* . Let X and Y denote the two random variables. Thus the expectations are defined for $X, Y : \Omega^* \rightarrow \mathbb{R}$ as:

$$\langle X \rangle = \int_{\Omega^*} X(s) dF_\xi(s). \quad (4.44)$$

The space of square integrable random variable on Ω^* , $L^2(\Omega^*)$:

$$\langle XY \rangle = (X, Y) = \int_{\Omega^*} X(s)Y(s) dF_\xi(s) \quad (4.45)$$

and the norm of X is defined by $\|X\|_{L^2(\Omega^*)} = (X, X)^{1/2} = \langle X^2 \rangle^{1/2}$. If we parameterize these random inputs with canonical random variables $\xi_i \stackrel{iid}{\sim} \mathcal{U}(-1, 1)$ and use the d -variate Legendre polynomials $\{\Psi_k\}_0^\infty$. Each Ψ_k is defined as follows.

$$\Psi_k(\boldsymbol{\xi}) = \prod_{i=1}^d \Psi_{\alpha_i^k}(\xi_i), \quad \boldsymbol{\xi} \in \Omega^* \quad (4.46)$$

where $\boldsymbol{\alpha}^k = (\alpha_1^k, \alpha_2^k, \dots, \alpha_d^k)$ is a multi index vector, with α_i^k being the order of 1D Legendre polynomial, Ψ , in ξ_i . With this basis, any $\mathbf{X} \in L^2(\Omega^*)$ can be written as a gPC as follows.

$$\mathbf{X} = \sum_{k=0}^{\infty} c_k \Psi_k. \quad (4.47)$$

For computation, we truncate Eq. 4.47 to a finite series

$$\mathbf{X}(\boldsymbol{\xi}) \doteq \sum_{k=0}^P c_k \Psi_k(\boldsymbol{\xi}) \quad (4.48)$$

P depends on the stochastic dimension d and expansion order, p , according to the

following relation

$$P = \frac{(d+p)!}{d!p!} - 1. \quad (4.49)$$

Let \mathbf{X} belongs to $L^2(\Omega^*)$. Since $\{\Psi_k\}_0^P$ form an orthogonal system, we can simplify the following dot product.

$$(X, \Psi_k) = \left(\sum_{l=0}^P c_l \Psi_l, \Psi_k \right) = \sum_{l=0}^P c_l (\Psi_l, \Psi_k) = c_k (\Psi_k, \Psi_k), \quad (4.50)$$

and the coefficient c_k is computed from

$$c_k = \frac{\langle X \Psi_k \rangle}{\langle \Psi_k^2 \rangle}. \quad (4.51)$$

Here the moments $\langle \Psi_k^2 \rangle$ of multivariate Legendre polynomials in 4.51 can be evaluated analytically [24] while $\langle X, \Psi_k \rangle$ require more elaborated computation. As we note that

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

Thus evaluating coefficient c_k involves computing the values of a set of $P+1$ integrals over $\Omega^* \subseteq \mathbb{R}^d$ which can be discretized as finite sums using an appropriate quadrature formula which takes the form

$$\int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_{\xi}(\mathbf{s}) \doteq \sum_{j=1}^{N_q} w_j X(\boldsymbol{\xi}_j) \Psi_k(\boldsymbol{\xi}_j), \quad (4.52)$$

where $\boldsymbol{\xi}_j \in \Omega^*$ and w_j are the nodes and weights. As mentioned previously, the complexity of NISP scales with $N_q = n^d$, the number of nodes n and number of uncertain parameters d . Thus this approach is limited to low d . Indeed, as the number of uncertain parameters increase the integral in Eq. (4.52) becomes more complex exponentially, creating the so-called 'the curse of dimensionality'. To this

end one way to mitigate the difficulty is to use Smolyak's formula [90] which leads to sparse quadrature technique [91, 92, 93]. In this approach, the set of integration nodes comprise what we call the NISP sample denoted by

$$\mathcal{S} = \{\boldsymbol{\xi}_j\}_{j=1}^{N_q} \subset \Omega^*. \quad (4.53)$$

Thus, to evaluate Eq. 4.52, one needs to compute $X(\boldsymbol{\xi}_q)$ for all $(\boldsymbol{\xi}_q \in \mathcal{S})$. Let $\boldsymbol{\Pi} \in \mathbb{R}^{(P+1) \times N_q}$ be the matrix called "the NISP projection matrix" which is given by

$$\boldsymbol{\Pi}_{k,j} = \frac{w_j \Psi_k(\boldsymbol{\xi}_j)}{\langle \Psi_k^2 \rangle}, \quad k = 0, \dots, P, \quad j = 1, \dots, N_q$$

Let $\boldsymbol{\zeta}$ be the vector with coordinates $\zeta_i = X(\boldsymbol{\xi}_i)$, then the vector $\mathbf{c} = \boldsymbol{\Pi}\boldsymbol{\zeta}$, or in component form

$$c_k = \sum_{j=1}^{N_q} \boldsymbol{\Pi}_{kj} \zeta_j = \sum_{j=1}^{N_q} \boldsymbol{\Pi}_{kj} X(\boldsymbol{\xi}_j), \quad k = 0, \dots, P. \quad (4.54)$$

4.3 ADCIRC uncertainties recast as stochastic variables

Let $\mathbf{p} = (p_1, p_2)^T$ be the vector of random model inputs having uniform distribution as specified in Table 4.2. Specifically, the inputs p_i are parameterized by

Table 4.2: The random input parameters for ADCIRC

Parameter	Description	Distribution
p_1	(Manning's n coefficient in the open ocean)	$\mathcal{U}(-1, 1)$
p_2	(Manning's n coefficient in the landlocked area)	$\mathcal{U}(-1, 1)$

$\xi_i \sim \mathcal{U}(-1, 1), i = 1, 2$ through

$$p_i(\boldsymbol{\xi}) = \mu_i + \sigma_i \xi_i, \quad i = 1, 2 \quad (4.55)$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2)^T$, $\mu_i = (1/2)(a_i + b_i)$, and $\sigma_i = (1/2)(b_i - a_i)$, so that $p_i \sim \mathcal{U}(a_i, b_i)$ as in Table 4.2. Now let \mathcal{G} be the physical domain in consideration. At a given time t and a point $\mathbf{x} \in \mathcal{G}$ and with a given vector of random inputs $\mathbf{p}(\boldsymbol{\xi})$ we have the model output denoted by $\mathbf{X}(t, \mathbf{x}, \boldsymbol{\xi})$. Here

$$\mathbf{X}(t, \mathbf{x}, \boldsymbol{\xi}) = \mathbf{A}(t, \mathbf{x}; \mathbf{p}(\boldsymbol{\xi})) \quad (4.56)$$

where $\mathbf{A}(t, \mathbf{x}; \mathbf{p}(\boldsymbol{\xi}))$ is the output of a deterministic ADCIRC solved at time t and point \mathbf{x} with the input parameters $\mathbf{p}(\boldsymbol{\xi})$. \mathbf{X} may correspond to any output considered the quantities of interest (QoIs).

4.3.1 PC representation of ADCIRC model for the idealized ebb shoal

Based on the results of the SEIK implementation in chapter 3, we will only apply the PC representation to the idealized Ebb Shoal case. The polynomial chaos is applied to the ADCIRC model, where the boundary conditions and initial conditions are assumed perfectly known as in chapter 3. We used NISP with a level 6 Smolyak quadrature to compute the spectral expansion of the model output in the PC basis. Figure 4.4 plots the water elevation in time for both original ADCIRC model and its surrogate counterpart at station one of the idealized Inlet Ebb Shoal. To ensure that level 6 quadrature is a good representation of ADCIRC in this specific case, we generated many PC version of the model with different number of quadrature levels ranging from 1 up to 6 and observed the convergence of these representations. This

is shown in Figure 4.3 alongside 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 $p = 6$ suggesting that a sixth-order expansion is sufficient. In addition, we also compute the RMSE between the water elevation from ADCIRC model and its counterpart. The RMSE is less than 5% of the wave amplitude for all quadrature points. The errors are shown in Table 4.3.

Table 4.3: The random input parameters for RMSE between the true model and the surrogate model (water elevations)

Quadrature point	RMSE (at station 1)
1 _{st} quadrature	0.088
10 _{th} quadrature	0.0011
20 _{th} quadrature	0.00036
30 _{th} quadrature	0.0012

4.4 Manning's n coefficient estimation using PC-based ADCIRC model and MCMC

The following results are the posterior distribution of Manning's n coefficients compared between SEIK and MCMC with the optimal β^2 . They were generated using 100000 iterations of MCMC. The accuracy of the estimations from MCMC are not affected by the initial guesses while SEIK is suffer the recovering of high value Manning's n coefficients from the low-value initial guesses. Since the inflation factor was used to improve the posterior from SEIK in chapter 3, we can show the comparison between the two schemes (SEIK and MCMC) with and without inflation factor for SEIK.

Figure 4.7 and 4.11 shows the comparison of the posterior pdfs of Manning's n coefficient between SEIK filter approach with 10 ensemble members and MCMC in a constant field and 2D parameterized field respectively. The pdfs produced by

SEIK is much more narrow in comparison to those produced by MCMC. This is due to our initial setup of background error covariance when we run SEIK. SEIK filter also underestimated the truths in many cases while MCMC experiences no difficulty in the mentioned manner. The distributions produced by MCMC cover the wide range of Manning's n coefficient and clearly portray to the Gaussian distribution with their means approximately at the truth values. It is concluded that SEIK filter with small ensembles size is not a good representation to propagate the uncertainty of the estimated parameters. However, the distribution can be improved to match the goal standard MCMC by using background error covariance inflation.

Figure 4.10 and 4.12 shows the comparison of the posterior pdfs of Manning's n coefficient between SEIK filter with the inflation factor of 2 and MCMC in a constant field and 2D parameterized field respectively. The effect of enforcing inflation factor is more spread among the samples of the joint pdfs produced by SEIK as well as the improvement in the estimation means of Manning's n coefficients in all cases. The parameter estimation with SEIK filter in this case appears to be a good representation of the true posterior distribution formulated by Bayes's rule.

