UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FORECAST UNCERTAINTY QUANTIFICATION USING MONTE CARLO, POLYNOMIAL CHAOS EXPANSION AND UNSCENTED TRANSFORMATION METHODS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

JUNJUN HU
Norman, Oklahoma
2015

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/21722

FORECAST UNCERTAINTY QUANTIFICATION USING MONTE CARLO, POLYNOMIAL CHAOS EXPANSION AND UNSCENTED TRANSFORMATION METHODS

A DISSERTATION APPROVED FOR THE
SCHOOL OF COMPUTER SCIENCE

BY

_____
Dr. S. Lakshmivarahan, Chair

_____
Dr. Meijun Zhu

_____
Dr. Sridhar Radhakrishnan

_____
Dr. Changwook Kim

_____
Dr. Sudarshan Dhall

© Copyright by JUNJUN HU 2015
All Rights Reserved.

Acknowledgements

The completion of my dissertation would never be possible without the guidance of my advisor and the committee members, help from friends, and support from my family.

I would like to express my deepest gratitude to my advisor Prof. S. Lakshmivarahan for his guidance, understanding, patience, encouragement and support during my graduate studies at University of Oklahoma. His mentorship was paramount in providing a well-rounded experience consistent with my long-term career goals. Not only the academic skills he has but also his structured working attitude and constant enthusiasm would affect me for a lifetime. For everything you've done for me, Prof. Varahan, I thank you.

I would also like to gratefully and sincerely thank my coadvisor Dr. John M. Lewis from National Severe Storm Laboratory (NSSL) and Desert Research Institute (DRI). His guidance, patience and encouragement make me confident in doing current research and future career. Special thanks go to him for always sharing his experiences either in academia or life with me, which are inspiring and great treasures to me.

I also owe my gratitude to my dissertation committee members: Professors Sudarshan Dhall, Changwook Kim, Sridhar Radhakrishnan, and Meijun Zhu. They have given me valuable suggestions and comments during the revision of this dissertation. My thanks also go to the other faculty and staff in the School of Computer Science at the University of Oklahoma.

Many thanks go to Dr. Jidong Gao from NSSL and my advisor Prof. S. Lakshmivarahan and Dr. John M. Lewis for their recommendation to the Summer Colloquium on Satellite Data Assimilation held by the NASA/NOAA/DoD Joint Center

for Satellite Data Assimilation (JCSDA) in 2015. It was a great opportunity to learn the latest technologies and applications together with the future on satellite data assimilation. Presentations from the scientists in satellite data assimilation field and discussions with the scientists and students there inspired my mind and stimulated my dissertation preparation.

I would like to give thanks to former and current colleagues in Oklahoma Geological Survey (OGS), Dr. G. Randy Keller, Dr. Kevin D. Crain, Dr. Vikram Jayaram, Shanika L. Wilson, Joyce A. Stiehler, etc. They have given their greatest support and help on my dissertation preparation.

Finally and most importantly, I want to thank my family members, my parents, two brothers, my husband, my son and my daughter. They are always supporting me and encouraging me with their best wishes and greatest love.

Table of Contents

Acknowledgements iv

List of Tables viii

List of Figures xi

Abstract xv

Chapter 1 Introduction 1

Chapter 2 Polynomial Chaos Expansion 8

2.1 The Hermite Polynomial Chaos 9

2.2 The Generalized Polynomial Chaos 10

2.3 Forecast Uncertainty Quantification Using Polynomial Chaos Expansion 13

2.3.1 Stochastic Galerkin Method 16

2.3.2 Stochastic Collocation Method 17

2.4 Discussions 22

Chapter 3 Unscented Transformation 23

3.1 The Basic Unscented Transformation 23

3.2 The Scaled Unscented Transformation 26

3.2.1 The Auxillary Random Variable 27

3.2.2 The Scaled Unscented Transform 27

3.3 Discussions 30

Chapter 4 Application of Stochastic Galerkin Method 31

4.1 The Two-variable Model 31

4.2 Univariate Hermite Polynomial Chaos Expansion 33

4.3 Multivariate Hermite Polynomial Chaos Expansion 38

4.4 Discussions 49

Chapter 5 Application of Stochastic Collocation Method 50

5.1 The Mixed-layer Model 51

5.2 Initial Condition Only 57

5.3 Parameter Only 72

5.4 Discussions 90

Chapter 6 Application of Unscented Transformation Approach 91

6.1 Initial Condition Only 91

6.2 Parameter Only 98

6.3 Discussions 104

Chapter 7 Discussions and Conclusions 105

References 111

Appendix A Hermite Polynomials 121

Appendix B Hermite Polynomial Chaos 128

Appendix C Gaussian Quadrature Rule 131

Appendix D Performance of Unscented Transformation 138

Appendix E Performance of Scaled Unscented Transformation 149

Appendix F Stochastic Galerkin for Mixed-layer Model 155

Appendix G Legendre Polynomials 163

List of Tables

Table 2.1 The correspondence between the random variables and the types of Wiener-Askey polynomial chaos (Xiu and Karniadakis 2002a) 12

Table 4.1 Inner products $\langle H_i H_j, H_0 \rangle$ 35

Table 4.2 Inner products $\langle H_i H_j, H_1 \rangle$ 35

Table 4.3 Inner products $\langle H_i H_j, H_2 \rangle$ 35

Table 4.4 Inner products $\langle H_i H_j, H_3 \rangle$ 36

Table 4.5 Inner products $\langle H_i H_j, H_4 \rangle$ 36

Table 4.6 Inner products $\langle H_i H_j, H_5 \rangle$ 36

Table 4.7 Inner products $\langle H_i H_j, H_6 \rangle$ 37

Table 4.8 Inner products $\langle H_{\mathbf{i}} H_{\mathbf{j}}, H_{0,0} \rangle$ 40

Table 4.9 Inner products $\langle H_{\mathbf{i}} H_{\mathbf{j}} H_{1,0} \rangle$ 40

Table 4.10 Inner products $\langle H_{\mathbf{i}} H_{\mathbf{j}}, H_{0,1} \rangle$ 40

Table 4.11 Inner products $\langle H_{\mathbf{i}} H_{\mathbf{j}}, H_{2,0} \rangle$ 40

Table 4.12 Inner products $\langle H_{\mathbf{i}} H_{\mathbf{j}}, H_{1,1} \rangle$ 41

Table 4.13 Inner products $\langle H_{\mathbf{i}} H_{\mathbf{j}}, H_{0,2} \rangle$ 41

Table 4.14 First moments of PC, Exact and MC at $t = 2$ (two-variable model) 43

Table 4.15 Second moments of PC, Exact and MC at $t = 2$ (two-variable model) 43

Table 4.16 Third moments of PC, Exact and MC at $t = 2$ (two-variable model) 43

Table 4.17 Moments of SG-M2-P2 and Exact at $t = 1$ (two-variable model) 44

Table 4.18 Moments of SG-M2-P2 and Exact at $t = 3$ (two-variable model) 44

Table 4.19 Moments of SG-M2-P2 and Exact at $t = 5$ (two-variable model) 45

Table 4.20 Moments of SG-M2-P2 and Exact at $t = 10$ (two-variable model) 45

Table 5.1 Upper-air Observations at $t = 0$, $t = 6h$ and $t = 9h$ 53

Table 5.2 Mean values and standard deviations for mixed-layer model initial conditions 56

Table 5.3 Mean values and ranges for mixed-layer model parameters 56

Table 5.4 Mean values and standard deviations for mixed-layer model boundary conditions 56

Table 5.5 Five-variable normalized Hermite polynomials (order no greater than 2)..... 58

Table 5.6 Sparse collocation points with weights (dimension 5, exact level 2), Gaussian-Hermite quadrature rule 60

Table 5.7 Sparse collocation points with weights (dimension 5, exact level 3), Gaussian-Hermite quadrature rule 60

Table 5.8 Covariance matrix, (mixed-layer model) IC only, PC vs. MC 66

Table 5.9 Six-variable normalized Legendre polynomials (order no greater than 2)..... 73

Table 5.10 Sparse collocation points with weights (dimension 6, exact level 2), Gauss-Legendre quadrature rule 74

Table 5.11 Sparse collocation points with weights (dimension 6, exact level 3), Gauss-Legendre quadrature rule 74

Table 6.1 Sigma points with weights used in UT, (mixed-layer model) IC only 92

Table 6.2 Covariance matrix, (mixed-layer model) IC only, UT vs. MC 97

Table 6.3 Sigma points with weights used in UT, (mixed-layer model) Parameter only 99

Table 6.4 Covariance matrix, (mixed-layer model) Parameter only, UT vs. MC 103

Table 7.1 Computing time, (mixed-layer model) IC only 108

Table 7.2 Computing time, (mixed-layer model) Parameter only 108

Table A.1 A list of $H_m(x)$, $0 \leq m \leq 4$ 122

Table A.2 Two-variate Hermite polynomials, degree less than or equal to 4 125

Table G.1 A list of $P_{\mathrm{m}}(\mathrm{x})$, $0 \leq \mathrm{m} \leq 6$ 165

Table G.2 Two-variate (n=2) Legendre polynomials, degree less than or equal to 4...167

List of Figures

Figure 4.1 Histogram at $t = 1$ (two-variable model) (a) MC (b) PC..... 46

Figure 4.2 Histogram at $t = 2$ (two-variable model) (a) MC (b) PC..... 46

Figure 4.3 Histogram at $t = 3$ (two-variable model) (a) MC (b) PC..... 46

Figure 4.4 Histogram at $t = 5$ (two-variable model) (a) MC (b) PC..... 47

Figure 4.5 Histogram at $t = 10$ (two-variable model) (a) MC (b) PC 47

Figure 4.6 Evolution of mean values of the amplitude pair derived from PC (two-variable model) (a) μ_1 (b) μ_2 48

Figure 4.7 Evolution of second moments of the amplitude pair derived from PC (two-variable model) (a) σ_1 (b)σ_1 (c) σ_{12} 48

Figure 5.1 A schematic diagram of the idealized mixed layer profiles of potential temperature and mixing ratio where basic variables are identified and symbolized (Lewis et al. 2015). 54

Figure 5.2 Evolution of the mean values, (mixed-layer model) IC only, PC vs. MC 64

Figure 5.3 Evolution of the standard deviations, (mixed-layer model) IC only, PC vs. MC 65

Figure 5.4 Histogram of θ at $t = 1$h, (mixed-layer model) IC only, (a) MC (b) PC.... 68

Figure 5.5 Histogram of h at $t = 1$h, (mixed-layer model) IC only, (a) MC (b) PC.... 68

Figure 5.6 Histogram of q at $t = 1$h, (mixed-layer model) IC only, (a) MC (b) PC.... 68

Figure 5.7 Histogram of θ at $t = 24$h, (mixed-layer model) IC only, (a) MC (b) PC . 69

Figure 5.8 Histogram of h at $t = 24$h, (mixed-layer model) IC only, (a) MC (b) PC . 69

Figure 5.9 Histogram of q at $t = 24$h, (mixed-layer model) IC only, (a) MC (b) PC . 69

Figure 5.10 Histogram of θ at $t = 48$h, (mixed-layer model) IC only, (a) MC (b) PC 70

Figure 5.11 Histogram of h at $t = 48$h, (mixed-layer model) IC only, (a) MC (b) PC 70

Figure 5.12 Histogram of q at $t = 48$h, (mixed-layer model) IC only, (a) MC (b) PC 70

Figure 5.13 The simulation of base state by PC, (mixed-layer model) IC only..... 71

Figure 5.14 Evolution of mean values, (mixed-layer model) Parameter only, PC (exact level 2) vs. MC 78

Figure 5.15 Evolution of standard deviations, (mixed-layer model) Parameter only, PC (exact level 2) vs. MC 79