4.5 Conclusion

The main focus of the theoretical part of this chapter is to portray of the link between sequential data assimilation and the general parameter estimation framework based on Bayesian theorem. We shown that any form of sequential data assimilation scheme including SEIK can be fundamentally derived from the Bayes's rule for the first-order Markov process. Given the sufficient amount of the observations, Bayes's rule gives the full pdfs of the estimated parameters with no Gaussian restriction as prescribed by Kalman filtering framework. The sampling of the posterior from the Bayes's rule using Metropolis-Hasting algorithm in MCMC is treated as the reference to

other data assimilation schemes. Thus, for the experimental part of this chapter, we compared the pdfs of Manning's n coefficient from SEIK filter both with and without inflation factor to the gold standard MCMC method. The result reflect that the uncertainty representation by SEIK for parameter estimation with inflation is a good approximation of the true pdfs computed from the Bayes's rule.

Figure 4.1: (Left) The distributions of the ADCIRC surrogate model output (water elevation) of idealized Ebb Shoal case with several polynomial orders at the selected points and times. The distributions are created by a million samples. (Right) Polynomial chaos coefficients at the corresponding points and times.

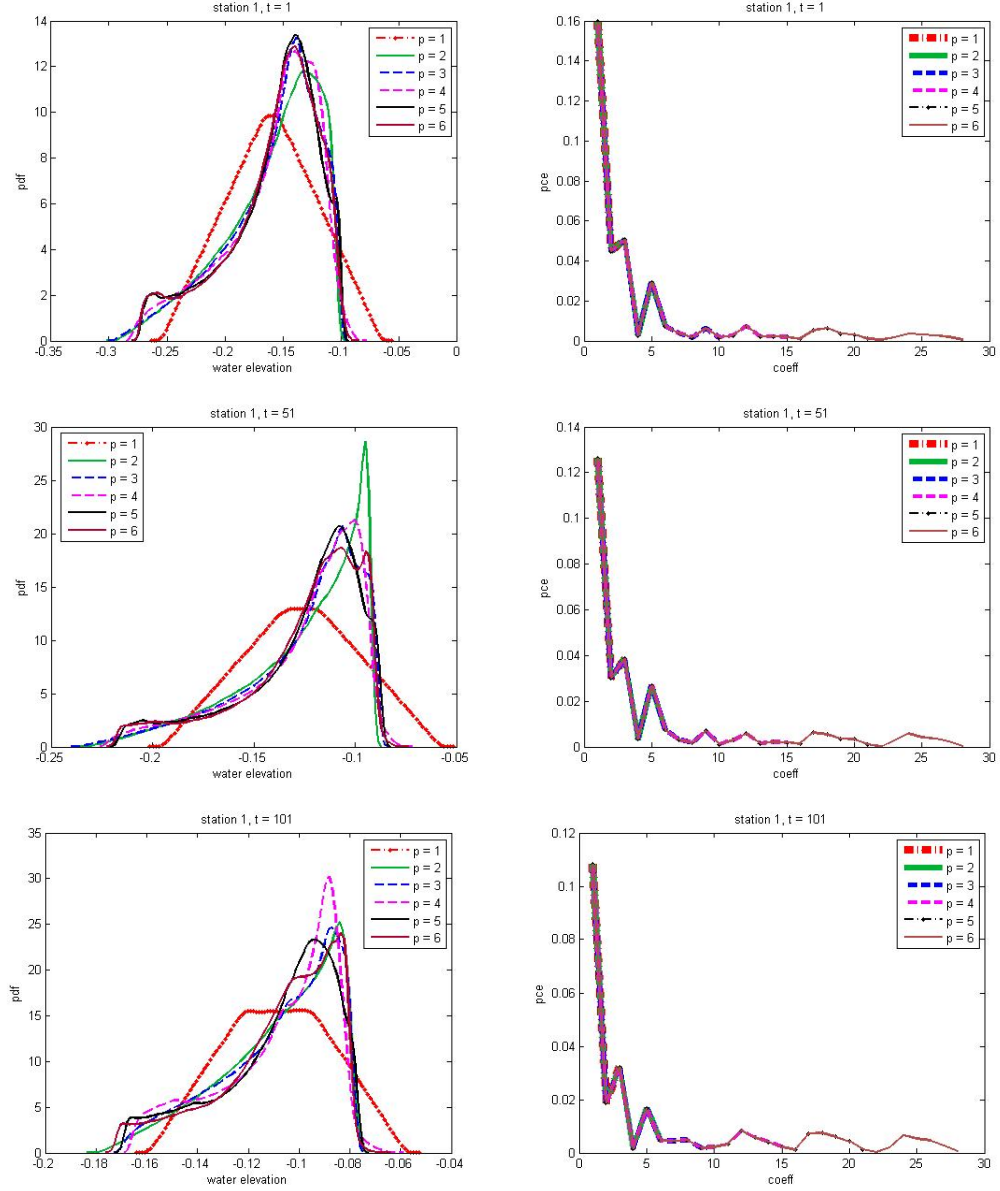


Figure 4.2: (Left) The distributions of the ADCIRC surrogate model output (water elevation) of idealized Ebb Shoal case with several polynomial orders at the selected points and times. The distributions are created by a million samples. (Right) Polynomial chaos coefficients at the corresponding points and times.

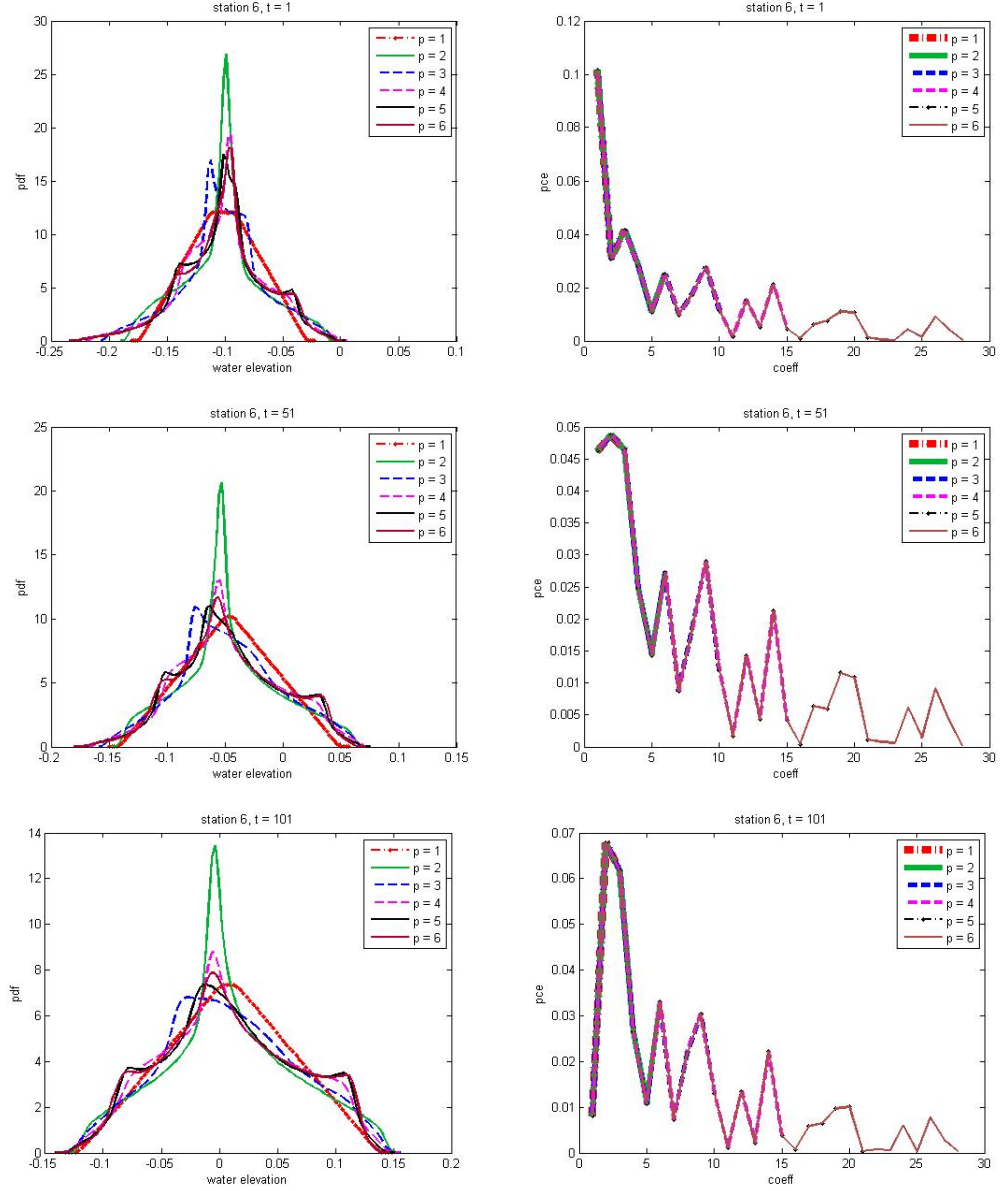


Figure 4.3: (Left) The distributions of the ADCIRC surrogate model output (water elevation) of idealized Ebb Shoal case with several polynomial orders at the selected points and times. The distributions are created by a million samples. (Right) Polynomial chaos coefficients at the corresponding points and times.

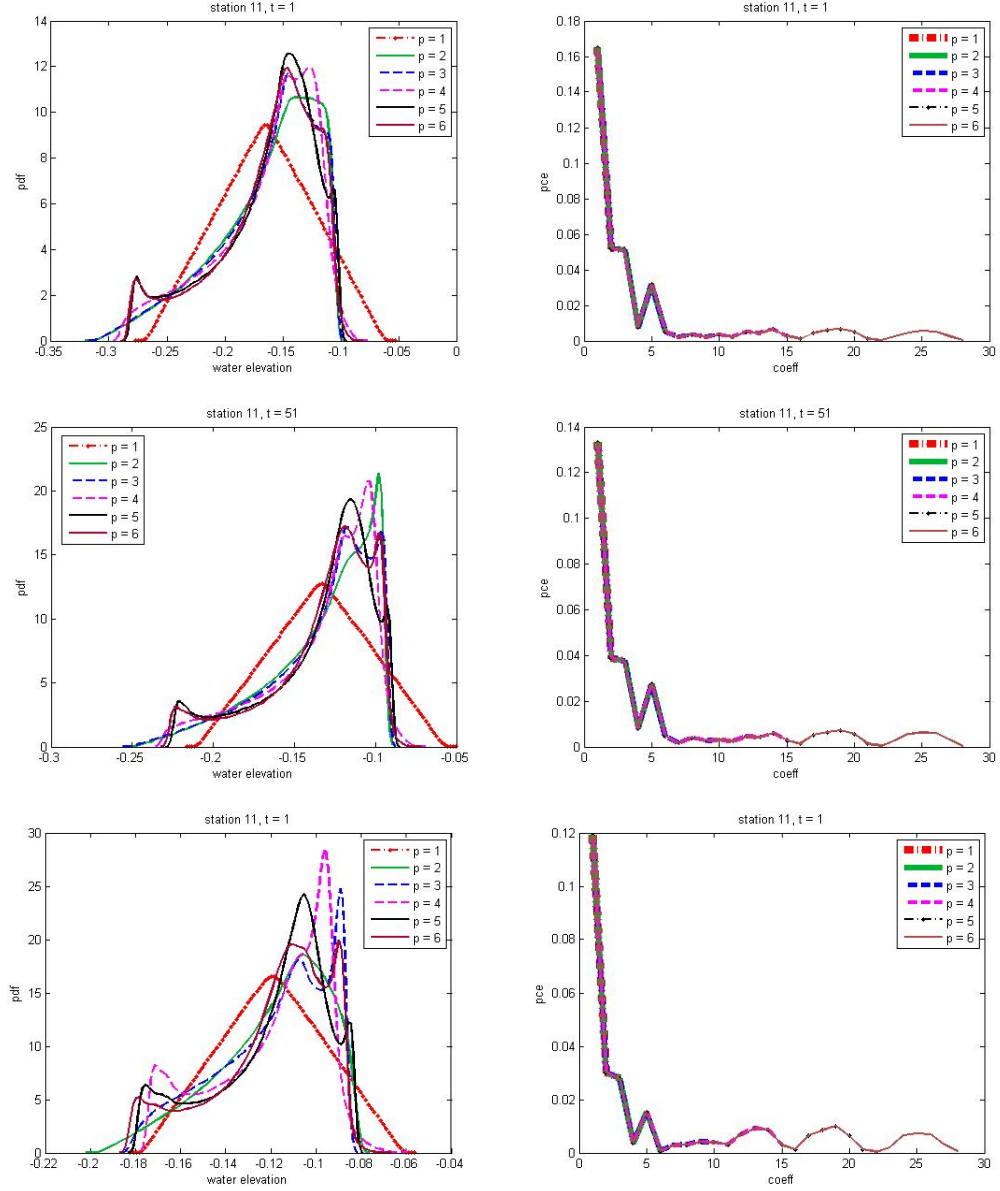


Figure 4.4: (Right) Water elevations simulated in time from the true model and the surrogate model at several quadrature points (different Manning's n coefficients). (Left) The residuals.

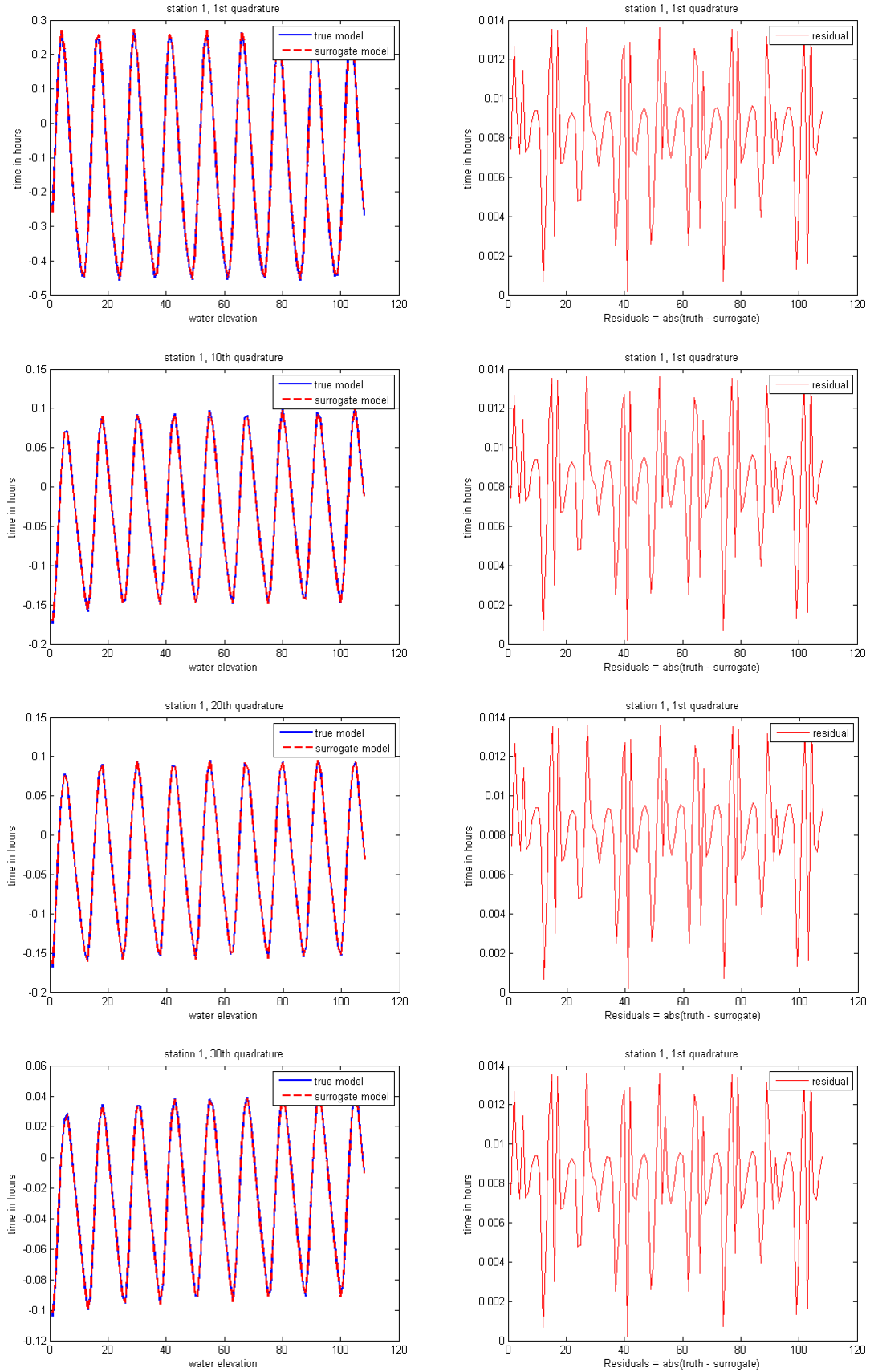


Figure 4.5: The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter(after 108 analysis iterations)and MCMC method.

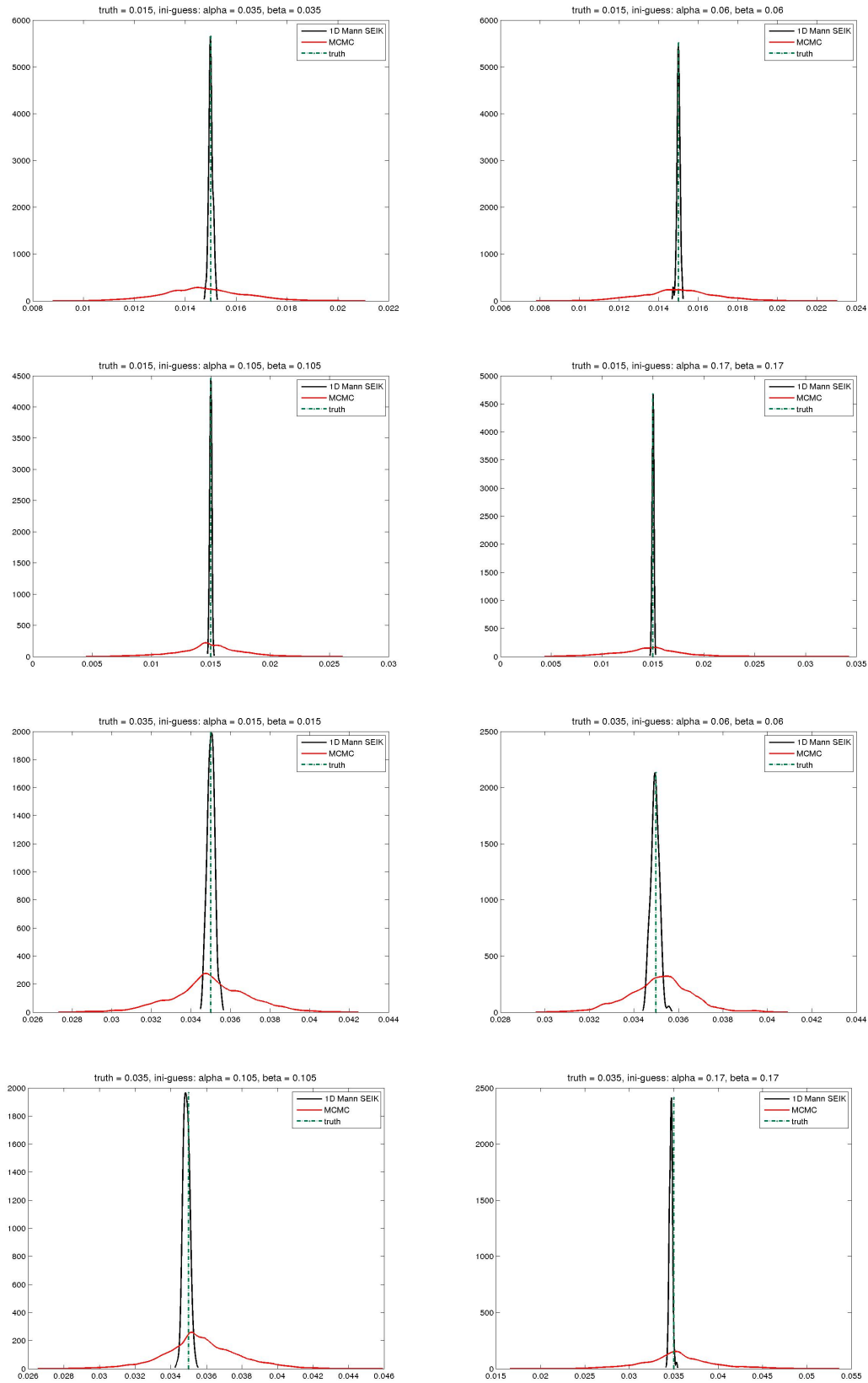


Figure 4.6: The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter(after 108 analysis iterations)and MCMC method.