Figure 5.16 Histogram of θ at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2) 80

Figure 5.17 Histogram of h at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2) 80

Figure 5.18 Histogram of q at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2) 80

Figure 5.19 Histogram of θ at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)..... 81

Figure 5.20 Histogram of h at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)..... 81

Figure 5.21 Histogram of q at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)..... 81

Figure 5.22 Histogram of θ at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)..... 82

Figure 5.23 Histogram of h at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)..... 82

Figure 5.24 Histogram of q at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2) 82

Figure 5.25 Evolution of mean values, (mixed-layer model) Parameter only, PC (exact level 3) vs. MC 84

Figure 5.26 Evolution of standard deviations, (mixed-layer model) Parameter only, PC (exact level 3) vs. MC 85

Figure 5.27 Histogram of θ at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 86

Figure 5.28 Histogram of h at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 86

Figure 5.29 Histogram of q at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 86

Figure 5.30 Histogram of θ at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 87

Figure 5.31 Histogram of h at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 87

Figure 5.32 Histogram of q at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 87

Figure 5.33 Histogram of θ at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 88

Figure 5.34 Histogram of h at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3) 88

Figure 5.35 Histogram of q at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)..... 88

Figure 5.36 The simulation of base state by PC (exact level 3), (mixed-layer model) Parameter only..... 89

Figure 6.1 Ensemble forecast using UT, (mixed-layer model) IC only 93

Figure 6.2 Evolution of mean values, (mixed-layer model) IC only, UT vs. MC 95

Figure 6.3 Evolution of standard deviations, (mixed-layer model) IC only, UT vs. MC 96

Figure 6.4 Ensemble forecast by UT, (mixed-layer model) Parameter only..... 100

Figure 6.5 Evolution of mean values, (mixed-layer model) Parameter only, UT vs. MC 101

Figure 6.6 Evolution of standard deviations, (mixed-layer model) Parameter only, UT vs. MC 102

Abstract

In the context of prediction science, the sources of uncertainty can be from the uncertainties of the experiments, modeling, model inputs, numerical analysis, etc. This study concentrates on quantifying the forecast uncertainty arising from the propagation of the uncertainties in the model inputs to the dynamical model. The uncertainties in the inputs include the randomness in (1) the initial conditions, (2) the forcing term (including both the external forcing and the boundary conditions), and (3) randomness in the parameters of the model. In order to quantify the uncertainties in the forecast, three uncertainty quantification (UQ) methods are studied, namely classical Monte Carlo (MC), polynomial chaos (PC) expansion and unscented transformation (UT). Using MC as the benchmark, two dynamical models are used in this study to examine the performance of PC expansion and UT. One is the low order (two components) spectral solution to the nonlinear advection equation, and the other one is the five-variable mixed-layer model which is used to describe the return flow event over the Gulf of Mexico during the cool season (between November and March) every year. The experimental results and the comparisons with MC have shown that both PC and UT can provide good estimates on the statistical information relating to the forecast, for example, the mean, variation (or standard deviation), covariance. The approach of UT utilizes a set of deterministically chosen sigma points to propagate the uncertainties contained in the inputs through the dynamical model. Only the first two moments of the forecast can be estimated by UT. Different from UT, the PC expansion represents the stochastic process in the form of a series expression (hence a surrogate approximation) in terms of the orthogonal polynomials whose type depends on the probability

distribution of the random inputs. Ensemble forecast can be achieved by sampling the random variable used in the PC expansion. Furthermore, the histogram of the forecast can be constructed using the ensemble forecast, and then one can estimate the probability density function (PDF) of the forecast. What's more, PC expansion can also give estimates on the statistics of higher order moments. The application of PC and UT in quantifying the forecast uncertainties in large scale system, the combination with data assimilation techniques and its real applications, and the ability to deal with nonGaussian distributions will be some of the topics for future study.

Chapter 1

Introduction

Uncertainty Quantification (UQ) is a growing field motivated by the synthesis of modeling, large-scale simulations, numerical analysis and experiments together with the science of probability and statistics. In the meantime, it is also as old as the disciplines of probability and statistics. In the context of predictive science, UQ is defined as the science of identifying, quantifying and reducing uncertainties associated with modeling, numerical algorithms, experiments and outcomes of the prediction (Smith 2013). It tries to determine how likely certain outcomes are if some aspects of the system are not exactly known. Smith (2013) uses five large-scale applications to show that the predictions with quantified uncertainties are critical to understand and predict certain physical phenomena as well as make decisions and designs based on the predictions. The five applications are weather models, climate models, subsurface hydrology and geology models, nuclear reactor designs and models for biological phenomena. Uncertainties can arise from different sources, some of which are listed below:

(1) Experiment uncertainties and limitations: in general, limited or incomplete data and limited sensor accuracy or resolution are two fundamental resources of uncertainties related to the experiment;

(2) Model and input uncertainties: model uncertainties are the errors or discrepancies that are induced by the approximation or imprecise presentation of the process to be studied. As is known that models usually have some parameters or some models (e.g., models described by differential equations) may have initial conditions, boundary conditions or sometimes external forcing and uncertainties exist in these

inputs. Moreover, uncertainties become more complex when using coupled system to quantify multiscale or multiphysics phenomena;

(3) Numerical errors and uncertainties: these are the uncertainties or errors related to numerical solution or algorithms. Some examples are errors in roundoff or discretization, etc.

This study mainly focuses on the uncertainties that lie in the model inputs. And the aim of this study is to quantify the forecast uncertainty resulting from the input uncertainties being propagated through a dynamical forecast model, which is like uncertainty propagation. As stated above, the input uncertainties of a dynamical forecast model may include the randomness in the initial conditions, the forcing term (including both the external forcing and the boundary conditions), and randomness in the parameters of the model. In each of these cases, the solution of the model is a stochastic process.

Generally, uncertainty quantification methods are classified into two groups: sampling-based and non-sampling techniques. The classical Monte Carlo (MC) and its variants, e.g., quasi-Monte Carlo (Fox 1999; Niederreiter 1992; Niederreiter et al. 1998) are the most commonly used sampling methods. In MC, independent samples (or realizations) of the random inputs are generated according to their prescribed probability distribution. Then for each realization, the problem becomes deterministic. After solving all realizations of the problem, an ensemble of the solutions are collected, which is called an ensemble forecast. The statistical information of the forecast (e.g., mean, variance, covariance, etc.) can then be calculated from the ensemble forecast. Although it is straightforward to apply MC, the statistics of the forecast converges

slowly. Therefore, typically a large number of executions are required, which means excessive computational burden especially for complex system that demands expensive computational resources even in its deterministic settings. Though the rapid development of the computation technology has released the burden to some extent, further efforts are still needed.

The unscented transformation (UT) method is another example of sampling-based approach. It is developed based on the intuition that the approximation of a probability distribution is easier than that of an arbitrary nonlinear function or transformation (Julier et al. 1995; Julier and Uhlmann 1996, 1997a, 1997b). It uses a set of samples (called sigma points) to approximate the probability distribution of the random inputs and then propagates these samples through the nonlinear transformation (the dynamical model in this context), the statistics are computed afterwards using the transformed sigma points. In spite of using similar idea as MC, different selection strategy of the sigma points is used. The number of the sigma points in UT depends on the dimension of the random inputs and the points are generated through deterministic formulas, whereas the samples for MC are randomly chosen.

In the non-sampling group, one representative is the polynomial chaos (PC) expansion. The earliest PC approach is the Wiener's (1938) polynomial chaos (PC) expansion. It is known that the solution of a model with random inputs is a stochastic process. The Wiener's PC expansion of the stochastic process is done by expressing the (unknown) solution of the model in an orthogonal expansion using a stochastic basis consisting of the set of all Hermite polynomials of the standard Gaussian random variable whose distribution is defined over the real line, where the coefficients (or the

strength of the modes) of the expansion are (unknown) deterministic functions of time. By exploiting the orthogonality property of the Hermite polynomial (with respect to the standard Gaussian as the weight function), the given model is reduced to a system of coupled nonlinear dynamics on the deterministic coefficient functions. By solving this reduced spectral dynamics numerically, one can then effectively reconstruct the stochastic solution of the original forecast model, based on which the probabilistic characterization of the model forecast can be provided. While this approach is quite similar in principle to the well-known Karhunen–Loève (K-L) expansion (Loève 1977), there is a major difference in the choice of the stochastic basis. In K-L expansion, the stochastic basis consists of the eigen functions of the known correlation function of the underlying stochastic process. In the case where the (stochastic) solution of the forecast model is unknown, let alone its underlying correlation structure, the more general approach based on Wiener's PC expansion can be used. However, like everything else in life, there is a price to pay for this lack of knowledge about the solution, namely, the solution based on K-L expansion is inherently optimal but the solution based on the Wiener's PC does not share this inherent optimality property (Loève 1977; Ghanem and Spanos 1991).

A succinct account of the role of Wiener's PC based approach in stochastic analysis is given in (Kallianpur 1980) and (Kuo 2006). Lototsky and Rozovskii (2006) develop a general framework for solving stochastic differential equations (Arnold 1974) using PC approach. Solution to the nonlinear filter (which is a general form of dynamic data assimilation for stochastic models) based on PC is developed in (Lototsky 2011). Mathematical generalization of Wiener's PC to include Askey scheme (called gPC) is

developed in (Xiu and Karniadakis 2002a). The monograph by Xiu (2010) contains an elegant presentation of PC, gPC and their applications.

Earliest application of Wiener's PC based approach to quantify uncertainty in engineering problems is due to Ghanem and Spanos (1991). Since then there is a virtual explosion of literature in this area. The review paper by Ghanem (1999) provides a very good presentation of the PC methodology and a roadmap for applications. Two recent books by Le Maitre and Knio (2010) and Grigoriu (2012) provide excellent presentation of both the theory of PC and its multi-faceted applications.

A note on the other non-sampling methods for quantifying the forecast uncertainty is in order. If the forecast uncertainty is only due to those in the initial condition, then the well-known partial differential equation known as the Liouville's equation (Saaty 1967) provides the complete solution by describing the evolution of the probability density function of the forecast with time. If the uncertainty in the forecast arises from two sources – those in the initial condition and in forcing, then the evolution of the probability density of the forecast is given by the celebrated Kolmogorov's forward equation (Jazwinski 1970). Soong (1973) describes several special methods to handle the uncertainty in the parameters in an otherwise deterministic model. But, when the uncertainty arises from all the three sources (initial condition, forcing and parameters), as is considered in this chapter, to my knowledge, the Wiener's polynomial chaos and its generalization are the only known approaches to quantify the model forecast uncertainty.

The related theory of nonlinear and non-Gaussian dynamic data assimilation is embodied in the contemporary theory of nonlinear filtering that deals with combining

an uncertain nonlinear model forecast with noisy (nonlinear) observations in a Bayesian framework (Crisan and Rozovskii 2011). In this case, the evolution of the posterior density that describes the evolution of the uncertainty in the analysis is given by the well-known Kushner-Zakai equation, which is a stochastic partial differential equation (Kushner 1962; Zakai 1969). There is a natural nesting between the three well-known classes of partial differential equations mentioned above in the sense Kushner-Zakai becomes Kolmogorov's forward equation when there is no noisy observation and the latter in turn becomes Liouville's equation when there is no random forcing. Notice that this well-known hierarchy does not handle uncertainty in parameters.