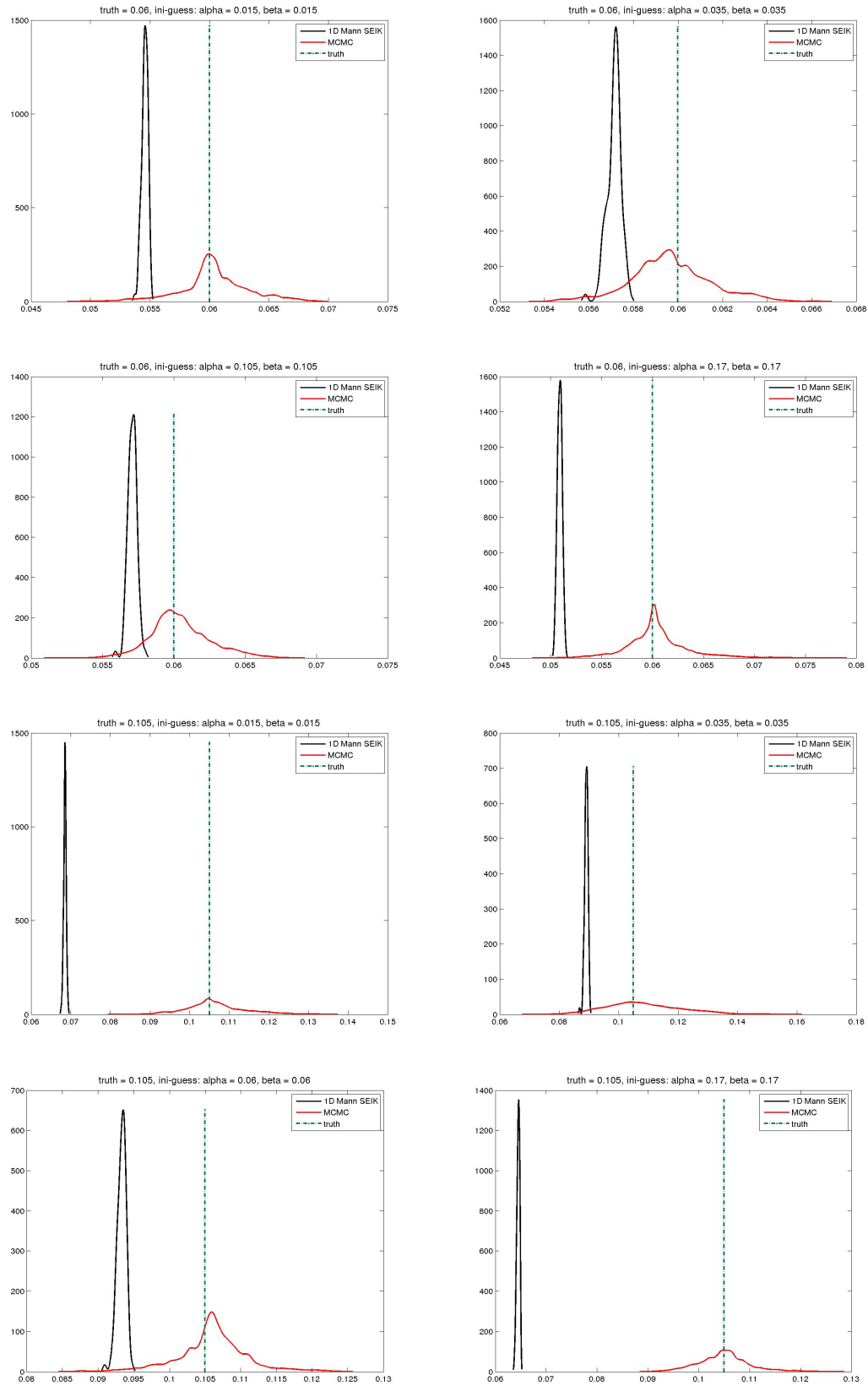


Figure 4.7: The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter(after 108 analysis iterations)and MCMC method.

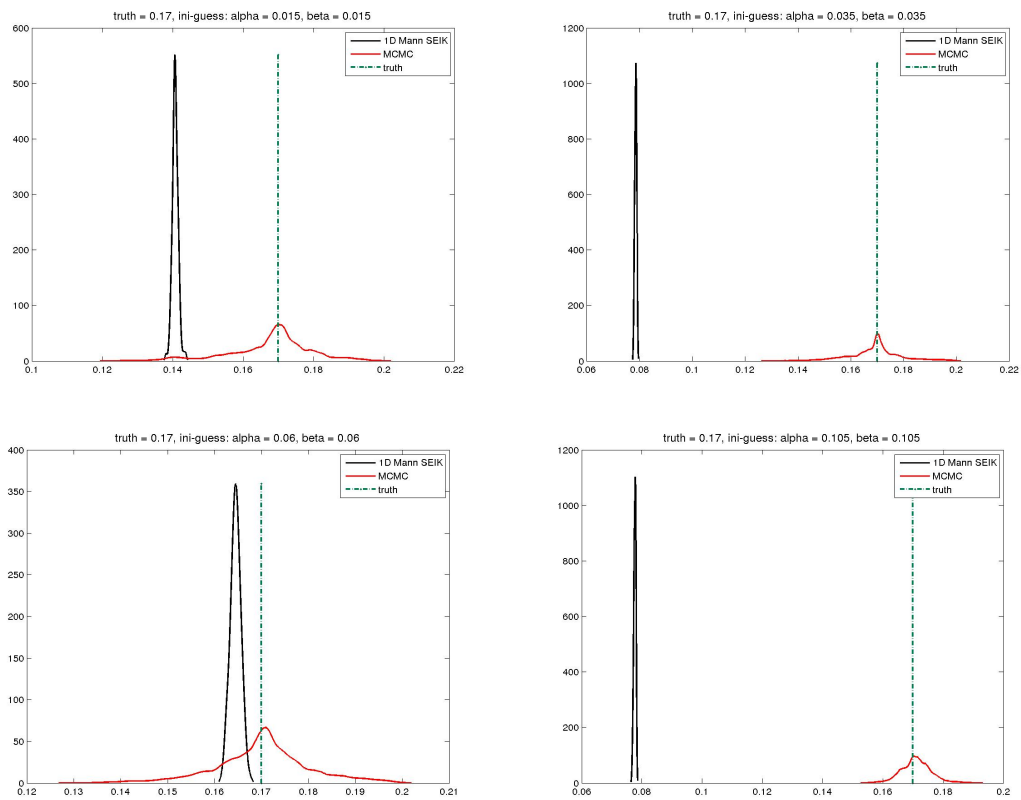


Figure 4.8: The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter with inflation(after 108 analysis iterations) and MCMC method.

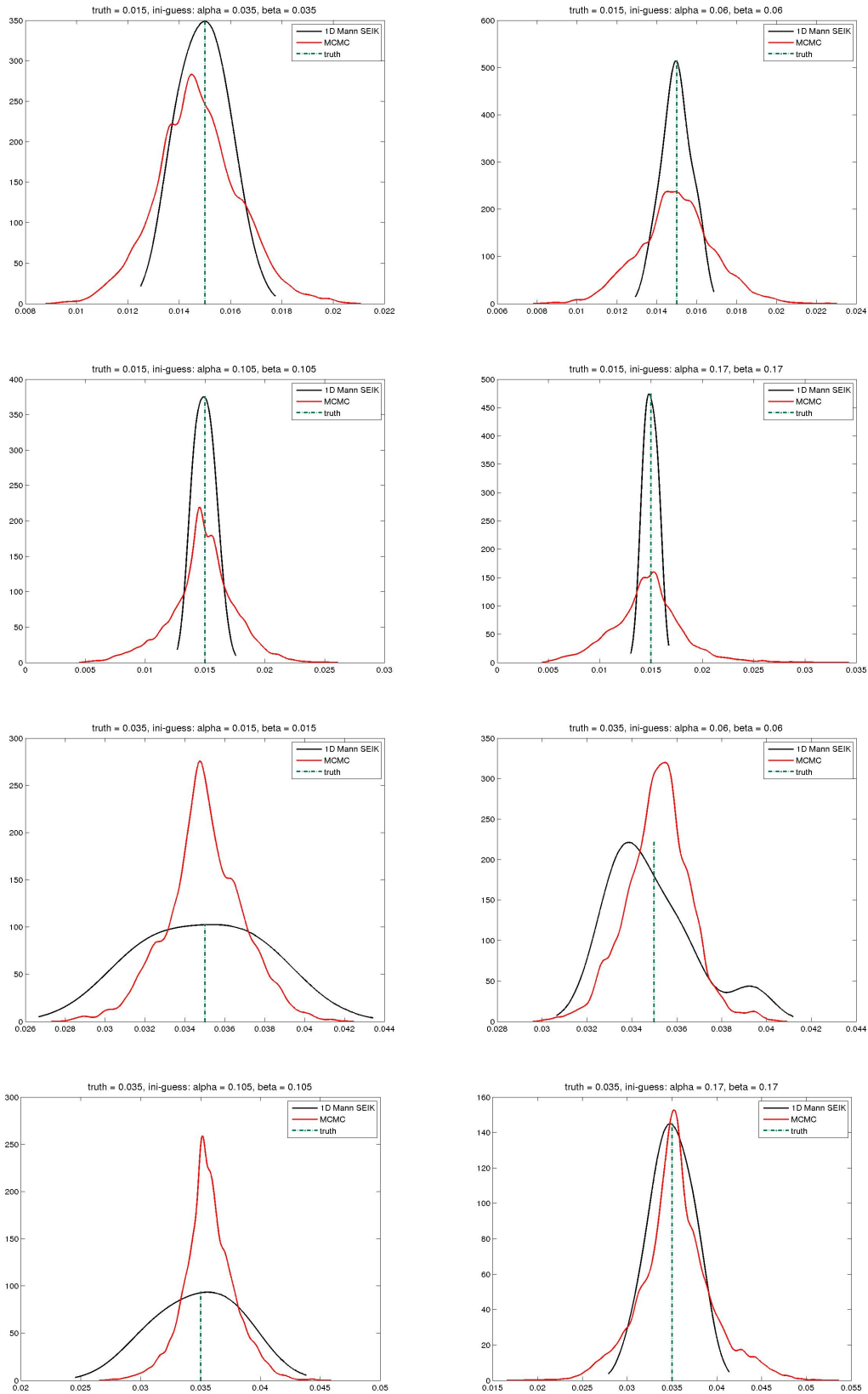


Figure 4.9: The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter with inflation(after 108 analysis iterations)and MCMC method.

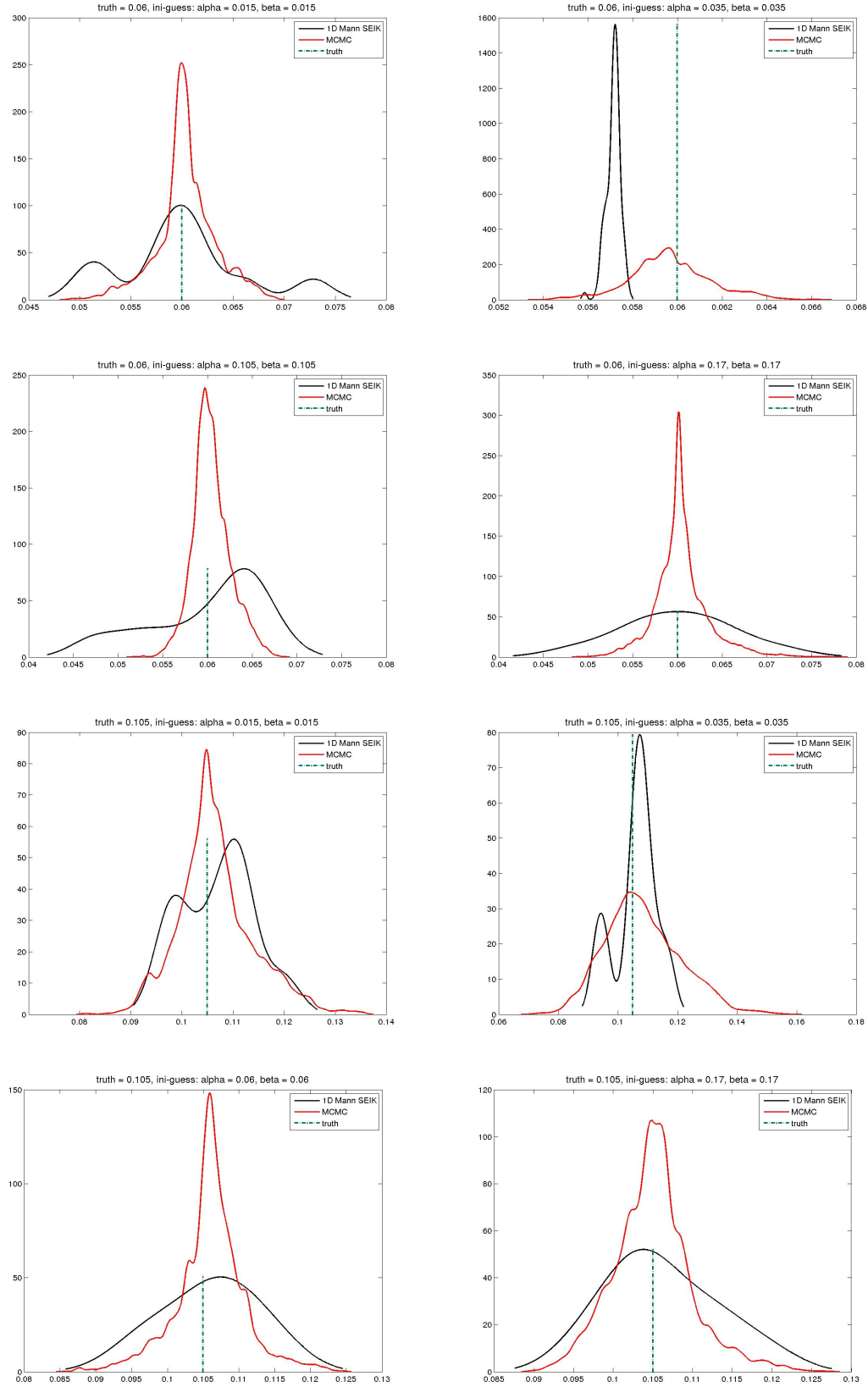


Figure 4.10: The comparisons between the posterior distributions of 1D Manning's n coefficient between SEIK filter with inflation(after 108 analysis iterations)and MCMC method.

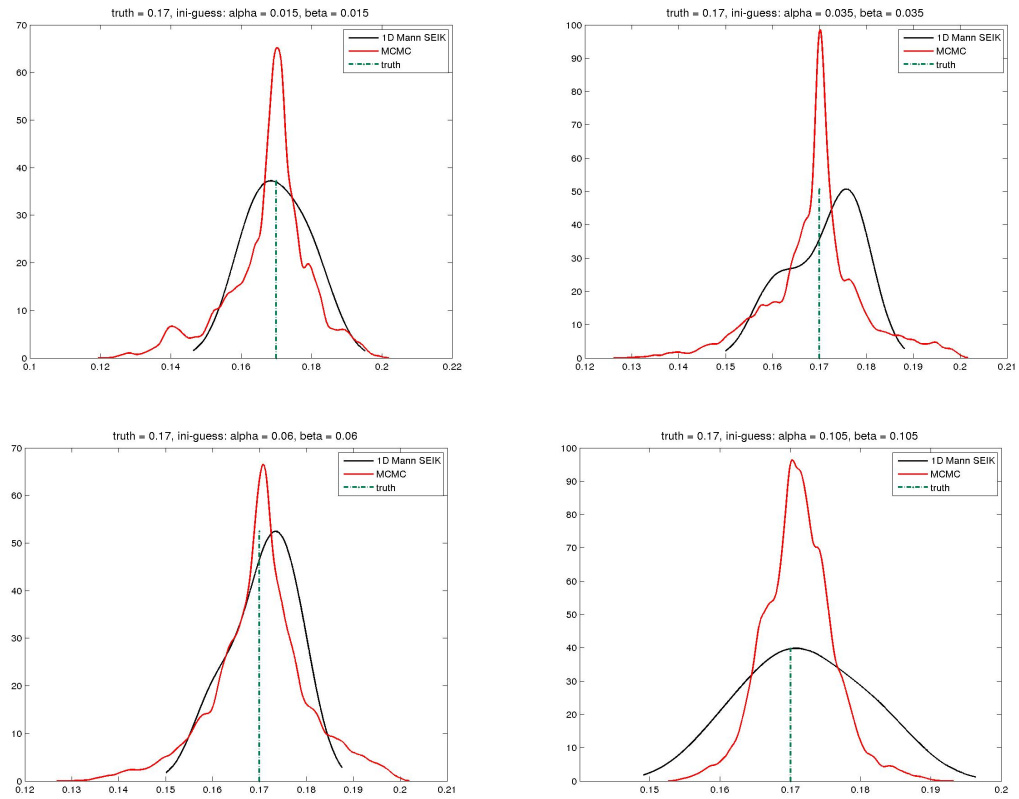


Figure 4.11: The comparisons between the posterior distributions of 2D Manning's n coefficient between SEIK filter(after 108th analysis iterations) and MCMC method, (Right) the distributions of α , (Left) the distribution of β .