This study aims to study the effectiveness and efficiency of UT and PC methods in quantifying the uncertainty of the forecast due to random inputs. The ensemble forecast and the statistical information from the ensemble forecast using the classical Monte Carlo (MC) approach is the benchmark. Therefore, the performance of UT and PC are compared with MC in this research. Two meteorological models are used to study the methods in detail. One is the low order (two components) spectral solution to the nonlinear advection equation found in (Platzmon 1964). This two-variable model is used to demonstrate using the stochastic Galerkin projection to obtain the expansion coefficients in PC expansion and only random initial conditions are considered. And the other one is a five-variable mixed-layer model which is used to describe a large-scale process termed as "return flow" by Keith Henry, a professor of meteorology at Texas A & M University (Henry 1979a, 1979b). In the late fall and winter, a rhythmic cycle of cold air penetrates into the Gulf of Mexico (GoM), these penetrations are then generally followed by return of modified air to land in response to circulation around an eastward-

moving cold anticyclone. Typically, 4–5 of these return-flow events (RFE's) occur per month between November and March each year (Crisp and Lewis 1992). The latter one is used to show using stochastic Collocation to obtain expansion coefficients in PC expansion and the application of UT. Uncertainty in the forecast arising from both random initial condition and random parameters are considered in this example.

The research is organized as follows: In Chapters 2 and 3, the mathematical background of polynomial chaos expansion and unscented transformation methods are presented respectively. The stochastic Galerkin approach using the two-variable model is demonstrated in Chapter 4. The description of the mixed-layer model and the application of PC with stochastic Collocation are followed in Chapter 5. In Chapter 6, UT is studied in the mixed-layer model. Finally in Chapter 7, PC and UT are compared with MC in quantifying the forecast uncertainty. A detailed discussion of these three methods and some conclusions are provided.

Chapter 2

Polynomial Chaos Expansion

As one of the most widely used non-sampling techniques, the polynomial chaos (PC) expansion approach offers a means of computing high-order information such as the mean, variance, and successive moments if the probability density function (PDF) of the input variable is well defined. The PC approach originates from Wiener's homogeneous chaos theory (Wiener 1938). Ghanem and Spanos (1991) demonstrated that PC is a feasible computational tool for scientific and engineering studies. Xiu and Karniadakis (2002a) then expanded the work by Ghanem and Spanos for Hermite-chaos expansion (Ghanem and Spanos 1991) and the work by Ogura for Charlier-chaos expansion (Ogura 1972), generalized the concept and proposed the generalized polynomial chaos (gPC) expansion or Wiener-Askey polynomial chaos expansion by setting the expansion basis as the orthogonal polynomials from the Askey-scheme class. They further demonstrated that the Wiener-Askey polynomial chaos exhibit exponential convergence rate when the optimal polynomial expansion is chosen according to the probability distribution of the random input. If the optimal polynomial chaos is not chosen, convergence can be assured but the exponential rate is not retained for any given type of random input.

This chapter first introduces the Hermite polynomial chaos expansion and the generalized polynomial chaos expansion of a random process. Then the framework of using polynomial chaos expansion to quantify the forecast uncertainty of a dynamical system is presented. To obtain the expansion coefficients in polynomial chaos

expansion, two approaches namely stochastic Galerkin (SG) and stochastic Collocation (SC) are illustrated in detail.

2.1 The Hermite Polynomial Chaos

The concept of homogeneous chaos expansion was originally proposed by Wiener (1938), in which the Hermite polynomials in terms of Gaussian random variables were employed. As homogeneous chaos expansion can approximate any functions in $L_2(C)$ and converges in the $L_2(C)$ sense (Cameron and Martin 1947), Hermite polynomial chaos can be adopted to expand second-order random process in terms of orthogonal polynomials. A second-order random process is a process with finite variance. Most physical processes are second-order random processes. Let $\mathbf{x}\epsilon R^n$ be a general second-order random process, it therefore can be represented as

$$\begin{aligned}\mathbf{x} &= \mathbf{u}_0 H_0 \\ &+ \sum_{i_1=1}^{\infty} \mathbf{u}_{i_1} H_1\left(\zeta_{i_1}\right) \\ &+ \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \mathbf{u}_{i_1 i_2} H_2\left(\zeta_{i_1}, \zeta_{i_2}\right) \\ &+ \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} \mathbf{u}_{i_1 i_2 i_3} H_3\left(\zeta_{i_1}, \zeta_{i_2}, \zeta_{i_3}\right) \\ &+ \cdots, \end{aligned} \tag{2.1}$$

where $H_N\left(\zeta_{i_1}, \zeta_{i_2} \ldots \zeta_{i_N}\right)$ denotes the Hermite polynomial of order N in terms of the multi-dimensional standard Gaussian variable $\boldsymbol{\zeta} = \left(\zeta_{i_1}, \zeta_{i_2} \ldots \zeta_{i_N}\right)$ with zero mean and unit variance. $\boldsymbol{\zeta}$ is a vector consisting of N independent Gaussian variables $\zeta_{i_1}, \zeta_{i_2} \ldots \zeta_{i_N}$. The above equation (2.1) is the discrete version of the original Wiener polynomial chaos expansion. For the continuous case, the summations in (2.1) will be replaced by the continuous integrals.

For notational convenience, (2.1) can be rewritten as

$$\mathbf{x} = \sum_{|\mathbf{i}|=0}^{\infty} \mathbf{u}_{\mathbf{i}} H_{\mathbf{i}}(\boldsymbol{\zeta}), \tag{2.2}$$

where $H_{\mathbf{i}}(\boldsymbol{\zeta}) = H_{p_1 p_2 \dots p_N}(\boldsymbol{\zeta})$ is a N-variate homogeneous Hermite polynomial. $\mathbf{i} = (p_1, p_2, \dots, p_N)$ is defined as a multi-index. The order of $H_{\mathbf{i}}(\boldsymbol{\zeta})$ is defined as $m = |\mathbf{i}| = p_1 + p_2 + \cdots p_N$ and $(p_1, p_2, \dots, p_N)$ is called an additive partition for m. For example, when $N = 2$, there are 3 partitions, $\{(2,0), (1,1), (0,2)\}$ for the order $m = 2$. The general form of N-variate homogeneous Hermite polynomials of degree m with partition $(p_1, p_2, \dots, p_N)$ is given by

$$\mathrm{H}_{p_1 p_2 \dots p_N}(\boldsymbol{\zeta}) = (-1)^m e^{\frac{\boldsymbol{\zeta}^T \boldsymbol{\zeta}}{2}} \frac{\partial^m}{\partial {\zeta_1}^{p_1} \partial {\zeta_2}^{p_2} \dots \partial {\zeta_N}^{p_N}} e^{-\frac{\boldsymbol{\zeta}^T \boldsymbol{\zeta}}{2}}. \tag{2.3}$$

Define the inner product in the Hilbert space as

$$\langle f(\boldsymbol{\zeta}) g(\boldsymbol{\zeta}) \rangle = \int f(\boldsymbol{\zeta}) g(\boldsymbol{\zeta}) W(\boldsymbol{\zeta}) d\boldsymbol{\zeta}, \tag{2.4}$$

with $W(\boldsymbol{\zeta})$ as the Gaussian probability density function, i.e.,

$$W(\boldsymbol{\zeta}) = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{\boldsymbol{\zeta}^T \boldsymbol{\zeta}}{2}}. \tag{2.5}$$

The Hermite polynomial basis $\{H_{\mathbf{i}}(\boldsymbol{\zeta})\}$ form a complete orthogonal polynomial basis, i.e.,

$$\langle H_{\mathbf{i}} H_{\mathbf{j}} \rangle = \langle H_{\mathbf{i}}^2 \rangle \delta_{\mathbf{ij}}, \tag{2.6}$$

where $\delta_{\mathbf{ij}}$ is the Kronecker delta function, i.e., $\delta_{\mathbf{ij}} = 0$, when $\mathbf{i} \neq \mathbf{j}$, otherwise 1. Here, $\mathbf{i} = \mathbf{j}$ means $\mathbf{i}_k = \mathbf{j}_k$ for every $k = 1, 2, \dots N$.

More details on Hermite polynomials in univariate and multivariate case can be found in Appendices A and B.

2.2 The Generalized Polynomial Chaos

Even though the Hermite polynomial chaos expansion is effective in solving stochastic differential equation with Gaussian random inputs as well as some certain types of non-

Gaussian inputs (Spanos and Ghanem 1989; Ghanem and Spanos 1991; Ghanem 1999; Xiu and Karniadakis 2003), the optimal convergence rate for general non-Gaussian random input is not obtained. In order to solve this problem, Xiu and Karniadakis (2002a) proposed the generalized polynomial chaos (gPC) expansion or Wiener-Askey polynomial chaos expansion by setting the expansion basis as the orthogonal polynomials from the Askey class. They have shown that the exponential convergence rate will be achieved if the optimal polynomials are selected in conjunction with the random distribution. Table 2.1 gives the correspondence between the random variables and the types of Wiener-Askey polynomial chaos.

As stated before, the idea of gPC approach originates from the Wiener's Hermite polynomial chaos expansion. The key idea of gPC approach is centered on the orthogonality of polynomials with respect to an inner product definition associated with a suitable weighting function. The PC approach lies in the fact that a second-order random process can be expressed as a series in terms of orthogonal polynomials (Xiu 2010). That is, the random process $\mathbf{x}$ can be expressed in terms of gPC expansion as follows:

$$\begin{aligned}\mathbf{x} = {} & \mathbf{v}_0\Phi_0 \\ & + \sum_{i_1=1}^{\infty} \mathbf{v}_{i_1}\Phi_1\left(\varsigma_{i_1}\right) \\ & + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} \mathbf{v}_{i_1 i_2}\Phi_2\left(\varsigma_{i_1},\varsigma_{i_2}\right) \\ & + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} \mathbf{v}_{i_1 i_2 i_3}\Phi_3\left(\varsigma_{i_1},\varsigma_{i_2},\varsigma_{i_3}\right) \\ & + \cdots, \end{aligned} \tag{2.7}$$

where $\Phi_N\left(\varsigma_{i_1},\varsigma_{i_2}\ldots\varsigma_{i_N}\right)$ denotes the Wiener-Askey polynomial chaos of order N in terms of the multi-dimensional random variable $\boldsymbol{\varsigma} = \left(\varsigma_{i_1},\varsigma_{i_2}\ldots\varsigma_{i_N}\right)$ with certain

distribution. Here, $\Phi_N(\varsigma_{i_1}, \varsigma_{i_2} \dots \varsigma_{i_N})$ is not restricted to Hermite polynomial chaos, but rather any type of polynomials from the Askey scheme. The paper presented by Xiu and Karniadakis (2002a) gives more details about Wiener-Askey polynomials chaos.

Table 2.1 The correspondence between the random variables and the types of Wiener-Askey polynomial chaos (Xiu and Karniadakis 2002a)

	Random variable	Wiener-Askey chaos	Support
Continuous	Gaussian	Hermite-Chaos	$(-\infty, \infty)$
	Gamma	Laguerre-Chaos	$[0, \infty)$
	Beta	Jacobi-Chaos	[a, b]
	Uniform	Legendre-Chaos	[a, b]
Discrete	Poisson	Charlier-Chaos	$\{0, 1, 2, \dots\}$
	Binomial	Krawtchouk-Chaos	$\{0, 1, \dots, N\}$
	Negative Binomial	Meixner-Chaos	$\{0, 1, 2, \dots\}$
	hypergeometric	Hahn-Chaos	$\{0, 1, \dots, N\}$

Likewise, equation (2.7) can be rewritten as

$$\mathbf{x} = \sum_{|\mathbf{i}|=0}^{\infty} \mathbf{v_i} \Phi_{\mathbf{i}}(\boldsymbol{\varsigma}). \tag{2.8}$$

The definition of the inner product follows the inner product in the Hilbert space supported by the random variable $\boldsymbol{\varsigma}$,

$$\langle f(\boldsymbol{\varsigma}) g(\boldsymbol{\varsigma}) \rangle = \int f(\boldsymbol{\varsigma}) g(\boldsymbol{\varsigma}) W(\boldsymbol{\varsigma}) d\boldsymbol{\varsigma}, \tag{2.9}$$

or in discrete case,

$$\langle f(\boldsymbol{\varsigma}) g(\boldsymbol{\varsigma}) \rangle = \sum_{\varsigma} f(\boldsymbol{\varsigma}) g(\boldsymbol{\varsigma}) W(\boldsymbol{\varsigma}), \tag{2.10}$$

where $W(\boldsymbol{\varsigma})$ is the weighting function. Usually for those orthogonal polynomials from the Wiener-Askey scheme, the weighting function is the same as the probability density function of corresponding distribution. More details can be found in (Xiu and Karniadakis 2002a).

Based on the inner product defined in equations (2.9) and (2.10), the Wiener-Askey polynomial basis $\{\Phi_{\mathbf{i}}(\boldsymbol{\varsigma})\}$ also have the orthogonality property, i.e.,

$$\langle \Phi_{\mathbf{i}} \Phi_{\mathbf{j}} \rangle = \langle \Phi_{\mathbf{i}}^2 \rangle \delta_{\mathbf{ij}}, \tag{2.11}$$

where $\delta_{\mathbf{ij}}$ is the Kronecker delta function, i.e., $\delta_{\mathbf{ij}} = 0$, when $\mathbf{i} \neq \mathbf{j}$, otherwise 1.

2.3 Forecast Uncertainty Quantification Using Polynomial Chaos Expansion

There are three main input uncertainty sources that can contribute to the uncertainty in the forecast from a dynamical model, i.e., the randomness in initial condition, the random forcing term (including both the external forcing and the boundary conditions), and the randomness exist in the parameters of the model. In this subsection, the application of polynomial chaos expansion approach for quantifying the forecast uncertainty based on a dynamic nonlinear forecast model is discussed.

Let $\mathbf{x}(k)$ denote the state of a discretized model at time $k \geq 0$ defined by a nonlinear stochastic difference equation

$$\mathbf{x}(k+1) = \mathbf{M}[(\mathbf{x}(k), \boldsymbol{\alpha})] + \mathbf{w}(k), \tag{2.12}$$

where $\mathbf{M}: R^n \times R^p \to R^n$ is the one-step state transition map or simply model map, $\mathbf{w}(k)$ is the random noise representing the model error, $\boldsymbol{\alpha} \in R^p$ is a random parameter vector, and $\mathbf{x}(0)$ is the random initial condition. It is further assumed that $\mathbf{x}(0)$, $\boldsymbol{\alpha}$ and $\mathbf{w}(k)$ are stochastically independent.

As discussed before, the random process $\mathbf{x}(k)$ can be represented in terms of orthogonal polynomial basis as

$$\mathbf{x}(k) = \sum_{|\mathbf{i}|=0}^{\infty} \mathbf{v}_{\mathbf{i}}(k) \Phi_{\mathbf{i}}(\boldsymbol{\xi}), \tag{2.13}$$

where $\boldsymbol{\xi} = \{\xi_1, \xi_2, \dots, \xi_N\}$ is a vector consisting of N independent random variables $\xi_1, \xi_2, \dots, \xi_N$, and $\{\Phi_{\mathbf{i}}(\boldsymbol{\xi})\}$ are orthogonal polynomial basis selected from Askey scheme according to the distribution of the random input. $\mathbf{i} = (p_1, p_2, \dots, p_N)$ is a multi-index with $|\mathbf{i}| = p_1 + p_2 + \cdots p_N$ as the degree (or order) of the corresponding polynomial

$\Phi_{\mathbf{i}}(\boldsymbol{\xi})$, $(p_1, p_2, \dots, p_n)$ is called an additive partition for the degree $|\mathbf{i}|$, and $\{\mathbf{v}_{\mathbf{i}}(k)\}$ are called expansion coefficients at time k.

Based on the assumption that the components $\xi_1, \xi_2, \dots, \xi_N$ of the random variable $\boldsymbol{\xi}$ are independent with each other, the N-variate orthogonal polynomial basis $\{\Phi_{\mathbf{i}}(\boldsymbol{\xi})\}$ are constructed as the products of N univariate polynomials $\phi_{i_j}(\xi_j), j = 1, \dots, N$, i.e.,

$$\Phi_{\mathbf{i}}(\boldsymbol{\xi}) = \prod_{j=1}^{N} \phi_{i_j}(\xi_j), \tag{2.14}$$

Here, $\{\phi_{i_j}(\xi_j)\}$ are the i_jth-order orthogonal polynomial basis in ξ_j dimension, which satisfy the following property of orthogonality with $w^{(j)}(\xi_j)$ the weighting function for random variable ξ_j:

$$\langle \phi_m(\xi_j)\phi_n(\xi_j)\rangle = \int \phi_m(\xi_j)\phi_n(\xi_j)w^{(j)}(\xi_j)d\xi_j = \langle \phi_m^2(\xi_j)\rangle\, \delta_{mn}, m, n \geq 0, \tag{2.15}$$

where δ_{mn} is the Kronecker delta function.

According to (2.15), the orthogonality of $\{\Phi_{\mathbf{i}}(\boldsymbol{\xi})\}$ can be obtained:

$$\langle \Phi_{\mathbf{i}}(\boldsymbol{\xi})\Phi_{\mathbf{j}}(\boldsymbol{\xi})\rangle = \int \Phi_{\mathbf{i}}(\boldsymbol{\xi})\Phi_{\mathbf{j}}(\boldsymbol{\xi})W(\boldsymbol{\xi})d\boldsymbol{\xi} = \langle \Phi_{\mathbf{i}}^2(\boldsymbol{\xi})\rangle\, \delta_{\mathbf{ij}}, \tag{2.16}$$

where $W(\boldsymbol{\xi})$ is the weighting function of random vector $\boldsymbol{\xi}$ constructed as the tensor product of the weighting function for each random variable $\xi_j, j = 1,2, \dots, N$, and

$$\delta_{\mathbf{ij}} = \prod_{m=1}^{N} \delta_{i_m j_m}. \tag{2.17}$$

For computational convenience, approximation is made by truncating the expansion (Xiu 2009) in (2.13), e.g., the M-th order PC approximation for $\mathbf{x}(k)$ is expressed as

$$\mathbf{x}_N^M(k) = \sum_{|\mathbf{i}|=0}^{M} \mathbf{v}_{\mathbf{i}}(k)\Phi_{\mathbf{i}}(\boldsymbol{\xi}). \tag{2.18}$$

Likewise, the parameter and random forcing term are expressed as PC expansion in terms of the random variable $\boldsymbol{\xi}$. Hu et al. (2015) gives an introduction on uncertainty quantification by Wiener's polynomial chaos expansion with single random variable.

Let the $M\text{-}th$ order PC approximation for the parameter $\mathbf{a}$ is expressed as

$$\mathbf{a}_N^M = \sum_{|\mathbf{i}|=0}^{M} \mathbf{a}_\mathbf{i} \Phi_\mathbf{i}(\boldsymbol{\xi}). \tag{2.19}$$

The $M\text{-}th$ order PC approximation for the forcing term $\mathbf{w}(k)$ is expressed as

$$\mathbf{w}_N^M(k) = \sum_{|\mathbf{i}|=0}^{M} \mathbf{u}_\mathbf{i}(k) \Phi_\mathbf{i}(\boldsymbol{\xi}). \tag{2.20}$$

The total number of polynomial chaos terms of $\mathbf{x}_N^M(k)$ is $\binom{M+N}{N}$.

According to the classical approximation theory, approximations given in (2.18)-(2.20) are the best approximations in $\mathcal{B}_N^M$, the linear space of N-variate polynomials of degree up to M in the mean-square sense.

When a sufficiently accurate PC approximation is available, all statistical information can be obtained in a straightforward manner. For example, the mean and covariance for $\mathbf{x}(k)$ are

$$E[\mathbf{x}(k)] = \mathbf{v_0}(k), P[\mathbf{x}(k)] = \sum_{|\mathbf{i}|=1}^{M} \mathbf{v}_\mathbf{i}(k)[\mathbf{v}_\mathbf{i}(k)]^T, \tag{2.21}$$

when the polynomials are normalized by their norms.

Estimation of the expansion coefficients is the vital part of using polynomial chaos approach. Ideally, the coefficients $\mathbf{v}_\mathbf{i}(k)$ can be obtained through an orthogonal projection,

$$\mathbf{v}_\mathbf{i}(k) = E[\mathbf{x}(k, \boldsymbol{\xi}) \Phi_\mathbf{i}(\boldsymbol{\xi})] = \int \mathbf{x}(k, \boldsymbol{\xi}) \Phi_\mathbf{i}(\boldsymbol{\xi}) W(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{2.22}$$

However, in practice, the projection in (2.22) is not available because of the lack of the knowledge of the solution $\mathbf{x}(k, \boldsymbol{\xi})$. The stochastic Galerkin (SG) method and the stochastic Collocation (SC) method are two typical methods to numerically approximate the expansion coefficients, more details about these two methods are given in the following paragraphs:

2.3.1 Stochastic Galerkin Method

The basic idea of stochastic Galerkin method is to seek approximations in the form of (2.18)–(2.20) which satisfy the model in (2.12) in a weak form, that is

$$\mathbf{x}_N^M(k+1) = M[(\mathbf{x}_N^M(k), \boldsymbol{\alpha}_N^M)] + \mathbf{w}_N^M(k). \tag{2.23}$$

Expanding (2.18)–(2.20), the resulted equation is a function of the coefficients and the polynomials. Galerkin projection of each polynomial $\Phi_{\mathbf{i}}(\boldsymbol{\xi})$ is then applied to the left and right sides of the resulted function. After calculation of the inner product defined in (2.16), a set of equations will be obtained. The resulting equations are usually a set of coupled deterministic functions for the coefficients $\mathbf{v_i}$, $\mathbf{a_i}$, and $\mathbf{u_i}$. Standard numerical techniques can be applied to solve these equations. SG method converges with the increased value of order (Xiu and Tartakovsky 2006).

To illustrate the details of SG method, a simple scalar example is used. The model is given by a scalar $(n = 1)$ ordinary differential equation (ODE)

$$\frac{dx}{dt} = ax, t > 0,$$

$$x(0) = x_0, \tag{2.24}$$

where the parameter a is assumed to be a random variable with certain distribution, and x_0 is the initial condition. Even though the PC expansion is introduced in a discretized time dependent dynamic model in previous sections, nonetheless the framework works equally well in continuous time dependent dynamic model.

The solution of x and parameter a are expressed by using gPC expansion as follows,

$$x(t) = \sum_{i=0}^{M} v_i(t)\,\phi_i(\xi), \quad a = \sum_{i=0}^{M} a_i\,\phi_i(\xi). \tag{2.25}$$

These expressions are then substituted into the model equation (2.24),

$$\sum_{i=0}^{M} \frac{dv_i(t)}{dt}\phi_i(\xi) = [\sum_{i=0}^{M} a_i\,\phi_i(\xi)][\sum_{i=0}^{M} v_i(t)\,\phi_i(\xi)], \tag{2.26}$$

which then can be simplified as

$$\sum_{i=0}^{M} \frac{dv_i(t)}{dt} \phi_i = \sum_{i=0}^{M} \sum_{j=0}^{M} a_i \, v_j(t) \, \phi_i \phi_j, \tag{2.27}$$

where the random variable ξ is ignored just for the sake of clarity.

Then Galerkin projection is applied, i.e., taking inner product with each polynomial basis $\phi_i, i = 0, \dots, M$ to both sides of (2.27), according to the orthogonality of the polynomial basis, a set of coupled equations will be obtained

$$\frac{dv_k(t)}{dt} = \sum_{i=0}^{M} \sum_{j=0}^{M} e_{ijk} a_i \, v_j(t) \, , k = 0, \dots, M, \tag{2.28}$$

where $e_{ijk} = \langle \phi_i \phi_j, \phi_k \rangle$.