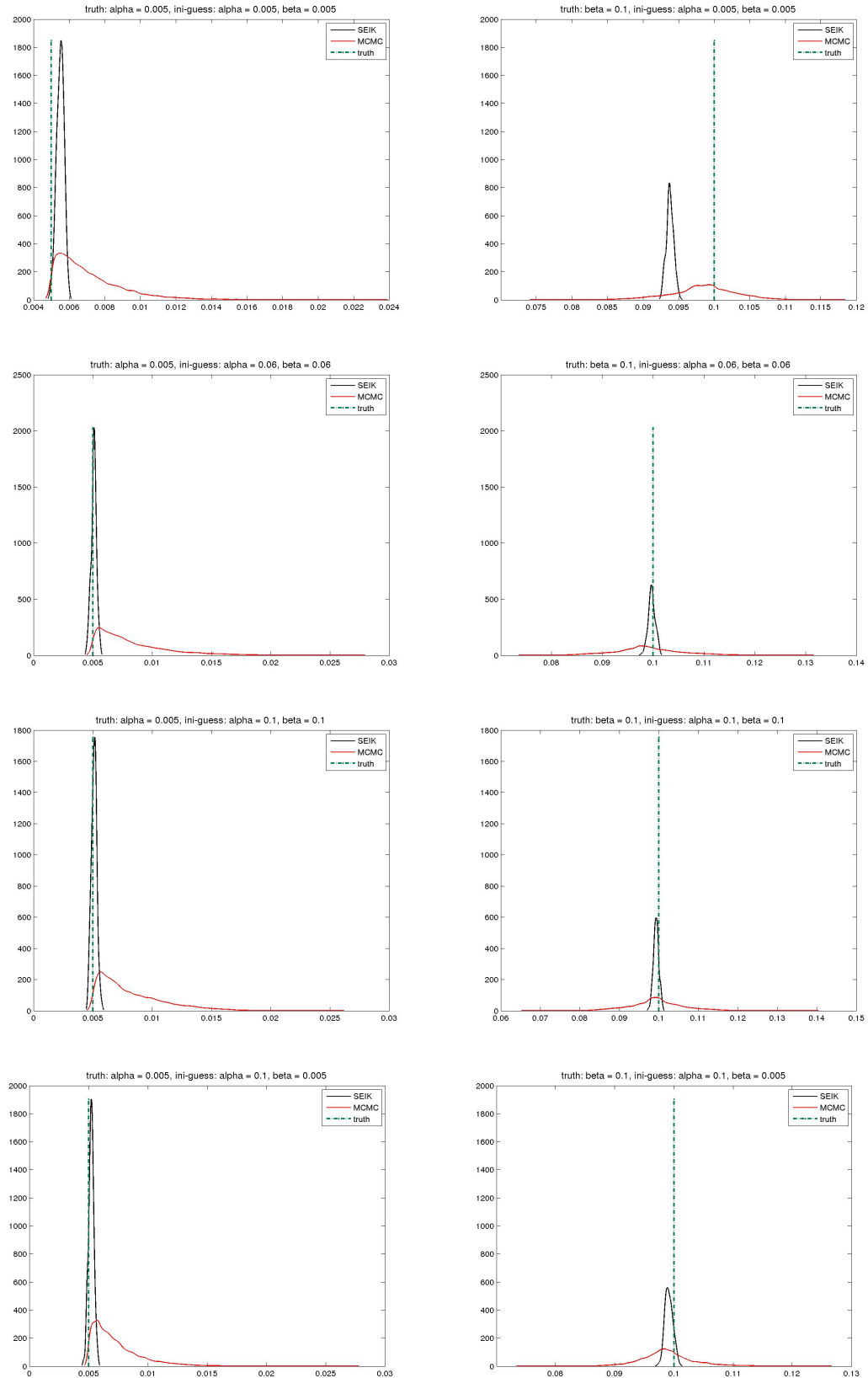
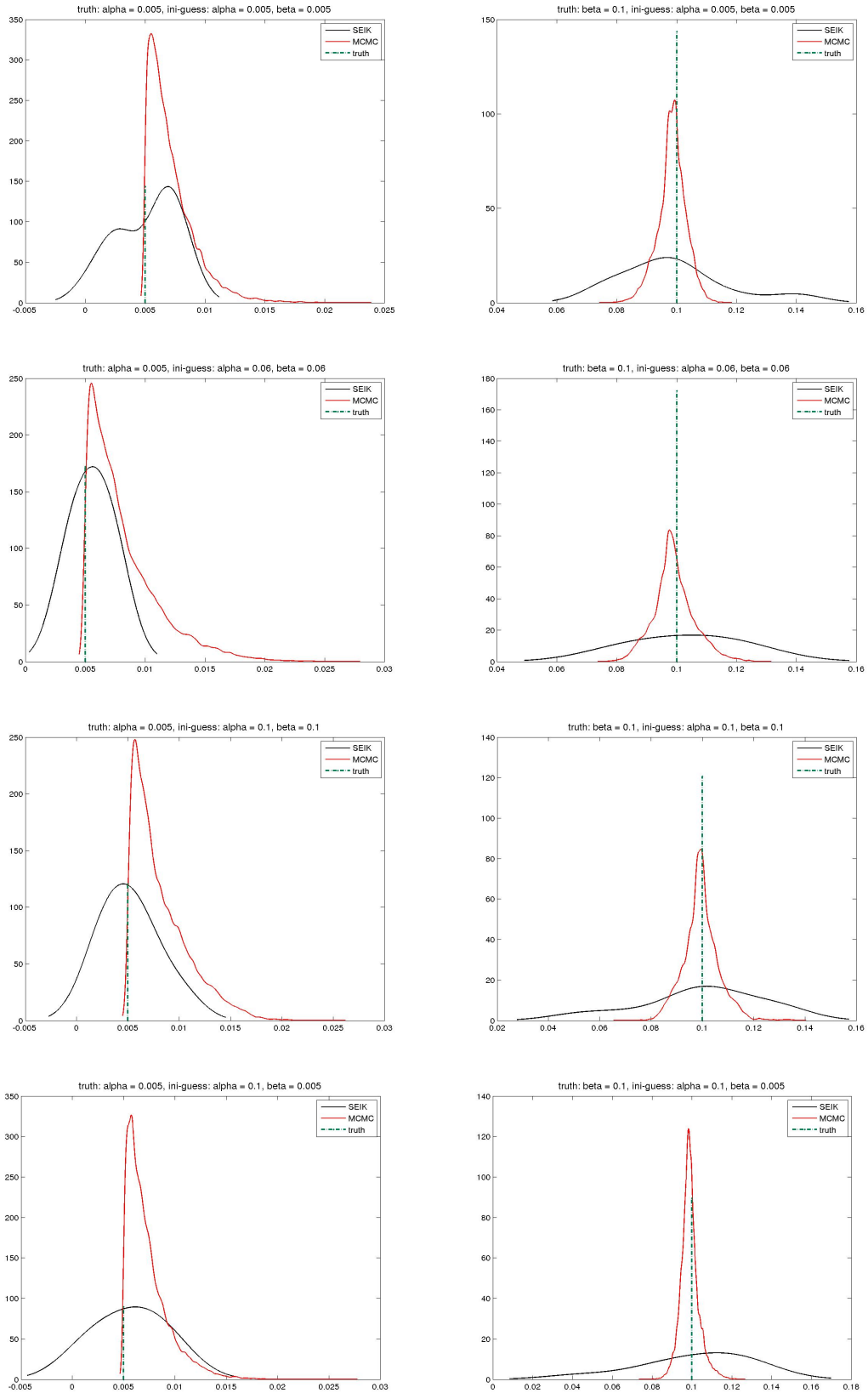


Figure 4.12: The comparisons between the posterior distributions of 2D Manning's n coefficient between SEIK filter(after 108th analysis iterations) and MCMC method, (Right) the distributions of α , (Left) the distribution of β .



Chapter 5

Concluding Remarks

The parameter estimation problem is the on-going research area requires that there is sufficient data provided from the real-world observation system. The inverse problem can then be formulated in order to estimate the parameter in the model that fit the observations. This thesis focus on determining the parameters related to the ocean modeling system. The accurate knowledge about the bottom stress terms in a physical domain is essential for the modeling of the hydrodynamics system. In the field of ocean modeling, bottom stress terms are embedded in the momentum equation of the shallow water model, which is used vastly in the prediction sea level, tides and storm surge under both moderate and extreme conditions. These stress terms is formulated using Manning's n coefficient which relating the bottom stress to the height of the water column above it. However, Manning's n coefficient is empirically defined constant and always subjected to large amount of uncertainty in the real-world physical simulations. The recovering of the true Manning's n coefficient from the observations is an active research area. An attempt to find the efficient framework for the parameter estimation in storm surge and tides modeling is served as the prior guaranty to the effective coastal ocean prediction system.

In this work, we made an attempt to validate the correctness in estimating Manning's n coefficient by SEIK filter, which is one of the most recent scheme in statistical data assimilation introduced in many literature, by comparing the posterior distri-

bution of the parameter we estimated from SEIK to the distribution from the gold standard approach, MCMC as our reference.

In the first part, we conducted the Observation Simulation System Experiments (OSSEs) on Advance Circulation (ADCIRC) model. We generated the synthetic water elevation data from several fields of Manning's n coefficient. Using the singular interpolated evolutive Kalman (SEIK) filter, we then recover these Manning's n coefficient fields from incorrect initial guesses. We separated our OSSEs in to two domains, one is idealized inlet with Ebb Shoal with comprising the open ocean area, landlocked area and the inlet, another is the Galveston Bay represents the more realistic version of the first domain. We estimated two configuration of Manning's n coefficient fields. The first was defined by a constant Manning's n coefficient across the domain. The second configuration is 2D parameterized piecewise Manning's n coefficient where the bottom stress in the bay was defined by one constant value, and the bottom stress in the open ocean was defined by the other. For the constant Manning's field, with small ensemble size (i.e. 10 ensemble members), there arise the difficulty in estimating the large Manning's field from the small initial guess for both domain. This is mainly due to the sensitivity difference of ADCIRC between the small Manning's n coefficient values and the large value. ADCIRC is more sensitive to the change in small Manning's n coefficient compare to the same amount of change in large Manning's n coefficient. Increasing the size of ensemble members improved the estimations. We have experimented with 10, 20 and 100 ensemble members for all cases. The estimation from the small initial guess increase by 200% when using 100 ensemble members for the most difficult case (i.e. estimating the largest Manning's n field from the smallest initial guess). For the 2D parameterized field of both domain, the use of small ensemble size is sufficient to recover the truths. Since the increasing of ensemble members requires greater computation, we instead used an inflation factor as the solution to increase the the background error covariance. We found that the

estimations converge to the truth faster with less computation compare to increasing ensemble size and the estimation of the means are accurate regardless of the accuracy of initial guesses. It also can be observed that the use of inflation factor has widen the posterior distribution of the parameter produced by SEIK.

In the second part, a spectral projection approach was implemented to propagate parametric uncertainties in a coastal ocean model. A non-intrusive spectral projection scheme was used to derive the PC representation of ADCIRC for selected Quantity Of Interests (QoIs), which in our case, the water elevation. Using PC expansion, we created the surrogate ADCIRC model of the idealized inlet with Ebb Shoal with the same forcing as the previous OSSEs with SIEK. The input of the surrogate ADCIRC model requires two inputs including the Manning's n coefficient in the landlocked area and another in the open ocean. The output of the surrogate model following the variation in these inputs are the water elevations at each node on the domain in time. The merit of applying spectral projection approach to the model is that its surrogate counterpart is much faster to execute, which allows the high frequency sampling of the model in practical application possible. For our presented setups, conditions and dimensions of stochastic problem, a level 6 sparse quadrature was found to be adequate for producing the desirable model response.