The remaining task is to solve the coupled ODE functions with the initial condition x_0.

Though Galerkin method is effective and has been adopted in various applications (Ghanem and Spanos 1991; Xiu and Karniadakis 2002b, 2003; Babuska et al. 2004; Le Maıtre et al. 2004; Frauenfelder et al. 2005), there are some limitations to SG method. From the implementation perspective, the process of deriving gPC equations is sometimes tedious and challenging. When the governing equations take complicated forms, e.g., highly complex and nonlinear equations, it is difficult to derive the explicit equations for the gPC coefficients. To overcome the disadvantage of SG method, the stochastic Collocation (SC) method was investigated (Xiu 2007). The comparison of SG and SC in the context of complex dynamic system can be found in (Xiu 2009).

2.3.2 Stochastic Collocation Method

Stochastic Collocation method acts like sampling method. In SC methods, one attempts to find solutions $\mathbf{x}\left(k, \boldsymbol{\xi}^{(j)}\right)$ at certain prescribed points or nodes and then does some approximation according to different strategies. There are two typical SC approaches, one is Lagrange interpolation approach and the other one is pseudo-spectral approach.

2.3.2.1 Lagrange Interpolation

Let $\Theta_N = \{\boldsymbol{\xi}^{(j)}\}_{j=1}^{Q} \in \Gamma$ be a set of (prescribed) nodes in the N-dimensional random space Γ which the random variable $\boldsymbol{\xi}$ belongs to. The Lagrange interpolation approach (Xiu 2009) is to approximate the solution $\mathbf{x}(k, \boldsymbol{\xi})$ in the form

$$\mathbf{x}(k, \boldsymbol{\xi}) \approx \sum_{i=1}^{Q} \tilde{\mathbf{x}}_i(k) L_i(\boldsymbol{\xi}), \tag{2.29}$$

where $L_i(\boldsymbol{\xi})$ are called Lagrange polynomials which satisfy

$$L_i(\boldsymbol{\xi}^{(j)}) = \delta_{ij}, 1 \le i, j \le Q, \tag{2.30}$$

and $\tilde{\mathbf{x}}_i(k)$ are the values of solution $\mathbf{x}(k)$ at nodes $\boldsymbol{\xi}^{(i)} \in \Theta_N$.

It can be seen that once those Q points (realization of random vector $\boldsymbol{\xi}$) are known, the approximation of $\mathbf{x}(k, \boldsymbol{\xi})$ can be calculated through (2.29). The only thing needs to be done is acquiring the values of the solution $\mathbf{x}(k, \boldsymbol{\xi})$ at each realization $\boldsymbol{\xi}^{(j)}$. And this work can be completed by solving the original governing equations. Therefore, the Lagrange interpolation approach is to some extent equivalent to solving Q deterministic equations with realization $\boldsymbol{\xi}^{(j)}$ of the random variable $\boldsymbol{\xi}$. A significant advantage of Lagrange interpolation approach is that no modification to the governing equations is required and any existing solvers to solve the original equations can be applied. This is in contrast to SG method, in which the original governing equations will be modified to a set of coupled equations with the expansion coefficients as unknowns.

Once the Lagrange interpolation form in (2.29) is obtained, the evaluation of the statistics of the random solution is straightforward. For example, the expectation of the solution $\mathbf{x}(k, \boldsymbol{\xi})$ is approximated as

$$E[\mathbf{x}(k, \boldsymbol{\xi})] \approx \sum_{i=1}^{Q} \tilde{\mathbf{x}}_i(k) \int L_i(\boldsymbol{\xi}) W(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{2.31}$$

where $\int L_i(\boldsymbol{\xi})W(\boldsymbol{\xi})d\boldsymbol{\xi}$ serves as the weights in discrete sum.

In spite of the advantage of the Lagrange interpolation scheme, the selection of the points or nodes is nontrivial, especially in multi-dimensional spaces due to the lack of the theoretical aspects of Lagrange interpolation. Although there are some "rules" to choose those nodes in engineering field, most of them are ad hoc and have no control over the interpolation errors. What's more, the manipulation of the multi-dimension case is not straightforward.

2.3.2.2 Pseudo-spectral approach

Another type of SC approach is the pseudo-spectral approach. Pseudo-spectral approach is an alternative approach to approximate the integration in (2.22) and gives the values of the gPC coefficients, the formula is given by

$$\hat{\mathbf{v}}_{\boldsymbol{i}}(k) = \int \mathbf{x}(k,\boldsymbol{\xi})\Phi_{\boldsymbol{i}}(\boldsymbol{\xi})W(\boldsymbol{\xi})d\boldsymbol{\xi} \approx \sum_{j=1}^{Q} \mathbf{x}\left(k,\boldsymbol{\xi}^{(j)}\right)\Phi_{\boldsymbol{i}}(\boldsymbol{\xi}^{(j)})W\left(\boldsymbol{\xi}^{(j)}\right), \tag{2.32}$$

where $\boldsymbol{\xi}^{(j)}$ and $W\left(\boldsymbol{\xi}^{(j)}\right)$ are called collocation points and weights, respectively, while Q denotes the number of collocation points. $\mathbf{x}\left(k,\boldsymbol{\xi}^{(j)}\right)$ are the values of the solution to the governing equations at given collocation points $\boldsymbol{\xi}^{(j)}$. Similar to Lagrange interpolation method, no modification to the original system is needed and any existing solver can be used to solve the original governing equations.

The key part is the selection of the points and weights to ensure the accuracy and efficiency of the approximation to the integration, especially for multi-dimensional problems. Many choices are available for one dimensional space, i.e., $N = 1$. For every single dimension $i = 1,2,\dots N$, a set of nodes

$$\Theta_i^1 = \left\{p_i^1, \dots, p_i^{q_i}\right\} \subset \Gamma_i, i = 1,2,\dots N, \tag{2.33}$$

with weights $\{\alpha_i^1, \dots, \alpha_i^{q_i}\}$ can be found to approximate the one-dimensional integration $E[f(y)]$ as $\mathcal{U}_i^{q_i}[f]$ based on a good one-dimensional rule

$$E[f(y)] = \int_{\Gamma_i} f(y)w(y)dy = \sum_{j=1}^{q_i} f\left(p_i^j\right) \cdot \alpha_i^j \triangleq \mathcal{U}_i^{q_i}[f]. \tag{2.34}$$

One optimal choice is usually to use the Gaussian quadrature rules based on the orthogonal polynomials discussed in Chapter 2. See Appendix C for details about Gaussian quadrature rule.

The challenge lies in multi-dimensional space, i.e., $N > 1$, especially for large dimension, i.e., $N \gg 1$. The aim is to find $\mathcal{U}^Q[f]$ to approximate the multi-dimensional integration $E[f(\mathbf{y})]$ given by

$$E[f(\mathbf{y})] = \int_{\Gamma} f(\mathbf{y})W(\mathbf{y})d\mathbf{y}. \tag{2.35}$$

The point selection strategies for multi-dimensional space are discussed below.

(1) Tensor product

One choice is to use the tensor product of the nodes selected for one-dimensional space. The tensor product formula is given as follows

$$\mathcal{U}^Q[f] = \left(\mathcal{U}_1^{q_1} \otimes \cdots \otimes \mathcal{U}_N^{q_N}\right)[f] = \sum_{j_1=1}^{q_1} \cdots \sum_{j_N=1}^{q_N} f\left(p_1^{j_1}, \cdots, p_N^{j_N}\right) \cdot \left(\alpha_1^{j_1} \otimes \cdots \otimes \alpha_N^{j_N}\right). \tag{2.36}$$

Clearly, the total number of points needed is

$$Q = \prod_{i=1}^{N} q_i. \tag{2.37}$$

If the same number of points are used in each dimension, i.e., $q_1 = \cdots = q_N = q$, then the total number is $Q = q^N$. One problem for tensor product is that the total number of points grows quickly for high dimensions. For example, if three points (i.e., $q = 3$) are selected in each dimension, then the total would be $Q = 3^N$ (e.g., $3^{10} = 59049$ for

$N = 10$). Because of the rapid growth of the number of points with high dimensions, it is proper to use tensor product approach only in lower dimensional problems, e.g., $N \leq 5$.

(2) Sparse Grid

The *N*-dimensional sparse grid quadrature rule also combines univariate quadrature rules but in a different way as tensor product rule does. Compared to the exponential growth rate of the tensor product rule, the computational cost of sparse grid rule rises considerably slower. The basic idea of sparse grid quadrature originates from Smolyak rule (Smolyak 1963). The Smolyak algorithm is a linear combination of product formulas. The linear combination is chosen in a way to preserve an integration property for $N = 1$ and for $N > 1$ as much as possible (Xiu 2007).

An *N*-dimensional sparse grid quadrature denoted as $\mathcal{U}_{N,L}$ has accuracy level L, which means it is exact for complete polynomials of order up to $2L - 1$, e.g., for all polynomials ${x_1}^{i_1}{x_2}^{i_2} \cdots {x_N}^{i_N}$ with $\sum_{j=1}^{N} i_j \leq 2L - 1$ (Heiss and Winschel 2008; Jia et al. 2012). The construction of $\mathcal{U}_{N,L}$ is defined as follows:

Let $\mathcal{U}_i^{q_i}, i = 1, \dots N$, be the univariate quadrature for i-th dimension with level q_i (q_i points), and $\mathcal{U}_i^0 = 0$. For each dimension, the difference of the approximation when increasing the accuracy level from $q_i - 1$ to q_i is defined as

$$\Delta_i^{q_i} = \mathcal{U}_i^{q_i} - \mathcal{U}_i^{q_i - 1}, i = 1, \dots N. \tag{2.38}$$

By introducing an auxillary number p, define

$$\mathbb{N}_p^N = \begin{cases} \left\{\mathbf{q} = (q_1, \dots, q_N) : q_j \geq 1 \text{ and } \sum_{j=1}^{N} q_j = N + q\right\} & p \geq 0 \\ \emptyset & p < 0 \end{cases}. \tag{2.39}$$

As an example, $\mathbb{N}_3^2 = \{(1,4),(2,3),(3,2),(4,1)\}$. $\mathcal{U}_{N,L}$ is then constructed based on the Smolyak rule as

$$\mathcal{U}_{N,L}[f] = \sum_{p=0}^{L-1} \sum_{\mathbf{q}\in\mathbb{N}_p^N} \left(\Delta_1^{q_1} \otimes \cdots \otimes \Delta_N^{q_N}\right)[f]. \quad (2.40)$$

Instead of using the differences, $\mathcal{U}_{N,L}$ can be rewritten in the terms of the univariate quadrature rules (Wasilkowski and Wozniakowski 1995) as

$$\mathcal{U}_{N,L}[f] = \sum_{p=L-N}^{L-1} (-1)^{L-1-p} \binom{N-1}{L-1-p} \sum_{\mathbf{q}\in\mathbb{N}_p^N} \left(\mathcal{U}_1^{q_1} \otimes \cdots \otimes \mathcal{U}_N^{q_N}\right)[f]. \quad (2.41)$$

From (2.41), the rule is a weighted sum of product rules with different combination of accuracy levels $q = (q_1, \ldots, q_N)$. Different combinations may have one same point, in this case the associated weight with the point will be the sum of the weights in all occurrences.

2.4 Discussions

In this chapter, the mathematical theories of Hermite polynomial chaos expansion and the generalized polynomial chaos expansion, and the framework of using polynomial chaos expansion to quantify the uncertainty in the forecast from a dynamical model with random inputs were introduced. To obtain the expansion coefficients in the expansion, two typical methods called stochastic Galerkin (SG) and stochastic Collocation (SC) were described in detail.

Chapter 3

Unscented Transformation

The unscented transformation (UT) is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation (Julier and Uhlmann 1996, 1997a, 1997b). It is developed based on the intuition that the approximation of a probability distribution is easier than that of an arbitrary nonlinear function or transformation (Julier et al. 1995). Similar to Monte Carlo approach, a set of samples are adopted to capture the prior probability distribution of the random inputs. These samples are then propagated through the dynamical model and the statistics of the posterior probability distribution are calculated using these transformed samples. However, different from MC, the samples called sigma points in UT are selected deterministically. In this chapter, the mathematical background of UT and an improved version scaled unscented transform (SUT) and their application in uncertainty quantification will be presented.

3.1 The Basic Unscented Transformation

To state the basic idea of UT, let $\mathbf{x} \in R^n$ be an n-dimensional random variable which has mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P_x}$, respectively. The m-dimensional random variable $\mathbf{y} \in R^m$ is related to $\mathbf{x}$ by a non-linear transformation $\mathbf{g}$, i.e.,

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \tag{3.1}$$

The objective is to predict the mean $\bar{\mathbf{y}}$ and covariance $\mathbf{P_y}$ of $\mathbf{y}$.

Following the intuition, UT seeks help from a discrete distribution which has the same mean and covariance (and possibly higher moments) with the random input $\mathbf{x}$ and each point in the discrete approximation can be directly propagated through the non-linear transformation. The mean and covariance of the transformed points are calculated

afterwards and treated as the estimate/approximation of the nonlinear transformation of the original distribution. The discrete distribution replacement is accomplished by selecting a set of points called sigma points to capture the mean and covariance of **x**. Though the idea of UT bears a superficial resemblance of Monte Carlo approach, the strategy is totally different. Instead of random sampling in Monte Carlo approach, the sampling in UT is deterministic. Besides, the distribution interpretation based on the sigma points is inconsistent with that for Monte Carlo method. For example, the weights put on the sigma points do not have to lie in the range [0, 1] and can be negative (Julier and Uhlmann 2004).

In UT, a set of sigma points containing $p+1$ vectors $\boldsymbol{\chi}_i, i = 0, \ldots, p$ with associated weights W_i are selected. The weights can be positive and negative but must obey the condition

$$\sum_{i=0}^{p} W_i = 1. \tag{3.2}$$

They also assure the mean and covariance through

$$\bar{\mathbf{x}} = \sum_{i=0}^{p} W_i \boldsymbol{\chi}_i,$$

$$\mathbf{P}_{\mathbf{x}} = \sum_{i=0}^{p} W_i (\boldsymbol{\chi}_i - \bar{\mathbf{x}})(\boldsymbol{\chi}_i - \bar{\mathbf{x}})^{\mathrm{T}}. \tag{3.3}$$

Each sigma point is then transformed through the nonlinear function **g**,

$$\boldsymbol{\mathcal{Y}}_i = \mathbf{g}(\boldsymbol{\chi}_i), i = 0, \ldots, p. \tag{3.4}$$

The mean and covariance of the transformed sigma points are calculated as

$$\bar{\mathbf{y}} = \sum_{i=0}^{p} W_i \boldsymbol{\mathcal{Y}}_i,$$

$$\mathbf{P}_{\mathbf{y}} = \sum_{i=0}^{p} W_i (\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}})(\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}})^{\mathrm{T}}. \tag{3.5}$$

One selection scheme (Julier and Uhlmann 1996) which satisfies the above requirement consists of $2n+1$ symmetric sigma points $\boldsymbol{\chi}_i, i = 0, \ldots, 2n$, centered at the mean $\bar{\mathbf{x}}$ and lies on the matrix square root is given as follows:

$$\boldsymbol{\chi}_0 = \bar{\mathbf{x}},$$

$$\boldsymbol{\chi}_i = \bar{\mathbf{x}} + \left(\sqrt{(n+\kappa)\mathbf{P}_\mathbf{x}}\right)_i, i = 1, \ldots, n,$$

$$\boldsymbol{\chi}_i = \bar{\mathbf{x}} - \left(\sqrt{(n+\kappa)\mathbf{P}_\mathbf{x}}\right)_i, i = n+1, \ldots, 2n, \tag{3.6}$$

where κ is a scaling parameter and $\left(\sqrt{(n+\kappa)\mathbf{P}_\mathbf{x}}\right)_i$ is the ith column of matrix $\sqrt{(n+\kappa)\mathbf{P}_\mathbf{x}}$ which is the matrix square root of $(n+\kappa)\mathbf{P}_\mathbf{x}$ and can be obtained through Cholesky factorization.

The weights associated with the sigma points are given by

$$W_0 = \kappa/(n+\kappa),$$

$$W_i = 1/2(n+\kappa), i = 1, \ldots, 2n. \tag{3.7}$$

After propagating each sigma point through the nonlinear transformation $\mathbf{g}$,

$$\boldsymbol{\mathcal{Y}}_i = \mathbf{g}(\boldsymbol{\chi}_i), i = 0, \ldots, 2n. \tag{3.8}$$

The mean and covariance of the transformed sigma points are calculated as

$$\bar{\mathbf{y}} = \sum_{i=0}^{2n} W_i \, \boldsymbol{\mathcal{Y}}_i,$$

$$\mathbf{P}_\mathbf{y} = \sum_{i=0}^{2n} W_i \, (\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}})(\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}})^{\mathrm{T}}. \tag{3.9}$$

The estimates of the mean and covariance of the transformed vector obtained by equations in (3.9) are accurate to the second order of the Taylor series expansion for any nonlinear transformation $\mathbf{g}(\mathbf{x})$. Errors are introduced in the third and higher order moments but are scaled by the choice of the parameter κ. Please refer to Appendix D for details of the accuracy.

3.2 The Scaled Unscented Transformation

One problem which lies in the basic UT is that with the increase of the dimension of the state space, the radius of the sphere that bounds all the sigma points will increase as well (Julier 2002; Van Der Merwe et al., 2000). Even though the sigma points can still capture the mean and covariance of the prior distribution correctly, it does so at the cost of sampling non-local effects. The difficulties become significant for strong nonlinearities. The parameter κ is a scaling parameter which is used to scale the sigma points towards or away from the mean of the prior distribution. As observed from the construction of the sigma points, the distance of ith sigma point $\boldsymbol{\chi}_i$ (except for $\boldsymbol{\chi}_0$) from the mean $\bar{\mathbf{x}}$, i.e., $|\boldsymbol{\chi}_i - \bar{\mathbf{x}}|$ is proportional to $\sqrt{(n+\kappa)}$. When $\kappa = 0$, the distance is proportional to $\sqrt{n}$. When $\kappa < 0$, the sigma points are scaled towards the mean and when $\kappa > 0$, the sigma points are scaled further from the mean. As a special case, when $\kappa = 3 - n$, the value of n has no influence on the distance. However, when $\kappa = 3 - n < 0$, the weight $W_0 < 0$, an non-positive semi-definite covariance matrix could be obtained. The scaled unscented transformation (SUT) developed in (Julier 2002) was to address this problem. SUT aims to overcome the dimensional scaling effects by using a set of scaled sigma points,

$$\boldsymbol{\chi}_i' = \boldsymbol{\chi}_0 + \alpha(\boldsymbol{\chi}_i - \boldsymbol{\chi}_0), \tag{3.10}$$

where α is a positive scaling parameter. The choice of α should guarantee the second order accuracy of the mean and covariance, and the positive semi-definiteness of the covariance. It can be made arbitrarily small to minimize the high order effects.

3.2.1 The Auxillary Random Variable

Before studying the mechanism of SUT, let's first examine an auxillary random variable $\mathbf{z}$, which is related to the original function by

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) = \frac{\mathbf{g}[\bar{\mathbf{x}}+\alpha(\mathbf{x}-\bar{\mathbf{x}})]-\mathbf{g}(\bar{\mathbf{x}})}{\alpha^2} + \mathbf{g}(\bar{\mathbf{x}}). \tag{3.11}$$

The objective can be achieved by performing the original UT to this auxillary random transformation, i.e., propagating each sigma point (the unscaled sigma points) through (3.11),

$$\boldsymbol{z}_i = \frac{\mathbf{g}[\bar{\mathbf{x}}+\alpha(\boldsymbol{\chi}_i-\bar{\mathbf{x}})]-\mathbf{g}(\bar{\mathbf{x}})}{\alpha^2} + \mathbf{g}(\bar{\mathbf{x}}), i = 0, \dots, p. \tag{3.12}$$

Then the mean $\bar{\mathbf{z}}$ and covariance $\mathbf{P}_{\boldsymbol{z}}$ are approximated by

$$\bar{\mathbf{z}} = \sum_{i=0}^{p} W_i \, \boldsymbol{z}_i,$$

$$\mathbf{P}_{\mathbf{z}} = \alpha^2 \sum_{i=0}^{2n} W_i \, (\boldsymbol{z}_i - \bar{\mathbf{z}})(\boldsymbol{z}_i - \bar{\mathbf{z}})^{\mathrm{T}}. \tag{3.13}$$

It can be verified that the Taylor series expansions of $\bar{\mathbf{z}}$ and $\mathbf{P}_{\boldsymbol{z}}$ agree with those of $\bar{\mathbf{y}}$ and $\mathbf{P}_{\boldsymbol{y}}$ approximated by equation (3.5). The proof of the accuracy is given in Appendix E and refer to (Julier 2002) for more details.

As seen from (3.12), all sigma points are propagated through the term $\mathbf{g}[\bar{\mathbf{x}} + \alpha(\mathbf{x} - \bar{\mathbf{x}})]$, the scaling effect given by equation (3.10) is implicitly achieved.

3.2.2 The Scaled Unscented Transform

Although the auxillary form of the unscented transform is able to meet the requirements mentioned at the beginning in this section, it requires the change of the transformation. The SUT yields the same results as the auxillary random variable problem does, but without modifying the transformation. As mentioned before, a set of scaled sigma points constructed in (3.10) by a positive scaling parameter α will be used in SUT, with a new set of weights given by

$$W_0' = \frac{W_0}{\alpha^2} + (1 - \frac{1}{\alpha^2}),$$

$$W_i' = \frac{W_i}{\alpha^2}, i = 1, \dots, p. \tag{3.14}$$

Appendix E gives the details about the selection of the weights.

With the new set of points (3.10) with weights (3.14), the SUT estimates the transformation and calculates its statistics as follows,

$$\boldsymbol{\mathcal{Y}_i'} = \mathbf{g}(\boldsymbol{\chi}_i'),$$

$$\bar{\mathbf{y}}' = \sum_{i=0}^{p} W_i' \, \boldsymbol{\mathcal{Y}_i'},$$

$$\mathbf{P}_{\boldsymbol{y}}' = \sum_{i=0}^{p} W_i' \, (\boldsymbol{\mathcal{Y}_i'} - \bar{\mathbf{y}}')(\boldsymbol{\mathcal{Y}_i'} - \bar{\mathbf{y}}')^{\mathrm{T}} + (1 - \alpha^2)(\boldsymbol{\mathcal{Y}_0'} - \bar{\mathbf{y}}')(\boldsymbol{\mathcal{Y}_0'} - \bar{\mathbf{y}}')^{\mathrm{T}}. \tag{3.15}$$

In Appendix E, it is shown that $\bar{\mathbf{y}}' = \bar{\mathbf{z}}$ and $\mathbf{P}_{\boldsymbol{y}}' = \mathbf{P}_{\mathbf{z}}$ for any sigma point distribution. In consequence, the SUT will carry all properties of the auxillary form. The mean and covariance estimated by (3.15) are accurate to the second order and the positive semi-definiteness of the covariance $\mathbf{P}_{\boldsymbol{y}}'$ is guaranteed if all the unscaled weights are positive. Besides, the numerical cost of SUT is the same as that for original UT. By comparing the covariance calculations in (3.5) and (3.15), the only difference is that a term $(1 - \alpha^2)$ is added to the zeroth sigma point. It can be interpreted simply as, when $\alpha = 1$, the form of (3.15) becomes that of (3.5); when $\alpha = 0$, (3.15) becomes the modified covariance calculation as in Julier and Uhlmann (2000).