In the last part of this study, we investigated more deeply into the Bayesian viewpoint of sequential data assimilation. We emphasized that Kalman filter can be viewed as the derivation of Bayesian's rule sequentially in time. The Bayes's rule compute joint probability density functions (pdfs) of the estimated parameters given the model and a set of observations. The surrogate version of ADCIRC created by PC expansion was used in MCMC sampling of the posterior distribution of Manning's n coefficients for the idealized inlet with Ebb Shoal. The implementation of PC and MCMC alleviate upon the restriction of Gaussian posterior distribution imposed by implementing Kalman filter to the parameter estimation problem. The Metropolis-

Hastings algorithm is used for our sampling of the posterior distribution for both a constant Manning's n field and 2D parameterized Manning's n field. We finally compared the posterior distribution produced by SEIK filter and MCMC on PC-based model, it is found that the Manning's n coefficient estimation with SEIK, when using the inflation factor, is a good approximation to the pdfs produced by the gold standard MCMC.

In conclusion, while the PC-based approach coupled with MCMC for parameter estimation is well-served for offline application with low dimension stochastic system, it is not designed for the usage in the online coastal ocean forecasting system which includes a large number of uncertain parameters. The sequential data assimilation scheme such as Kalman filter provides a robust solution to approximate the pdfs in real-time. However with the Guassion restriction, the parameter uncertainties propagation by Kalman filter can be deceiving. The result of our study shows that implementing the SEK filter to estimate Manning's n coefficient in ADCIRC model, in additional with appropriate inflation factor, produces the pdfs close to that of the goal standard MCMC. This indicates that the use SEIK filter is valid and efficient to estimate the uncertainty of low dimensional bottom stress term in ADCIRC model.

REFERENCES

- [1] T. Yanagi, *Coastal oceanography, Vol. 1.* Springer, 1999.
- [2] T. Mayo, T. Butler, C. Dawson, and I. Hoteit, “Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of manning’s friction coefficient,” *Ocean Modelling*, vol. 76, pp. 43–58, April 2014.
- [3] J. Dietrich, J. Westerink, A. Kennedy, J. Smith, R. Jensen, M. Zijlema, L. Holthuijsen, C. Dawson, R. L. Jr, M. Powell, and et al., “Hurricane gustav (2008) waves and storm surge: Hindcast, synoptic analysis, and validation in southern louisiana,” *Monthly Weather Review*, vol. 139, no. 8, pp. 2488–2522, 2011.
- [4] T. Mayo, “Data assimilation for parameter estimation i coastal ocean hydrodynamics modeling.”
- [5] A. B. Kennedy, U. Gravois, B. C. Zachry, J. J. Westerink, M. E. Hope, J. C. Dietrich, M. D. Powell, A. T. Cox, R. A. Luettich, and R. G. Dean, “Origin of the hurricane ike forerunner surge,” *Geophysical Research Letters*, vol. 38, no. 8, 2011.
- [6] C. P. Jelesnianski, “Numerical computations of storm surges without bottom stress,” *Monthly Weather Review*, vol. 94, no. 6, pp. 379–394, 1966.
- [7] S. E. Kaipio J., “Statistical and computational inverse problems,” *Applied Mathematical Sciences*, vol. 160, no. ISBN 0-387-22073-9, p. 344, 2005.
- [8] G. Robert, P. Casella, “Monte carlo statistical methods,” *Applied Mathematical Sciences*, vol. 160, no. ISBN 0-387-22073-9, p. 344, 2004.

- [9] R. Kalman, “a new approach to linear filtering and prediction problems 1,” *Transactions of the ASME Journal of Basic Engineering*, vol. 82, no. Series D., pp. 35–45, 1960.
- [10] J. V. T. Sorensen and H. Madsen, “Parameter sensitivity of three kalman filter schemes for assimilation of water levels in shelf sea models,” *Ocean Modelling*, vol. 11, pp. 441–463, 2006.
- [11] G. Y. H. E. Serafy and A. E. Mynett, “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*, vol. 44, pp. W06 416, doi:10.1029 2006WR005 412, 2008.
- [12] A.W.Heemink and H. Kloosterhuis, “Data assimilation for non-linear tidal models,” *International Journal of numerical methods in fluids*, vol. 11, pp. 1097–1112, 1990.
- [13] T. Butler, M. U. Altaf, C. Dawson, I. Hoteit, X. Luo, and T. Mayo, “Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting,” *Monthly Weather Review*, vol. 140, pp. 2215–2231, 2012.
- [14] A. Bennet, *Inverse Methods in Physical Oceanography*. Cambridge University Press, 1992.
- [15] I. Hoteit, G. Triantafyllou, and G. Patihakis, “Efficient data assimilation into a complex 3d physical-biogeochemical model using a partially local kalman filter,” *Annales Geophysicae*, vol. 23, pp. 1–15, 2005.
- [16] P. M. Tagade and H.-L. Choia, “A generalized polynomial chaos-based method for efficient bayesian calibration of uncertain computational models,” *Inverse Problems in Science and Engineering*, vol. 22, no. 4, pp. 602–624, 2014.
- [17] N. Oreskes, K. Shrader-Frechett, and K. Belitz, “Verification, validation and confirmation of numerical models in earth sciences, science,” *Science*, vol. 263, pp. 641–647, 1994.

- [18] U. Mehta, “Some aspects of uncertainty in computational fluid dynamics results,” *J. of Fluid Engg.*, vol. 113, pp. 538–543, 1991.
- [19] S. Cheung, T. Oliver, E. Prudencio, S. Prudhomme, and R. Moser, “Bayesian uncertainty analysis with applications to turbulence modeling,” *Rel. Engg. and Syst. Safety*, vol. 96, pp. 1137–1149, 2011.
- [20] M. Kennedy and A. OHagan, “Bayesian calibration of computer models,” *J. of the Royal Stat. Soc. Series B (Stat. Method.) 63 (2001)*, vol. 63, pp. 425–464, 2011.
- [21] D. H. J. Besag, P. Green and K. Mengersen, “Bayesian computation and stochastic systems,” *Stat. Sci.*, vol. 10, pp. 3–41, Mon. 1995.
- [22] D. Gamerman and H. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Boca Raton: Chapman and Hall/CRC,, 2006.
- [23] R. Ghanem and P. Spanos, : *Stochastic Finite Elements: A Spectral Approach*, 2nd ed. New York: Dover, 2002.
- [24] O. Matre and O. Knio, *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*. Berlin: Springer, 2010.
- [25] O. L. Matre, O. Knio, H. Najm, and R. Ghanem, “A stochastic projection method for fluid flow. i. basic formulation,” *Journal of Computational Physics*, vol. 173, pp. 481–511, 2001.
- [26] O. L. Matre, H. Najm, P. Pbay, R. Ghanem, and O. Knio, “Multi-resolution-analysis scheme for uncertainty quantification in chemical systems,” *SIAM Journal Scientific Computation*, vol. 29, no. 2, pp. 864–889, 2007.
- [27] H. Najm, B. Debusschere, Y. Marzouk, S. Widmer, and O. L. Matre, “Multi-resolution-analysis scheme for uncertainty quantification in chemical systems,” *International Journal of Numerical Methods in Engineering*, vol. 80, no. 6, pp. 789–814, 2009.

- [28] B. Phenix, J. Dinaro, M. Tatang, J. Tester, J. Howard, and G. McRae, “Incorporation of parametric uncertainty into complex kinetic mechanisms: application to hydrogen oxidation in supercritical water,” *Combustion and Flame*, vol. 112, pp. 132–146, 1998.
- [29] M. Reagan, H. Najm, P. Pbay, O. Knio, and R. Ghanem, “Incorporation of parametric uncertainty into complex kinetic mechanisms: application to hydrogen oxidation in supercritical water,” *International Journal of Chemical Kinetics*, vol. 37, no. 6, pp. 368–382, 2005.
- [30] W. Meecham and D. Jeng, “Use of the wiener-hermite expansion for nearly normal turbulence,” *J. of Fluid Mech.*, vol. 32, pp. 225–249, 1968.
- [31] A. Chorin, “Gaussian fields and random flow,” *J. of Fluid Mech.*, vol. 85, pp. 325–347, 1974.
- [32] R. Ghanem and J. Red-Horse, “Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach,” *Fluid Dyn. Res.*, vol. 133, pp. 137–144, 1999.
- [33] O. Knio and O. Maitre, “Uncertainty propagation in cfd using polynomial chaos decomposition,” *Physica D*, vol. 38, pp. 616–640, 2006.
- [34] D. Xiu and G. Karniadakis, “The weiner-askey polynomial chaos for stochastic differential equations,” *SIAM J. of Sci. Comp.*, vol. 24, pp. 619–644, 2002.
- [35] —, “Modeling uncertainty in flow simulations via generalized polynomial chaos,” *J. of Comp. Phys.*, vol. 187, pp. 137–167, 2003.
- [36] D. Xiu, *Numerical methods for stochastic computation: A spectral approach*. New Jersey: Princeton University Press, 2010.
- [37] N. Wiener, “The homogeneous chaos,” *American Journal of Mathematics*, vol. 60, no. 4, pp. 897–936, 1938.
- [38] R. Ghanem, *Stochastic finite elements: a spectral approach*.