Although the sigma points only capture the first two moments (the mean and covariance) accurate to the second order, the scaled unscented transform can be extended to include partial information of the higher order terms in the Taylor series expansion of the covariance. Adding an extra weighting parameter β to the zeroth sigma point, further higher order effects can be incorporated at no additional computational cost, the estimation of the covariance becomes

$$\mathbf{P}'_{\boldsymbol{y}} = \sum_{i=0}^{p} W_i' \,(\boldsymbol{\mathcal{Y}}'_{\boldsymbol{i}} - \bar{\mathbf{y}}')(\boldsymbol{\mathcal{Y}}'_{\boldsymbol{i}} - \bar{\mathbf{y}}')^{\mathrm{T}} + (1 - \alpha^2 + \beta\,)(\boldsymbol{\mathcal{Y}}'_{\mathbf{0}} - \bar{\mathbf{y}}')(\boldsymbol{\mathcal{Y}}'_{\mathbf{0}} - \bar{\mathbf{y}}')^{\mathrm{T}}. \quad (3.16)$$

For a special case, when the input **x** is Gaussian distributed, the optimal choice for the parameter is $\beta = 2$ (Julier and Uhlmann 2004).

Further, in order to reduce the number of calculations, the sigma point selection and scaling can be combined into one single step (*Van Der Merwe et al.* 2000; *Van Der Merwe* 2004). For example, corresponding to the sigma point selection scheme in (3.6) with weights in (3.7), the scaled sigma points with weights can be summarized as follows.

Let

$$\lambda = \alpha^2(n + \kappa) - n. \quad (3.17)$$

The $2n + 1$ scaled sigma points $\boldsymbol{\chi}_i, i = 0, \ldots, 2n,$ and respective mean (m) and covariance (c) weights $W_i^{(j)}, j \in (m, c)$, are selected as follows:

$$\begin{gathered}
\boldsymbol{\chi}_0 = \bar{\mathbf{x}}, \\
\boldsymbol{\chi}_\mathrm{i} = \bar{\mathbf{x}} + \left(\sqrt{(n + \lambda)\mathbf{P}_\mathbf{x}}\right)_i, i = 1, \ldots, n, \\
\boldsymbol{\chi}_\mathrm{i} = \bar{\mathbf{x}} - \left(\sqrt{(n + \lambda)\mathbf{P}_\mathbf{x}}\right)_i, i = n + 1, \ldots, 2n, \\
W_0^{(m)} = \lambda/(n + \lambda), \\
W_0^{(c)} = \lambda/(n + \lambda) + (1 - \alpha^2 + \beta), \\
W_i^{(m)} = W_i^{(c)} = 1/\{2(n + \lambda)\}, i = 1, \ldots, 2n,
\end{gathered} \quad (3.18)$$

where $\left(\sqrt{(n + \lambda)\mathbf{P}_\mathbf{x}}\right)_i$ is the ith column of matrix $\sqrt{(n + \lambda)\mathbf{P}_\mathbf{x}}$ (matrix square root, e.g., Cholesky factorization). α, β, and κ are three parameters, $0 \leq \alpha \leq 1$ controls the spread of the sigma points and ideally should be small to avoid sampling non-

local effects for strong nonlinearities, $\beta \geq 0$ incorporates knowledge of the higher order moments of the distribution and its optimal choice for a Gaussian distribution is $\beta = 2$, and $\kappa \geq 0$ is used to guarantee positive semi-definiteness of the covariance matrix and its value is not critical, e.g., usually $\kappa = 0$.

Consider the problem of propagating the random variable **x** through a nonlinear function **g**(**x**) given in Equation (3.1), the complete procedure of the SUT is as follows (Van Der Merwe et al. 2000):

(1) Choose values for parameters α, β, and κ.

(2) Determine the sigma point set $\mathbf{S}_i = \left\{\boldsymbol{\chi}_i, \mathrm{W}_i^{(j)}, i = 0, \ldots, 2n, j \in (m, c)\right\}$ according to equation (3.18).

(3) Propagate every sigma point through the nonlinear transformation **g**, i.e.,

$$\boldsymbol{\mathcal{Y}}_i = \mathbf{g}(\boldsymbol{\chi}_i), i = 0, \ldots, 2n. \tag{3.19}$$

(4) Calculate the mean and covariance of **y** as

$$\bar{\mathbf{y}} = \sum_{i=0}^{2n} W_i^{(m)} \boldsymbol{\mathcal{Y}}_i, \tag{3.20}$$

$$\mathbf{P}_{\mathbf{y}} = \sum_{i=0}^{2n} W_i^{(c)} (\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}})(\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}})^{\mathrm{T}}. \tag{3.21}$$

3.3 Discussions

As an alternative sampling method to quantify the uncertainty in the forecast, the unscented transformation method uses a set of deterministically chosen samples (called sigma points) to represent and propagate the uncertainty through a dynamic forecast model. In this chapter, the theory and implementation details of UT and its improved version scaled unscented transformation (SUT) were presented. The accuracy was also discussed in this chapter.

Chapter 4

Application of Stochastic Galerkin Method

In this chapter, the ability of polynomial chaos expansion approach with expansion coefficients solved by stochastic Galerkin scheme will be studied. The dynamic model studied in (Lewis 2014) is used as an example to investigate its application and performance.

4.1 The Two-variable Model

The model is given as

$$\frac{du_1}{dt} = -\frac{1}{2}u_1u_2, \tag{4.1}$$

$$\frac{du_2}{dt} = \frac{1}{2}[u_1]^2. \tag{4.2}$$

These two equations govern the truncated two-component spectral form of solution to the nonlinear advection equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0, \tag{4.3}$$

where $u(x,t)$ is the flow field and must be initialized in the spatial domain, $0 \leq x \leq 2\pi$, at initial time. The spectral form of the solution is given in (Platzman 1964) as

$$u(x,t) = -\sum_{n=1}^{\infty} u_n(t)\, sin(nx). \tag{4.4}$$

Equations (4.1) and (4.2) are the truncated two-mode spectral form of the solution for the longest waves ($n = 1, 2$) which are referred to as the primary ($n = 1$) and secondary wave ($n = 2$). As stated in (Lewis 2014), an "energy conservation" principal is processed in the dynamical system and expressed as

$$\frac{d}{dt}\{u_1(t)^2 + u_2(t)^2\} = 0, \tag{4.5}$$

which is equivalent to

$$u_1(t)^2 + u_2(t)^2 = u_1(0)^2 + u_2(0)^2 = c^2. \tag{4.6}$$

where $u_1(0)$ and $u_2(0)$ represent the initial conditions.

Lewis (2014) gives the analytic solution to (4.1) and (4.2) as

$$u_1(t) = u_1(0)\frac{(1+\alpha)e^{\frac{ct}{2}}}{(1+\alpha e^{ct})}, \tag{4.7}$$

$$u_2(t) = \frac{c(\alpha e^{ct}-1)}{(1+\alpha e^{ct})}, \tag{4.8}$$

where

$$c = \sqrt{[u_1(0)]^2 + [u_2(0)]^2}, \tag{4.9}$$

$$\alpha = \frac{c+u_2(0)}{c-u_2(0)}. \tag{4.10}$$

And if the initial condition (IC) follows the bivariate normal distribution and the initial probability density function (PDF) is $D(u_1(0), u_2(0), 0)$, using Liouville's equation the exact PDF at time t can be obtained as

$$D(u_1(t), u_2(t), t) = D(u_1(0), u_2(0), 0) \times \frac{\sqrt{(\alpha+e^{-ct})(1+\alpha e^{ct})}}{(1+\alpha)}. \tag{4.11}$$

Same as the study in (Lewis 2014), the assumption of the experiment is that IC follows bivariate normal distribution with means and variances given by

$$\mu_1(0) = 1.25, \mu_2(0) = -0.35,$$

$$\sigma_1 = \sigma_2 = 0.09, \sigma_{12} = 0. \tag{4.12}$$

The remaining task in this chapter will be using PC approach to quantify the uncertainty lying in the forecast by the model of equations (4.1) and (4.2) starting from initial conditions given in (4.12).

Let $\mathbf{u} = (u_1, u_2)^{\mathrm{T}}$, the aim is to find an approximation of $\mathbf{u}(t)$ in the PC expansion form of

$$\mathbf{u}_N^M(t,\boldsymbol{\xi}) = \sum_{|\mathbf{i}|=0}^{M} \mathbf{v}_{\mathbf{i}}(t)\Phi_{\mathbf{i}}(\boldsymbol{\xi}), \tag{4.13}$$

where $\{\mathbf{v}_{\mathbf{i}}(t) = [\mathbf{v}_{\mathbf{i},1}(t), \mathbf{v}_{\mathbf{i},2}(t)]^{\mathrm{T}}\} \in R^2$ are the expansion coefficients, N is the number of random variable used in the expansion and M is the highest order of the polynomials. Since the initial condition is assumed to follow bivariate normal distribution, the optimal choice for $\Phi_{\mathbf{i}}(\boldsymbol{\xi})$ in (4.13) will be the Hermite polynomials according to Table (2.1). In this section, both the univariate Hermite polynomial chaos expansion ($N = 1$) and multivariate Hermite polynomial chaos expansion ($N = 2$) are examined.

4.2 Univariate Hermite Polynomial Chaos Expansion

When ($N = 1$), the equation (4.13) becomes

$$\mathbf{u}_1^M(t,\xi) = \sum_{i=0}^{M} \mathbf{v}_i(t)H_i(\xi). \tag{4.14}$$

Substitute equation (4.14) into the original equations (4.1) and (4.2), and obtain

$$\begin{aligned}\sum_{i=0}^{M} \frac{\mathrm{d}v_{i,1}(t)}{dt} H_i(\xi) &= -\frac{1}{2}\left(\sum_{i=0}^{M} v_{i,1}(t)H_i(\xi)\right)\left(\sum_{i=0}^{M} v_{i,2}(t)H_i(\xi)\right) \\ &= -\frac{1}{2}\sum_{i=0}^{M}\sum_{j=0}^{M} v_{i,1}(t)v_{j,2}(t)H_i(\xi)H_j(\xi), \end{aligned} \tag{4.15}$$

$$\begin{aligned}\sum_{i=0}^{M} \frac{\mathrm{d}v_{i,2}(t)}{dt} H_i(\xi) &= \frac{1}{2}\left(\sum_{i=0}^{M} v_{i,1}(t)H_i(\xi)\right)^2 \\ &= \frac{1}{2}\sum_{i=0}^{M}\sum_{j=0}^{M} v_{i,1}(t)v_{j,1}(t)H_i(\xi)H_j(\xi). \end{aligned} \tag{4.16}$$

For convenience, $\mathbf{v}_i(t)$ and $H_i(\xi)$ are simplified as $\mathbf{v}_i$ and H_i, respectively in following paragraphs. Then Galerkin projection (i.e., multiply both sides of each equation by $H_k(\xi), k = 0, \dots, M$) is applied on equations (4.15) and (4.16), then the following equations are obtained

$$\begin{aligned}&\sum_{i=0}^{M} \frac{\mathrm{d}v_{i,1}}{dt} \langle H_i, H_k\rangle \\ &= -\frac{1}{2}\sum_{i=0}^{M}\sum_{j=0}^{M} v_{i,1}v_{j,2}\langle H_iH_j, H_k\rangle, \end{aligned} \tag{4.17}$$

$$\sum_{i=0}^{M} \frac{\mathrm{d}v_{i,2}}{dt} \langle H_i, H_k \rangle$$

$$= \frac{1}{2} \sum_{i=0}^{M} \sum_{j=0}^{M} v_{i,1} v_{j,1} \langle H_i H_j, H_k \rangle. \tag{4.18}$$

It is assumed that the normalized Hermite polynomials are used in the expansion. According to the orthogonality property of Hermite polynomials, $(2M+2)$ equations for the expansion coefficients are obtained as follows,

$$\frac{\mathrm{d}v_{k,1}}{dt} = -\frac{1}{2} \sum_{i=0}^{M} \sum_{j=0}^{M} v_{i,1} v_{j,2} \langle H_i H_j, H_k \rangle, k = 0, \dots, M,$$

$$\frac{\mathrm{d}v_{k,2}}{dt} = \frac{1}{2} \sum_{i=0}^{M} \sum_{j=0}^{M} v_{i,1} v_{j,1} \langle H_i H_j, H_k \rangle, k = 0, \dots, M. \tag{4.19}$$

After computing the inner products $\langle H_i H_j, H_k \rangle, k = 0, \dots, M$, the explicit form of the equations will be obtained.