- [39] D. Pham, “Stochastic methods for sequential data assimilation in strongly non-linear systems,” *Monthly weather review*, vol. 12, no. 5, pp. 1194–1207, 2001.
- [40] J. B. I. Hoteit, D. Pham, “A simplified reduced order kalman filtering and application to altimetric data assimilation in tropical pacific,” *Journal of Marine systems*, vol. 36, no. 1, pp. 101–127, 2002.
- [41] G. P. G. Triantafyllou, I. Hoteit, “A singular evolutive interpolated kalman filter for efficient data assimilation in a 3-d complex physicalbiogeochemical model of the cretan sea,” *Journal of Marine systems*, vol. 40, pp. 213–231, 2003.
- [42] A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. C. Thacker, and O. M. Knios, “Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach,” *Computational Geoscience*, vol. 16, pp. 757–778, 2012.
- [43] R. Luettich and J. Westerink, *Formulation and numerical implementation of the 2D/3D ADCIRC finite element model version 44. XX*. R. Luettich, 2004.
- [44] C. B. N. S. J.J.Westerink, R.A. Luettich, “Adcirc: an advanced three-dimensional circulation model for shelves, coasts and estuaries. report 1: theory and methodology of adcirc-2ddi and adcirc-3dl,” *Technical Report DRP-92-6*, 1992.
- [45] C. Dawson and J. Proft, “Coupled discontinuous and continuous galerkin finite element methods for the depth-integrated shallow water equations,,” *Computer Methods in Applied Mechanics and Engineering*, vol. 193, no. 3, pp. 289–318, 2004.
- [46] D. Hill, “Tidal modeling of glacier bay, alaska - methodology, results, and applications,” Department of Civil, Environmental Engineering, The Pennsylvania State University, Tech. Rep., 2007.
- [47] S. Bunya, J. Dietrich, J. Westerink, B. Ebersole, J. Smith, J. Atkinson, R. Jensen, D. Resio, R. Luettich, C. Dawson, and 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*, vol. 138, no. 2, pp. 345–377, 2010.
- [48] J. Dietrich, S. Bunya, J. Westerink, B. Ebersole, J. Smith, J. Atkinson, R. Jensen, D. Resio, R. Luettich, C. Dawson, and 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*, vol. 138, no. 2, pp. 378–404, 2010.
- [49] J. J. Westerink, R. A. Luettich, J. C. Feyen, J. H. Atkinson, C. Dawson, H. J. Roberts, M. D. Powell, J. P. Dunion, E. J. Kubatko, and H. Pourtaheri, “A basin-to channel-scale unstructured grid hurricane storm surge model applied to southern louisiana,” *Monthly Weather Review*, vol. 136, no. 3, pp. 833–864, 2008.
- [50] J. J. W. Ethan J Kubatko and C. Dawson, “Discontinuous galerkin methods for advection dominated problems in shallow water ow,” *Computer Methods in Applied Mechanics and Engineering*, vol. 196(1), pp. 437–451, 2006.
- [51] R. Berg, “Tropical cyclone report: Hurricane ike (al092008),1-14 september 2008,” *National Hurricane, Center,,* 2009.
- [52] P. T. Brummelhuis and A. Heemink, “Parameter identification in tidal models with uncertain boundary conditions,” *Stochastic Hydrology and Hydraulics*, vol. 4, no. 43, pp. 193–208, 1990.
- [53] P. T. Brummelhuis, A. Heemink, and H. V. D. Boogaard, “Identification of shallow sea models,” *International journal for numerical methods in fluids*, vol. 17, no. 8, pp. 637–665, 1993.
- [54] S. Das and R. Lardner, “On the estimation of parameters of hydraulic models by assimilation of periodic tidal data,” *Journal of Geophysical Research: Oceans (19782012)*, vol. 96, no. C8, pp. 15 187–15 196, 1991.
- [55] A. W. Lal, “Calibration of riverbed roughness,” *Journal of Hydraulic Engineering*, vol. 121, no. 9, pp. 664–671, 1995.

- [56] R. H. Khatibi, J. J. Williams, and P. R. Wormleaton, “Identification problem of open-channel friction parameters,” *Journal of Hydraulic Engineering*, vol. 123, no. 12, pp. 1078–1088, 1997.
- [57] G. A. Atanov, E. G. Evseeva, and E. A. Meselhe, “Estimation of roughness profile in trapezoidal open channels,” *Journal of Hydraulic Engineering*, vol. 8, no. 1, pp. 135–154, 2005.
- [58] R. Ramesh, B. Datta, S. M. Bhallamudi, and A. Narayana, “Optimal estimation of roughness in open-channel flows,” *Journal of Hydraulic Engineering*, vol. 126, no. 4, pp. 299–303, 2000.
- [59] A. Ishii and M. Kawahara, “Parameter identification of manning roughness coefficient using analysis of hydrauric jump with sediment transport,” *Proceeding of the first Asian-Pacific Congress on Computational Mechanics*, vol. 2, pp. 1071–1076, 2006.
- [60] I. Navon, “Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography,” *Dynamics of Atmospheres and Oceans*, vol. 27, no. 1, pp. 55–79, 1998.
- [61] W. Budgell, “Stochastic filtering of linear shallow water wave processes,” *SIAM journal on scientific and statistical computing*, vol. 8, no. 2, pp. 152–170, 1987.
- [62] M. Verlaan and A. Heemink, “Tidal flow forecasting using reduced rank square root filters,” *Stochastic Hydrology and Hydraulics*, vol. 11, no. 5, pp. 349–368, 1997.
- [63] J. V. T. Srensen and H. Madsen, “Efficient kalman filter techniques for the assimilation of tide gauge data in three-dimensional modeling of the north sea and baltic sea system,” *Journal of Geophysical Research: Oceans (1978-2012)*, vol. 109, no. C3, 2005.
- [64] M. U. Altaf, T. Butter, X. Luo, C. Dawson, T., and I. Hotit, “Improving short-range ensemble kalman storm surge forecasting using robust adaptive inflation,” *Monthly Weather Reviews*, vol. 141, pp. 2705–2720, 2013.

- [65] O.-P. Tossavainen, J. Percelay, A. Tinka, Q. Wu, and fAlexandre M Bayen, “Efficient kalman filter techniques for the assimilation of tide gauge data in three-dimensional modeling of the north sea and baltic sea system,” *In Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pp. 1783–1790, 2008.
- [66] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Verlag: Springer, 2009.
- [67] ———, “The ensemble kalman filter: Theoretical formulation and practical implementation,” *Ocean dynamics*, vol. 53, no. 4, pp. 343–367, 2013.
- [68] A. Gelb, *Applied Optimal Estimation*. MIT Press, 1974.
- [69] G. Evensen, “Efficient kalman filter techniques for the assimilation of tide gauge data in three-dimensional modeling of the north sea and baltic sea system,” *Journal of Geophysical Research: Oceans (19782012)*, vol. 99, no. C5, pp. 10 143–10 162, 1994.
- [70] L. Nerger, T. Janjic, J. Schroter, and W. Hiller, “A unification of ensemble square root kalman filters,” *Monthly Weather Review*, vol. 140, no. 7, pp. 2335–2345, 2012.
- [71] C. H. B. T. M. H. J. S. W. M.K. Tippett, J. L. Anderson, “Ensemble square root filter*,” *Monthly Weather Review*.
- [72] J. Anderson, “An ensemble adjustment kalman filter for data assimilation,” *Monthly weather review*, vol. 129, no. 12, pp. 2884–2903, 2001.
- [73] A. Aksoy, F. Zhang, and J. Nielsen-Gammon, “Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model,” *Monthly weather review*, vol. 134, no. 10, pp. 2951–2970, 2006.
- [74] J. Annan, J. Hargreaves, N. Edwards, and R. Marsh, “Parameter estimation in an intermediate complexity earth system model using an ensemble kalman filter.” *Ocean Modelling*, vol. 8, no. 1, pp. 135–154, 2005.

- [75] H. H. Franssen and W. Kinzelbach, “Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem,” *Water Resources Research*, vol. 44, no. 9, 2008.
- [76] X. Luo and I. Hoteit, “Robust ensemble filtering and its relation to covariance inflation in the ensemble kalman filter,” *Monthly Weather Reviews*, vol. 139, p. 39383953, 2011.
- [77] R. Madankan, “Polynomial chaos based method for state and parameter estimation.”
- [78] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte-Carlo Methods in Practice*. Springer-Verlag, 2001.
- [79] R. N. Iyengar and P. K. Dash, “Study of the random vibration of nonlinear systems by the gaussian closure technique,” *Journal of Applied Mechanics*, vol. 45, pp. 393–399, 1978.
- [80] J. B. Roberts and P. D. Spanos, *Random Vibration and Statistical Linearization*. Wiley, 1990.
- [81] T. Lefebvre, H. Bruyninckx, and J. Schutter, “Comment on a new method for the nonlinear transformations of means and covariances in filters and estimators,” *IEEE Transactions on Automatic Control*, vol. 48, no. 8, 2002.
- [82] ———, “Kalman filters of non-linear systems: A comparison of performance,” *International journal of Control*, vol. 77, no. 7, pp. 639–653, 2004.
- [83] K. Dalbey, A. K. Patra, E. B. Pitman, M. I. Bursik, and M. Sheridan, “Input uncertainty propagation methods and hazard mapping of geophysical mass flows,” *Journal of Geophysical Research*, vol. 113, pp. 1–16, 2008.
- [84] C. W. Clenshaw and A. R. Curtis, “A method for numerical integration on an automatic computer,” *NUMERISCHE MATHEMATIK*, vol. 2, no. 1, pp. 197–205, 1960.

- [85] E. Cheney and D. Kincaid, *Numerical mathematics and computing*, 5th ed. CA: Brooks/Cole, Pacific Grove, 1999.
- [86] M. Berveiller, B. Sudret, and M. Lemaire, “Stochastic finite element: a non intrusive approach by regression,” *European Journal of Computational Mechanics*, vol. 15, pp. 81–2, 2006.
- [87] I. Babuka, F. Nobile, and R. Tempone, “A stochastic collocation method for elliptic partial differential equations with random input data,” *SIAM Journal of Numerical Analysis*, vol. 45, no. 3, pp. 1005–1034, 2007.
- [88] L. Mathelin and M. Hussaini, “A stochastic collocation algorithm for uncertainty analysis.”
- [89] D. Xiu and J. Hesthaven, “High-order collocation methods for differential equations with random inputs,” *SIAM Journal of Scientific Computation*, vol. 27, no. 3, pp. 1118–1139, 2005.
- [90] S. Smolyak, “Quadrature and interpolation formulas for tensor products of certain classes of functions,” *Doklady Akademii Nauk SSSR*, vol. 4, pp. 240–243, 1963.
- [91] T. Gerstner and M. Griebel, “Numerical integration using sparse grids,” *Numerical Algorithms*, vol. 18, pp. 209–232, 1998.
- [92] K. Petras, “On the smolyak cubature error for analytic functions,” *Advances in Computational Mathematics*, vol. 12, pp. 71–93, 2000.
- [93] —, “Fast calculation in the smolyak algorithm,” *Numerical Algorithms*, vol. 26, pp. 93–109, 2001.
- [94] —, “Smolyak cubature of given polynomial degree with few nodes for increasing dimension.” *Numerical Mathematics*, vol. 93, no. 4, pp. 729–753, 2003.