The initial values for the expansion coefficients are easily derived from the initial condition in (4.12) as,

$$\mathbf{v}_0(0) = [\mu_1(0), \mu_2(0)]^{\mathrm{T}} = [1.25, -0.35]^{\mathrm{T}},$$

$$\mathbf{v}_1(0) = [\sqrt{\sigma_1}, \sqrt{\sigma_2}]^{\mathrm{T}} = [0.3, 0.3]^{\mathrm{T}},$$

$$\mathbf{v}_i(0) = [0,0]^{\mathrm{T}}, i = 2, \dots, M. \tag{4.20}$$

The remaining task is to solve the equations given in (4.19) with the initial condition in (4.20) to obtain the expansion coefficients $\mathrm{v}_i(t), i = 0, \dots, M$, at different times. Once the coefficient values at time t are available, a surrogate in the form of (4.13) for the stochastic process $\mathbf{u}(t)$ will be obtained. By sampling the random variable $\boldsymbol{\xi}$, an ensemble of forecast and the histogram will be generated. In addition, the statistics of the forecast can be calculated either through the coefficients (in theory) or through the samples afterwards.

As an example, Tables 4.1-4.7 show the inner products in (4.15) and (4.16) up to $M = 6$.

Table 4.1 Inner products $\langle \boldsymbol{H_i H_j, H_0} \rangle$

	H_0	H_1	H_2	H_3	H_4	H_5	H_6
H_0	1	0	0	0	0	0	0
H_1	0	1	0	0	0	0	0
H_2	0	0	1	0	0	0	0
H_3	0	0	0	1	0	0	0
H_4	0	0	0	0	1	0	0
H_5	0	0	0	0	0	1	0
H_6	0	0	0	0	0	0	1

Table 4.2 Inner products $\langle \boldsymbol{H_i H_j, H_1} \rangle$

	H_0	H_1	H_2	H_3	H_4	H_5	H_6
H_0	0	1	0	0	0	0	0
H_1	1	0	1.4142	0	0	0	0
H_2	0	1.4142	0	1.7321	0	0	0
H_3	0	0	1.7321	0	2	0	0
H_4	0	0	0	2	0	2.2361	0
H_5	0	0	0	0	2.2361	0	2.4495
H_6	0	0	0	0	0	2.4495	0

Table 4.3 Inner products $\langle \boldsymbol{H_i H_j, H_2} \rangle$

	H_0	H_1	H_2	H_3	H_4	H_5	H_6
H_0	0	0	1	0	0	0	0
H_1	0	1.4142	0	1.7321	0	0	0
H_2	1	0	2.8284	0	2.4495	0	0
H_3	0	1.7321	0	4.2426	0	3.1623	0
H_4	0	0	2.4495	0	5.6569	0	3.8730
H_5	0	0	0	3.1623	0	7.0711	0
H_6	0	0	0	0	3.8730	0	8.4853

Table 4.4 Inner products $\langle H_i H_j, H_3 \rangle$

	H_0	H_1	H_2	H_3	H_4	H_5	H_6
H_0	0	0	0	1	0	0	0
H_1	0	0	1.7321	0	2	0	0
H_2	0	1.7321	0	4.2426	0	3.1623	0
H_3	1	0	4.2426	0	7.3485	0	4.4721
H_4	0	2	0	7.3485	0	10.9545	0
H_5	0	0	3.1623	0	10.9545	0	15
H_6	0	0	0	4.4721	0	15	0

Table 4.5 Inner products $\langle H_i H_j, H_4 \rangle$

	H_0	H_1	H_2	H_3	H_4	H_5	H_6
H_0	0	0	0	0	1	0	0
H_1	0	0	0	2	0	2.2361	0
H_2	0	0	2.4495	0	5.6569	0	3.8730
H_3	0	2	0	7.3485	0	10.9545	0
H_4	1	0	5.6569	0	14.6969	0	17.8885
H_5	0	2.2361	0	10.9545	0	24.4949	0
H_6	0	0	3.8730	0	17.8885	0	36.74230

Table 4.6 Inner products $\langle H_i H_j, H_5 \rangle$

	H_0	H_1	H_2	H_3	H_4	H_5	H_6
H_0	0	0	0	0	0	1	0
H_1	0	0	0	0	2.2361	0	2.4495
H_2	0	0	0	3.1623	0	7.0711	0
H_3	0	0	3.1623	0	10.9545	0	15
H_4	0	2.2361	0	10.9545	0	24.4949	0
H_5	1	0	7.0711	0	24.4949	0	44.7214
H_6	0	2.4495	0	15	0	44.7214	0

Table 4.7 Inner products $\langle H_i H_j, H_6 \rangle$

	H_0	H_1	H_2	H_3	H_4	H_5	H_6
H_0	0	0	0	0	0	0	1
H_1	0	0	0	0	0	2.4495	0
H_2	0	0	0	0	3.8730	0	8.4853
H_3	0	0	0	4.4721	0	15	0
H_4	0	0	3.8730	0	17.8885	0	36.7423
H_5	0	2.4495	0	15	0	44.7214	0
H_6	1	0	8.4853	0	36.7423	0	89.4427

For example, when $M = 4$, define a new vector $\mathbf{V}(t)$ by concatenating vectors $\mathbf{v}_0(t)$, $\mathbf{v}_1(t)$, …, $\mathbf{v}_4(t)$ as

$$\mathbf{V}(t) = [\mathbf{v}_0^{\mathrm{T}}(t), \mathbf{v}_1^{\mathrm{T}}(t), \dots, \mathbf{v}_4^{\mathrm{T}}(t)]^{\mathrm{T}} = [V_1, V_2, \dots, V_{10}]^{\mathrm{T}}, \tag{4.21}$$

that is,

$$[V_1, V_2] = \mathbf{v}_0^{\mathrm{T}}, [V_3, V_4] = \mathbf{v}_1^{\mathrm{T}}, \dots, [V_9, V_{10}] = \mathbf{v}_4^{\mathrm{T}}. \tag{4.22}$$

Then the resulting coefficient equations are:

$$\frac{dV_1}{dt} = -\frac{1}{2}(V_1V_2 + V_3V_4 + V_5V_6 + V_7V_8 + V_9V_{10}),$$

$$\frac{dV_2}{dt} = \frac{1}{2}(V_1^2 + V_3^2 + V_5^2 + V_7^2 + V_9^2),$$

$$\frac{dV_3}{dt} = -\frac{1}{2}(V_1V_4 + V_3V_2 + 1.4142V_3V_6 + 1.4142V_5V_4 + 1.7321V_5V_8 + 1.7321V_7V_6 + 2V_7V_{10} + 2V_9V_8),$$

$$\frac{dV_4}{dt} = \frac{1}{2}(V_1V_3 + V_3V_1 + 1.4142V_3V_5 + 1.4142V_5V_3 + 1.7321V_5V_7 + 1.7321V_7V_5 + 2V_7V_9 + 2V_9V_7),$$

$$\frac{dV_5}{dt} = -\frac{1}{2}(V_1V_6 + V_5V_2 + 1.4142V_3V_4 + 1.7321V_3V_8 + 1.7321V_7V_4 + 2.8284V_5V_6 + 2.4495V_5V_{10} + 2.4495V_9V_6 + 4.2426V_7V_8 + 5.6569V_9V_{10}),$$

$$\frac{dV_6}{dt} = \frac{1}{2}(V_1V_5 + V_5V_1 + 1.4142V_3^2 + 1.7321V_3V_7 + 1.7321V_7V_3 + 2.8284V_5^2 + 2.4495V_5V_9 + 2.4495V_9V_5 + 4.2426V_7^2 + 5.6569V_9^2),$$

$$\frac{dV_7}{dt} = -\frac{1}{2}(V_1V_8 + V_7V_2 + 1.7321V_3V_6 + 1.7321V_5V_4 + 2V_3V_{10} + 2V_9V_4 + 4.2426V_5V_8 + 4.2426V_7V_6 + 7.3485V_7V_{10} + 7.3485V_9V_8),$$

$$\frac{dV_8}{dt} = \frac{1}{2}(V_1V_7 + V_7V_1 + 1.7321V_3V_5 + 1.7321V_5V_3 + 2V_3V_9 + 2V_9V_3 + 4.2426V_5V_7 + 4.2426V_7V_5 + 7.3485V_7V_9 + 7.3485V_9V_7),$$

$$\frac{dV_9}{dt} = -\frac{1}{2}(V_1V_{10} + V_9V_2 + 2V_3V_8 + 2V_7V_4 + 2.4495V_5V_6 + 5.6569V_5V_{10} + 5.6569V_9V_6 + 7.3485V_7V_8 + 14.6969V_9V_{10}),$$

$$\frac{dV_{10}}{dt} = \frac{1}{2}(V_1V_9 + V_9V_1 + 2V_3V_7 + 2V_7V_3 + 2.4495V_5^2 + 5.6569V_5V_9 + 5.6569V_9V_5 + 7.3485V_7^2 + 14.6969V_9^2),$$

(4.23)

and the initial condition is

$$\mathbf{V}(0) = [1.25 \quad -0.35 \quad 0.3 \quad 0.3 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^{\mathrm{T}}. \tag{4.24}$$

4.3 Multivariate Hermite Polynomial Chaos Expansion

When ($N = 2$), the equation (4.13) becomes

$$\mathbf{u}_2^M(t, \boldsymbol{\xi}) = \sum_{|\mathbf{i}|=0}^{M} \mathbf{v}_{\mathbf{i}}(t)\Phi_{\mathbf{i}}(\boldsymbol{\xi}). \tag{4.25}$$

Substitute equation (4.25) into the original equations (4.1) and (4.2), and obtain

$$\sum_{|\mathbf{i}|=0}^{M} \frac{dv_{\mathbf{i},1}(t)}{dt} H_{\mathbf{i}}(\boldsymbol{\xi}) = -\frac{1}{2}\left(\sum_{|\mathbf{i}|=0}^{M} v_{\mathbf{i},1}(t)H_{\mathbf{i}}(\boldsymbol{\xi})\right)\left(\sum_{|\mathbf{i}|=0}^{M} v_{\mathbf{i},2}(t)H_{\mathbf{i}}(\boldsymbol{\xi})\right)$$

$$= -\frac{1}{2}\sum_{|\mathbf{i}|=0}^{M}\sum_{|\mathbf{j}|=0}^{M} v_{\mathbf{i},1}(t)\, v_{\mathbf{j},2}(t)H_{\mathbf{i}}(\boldsymbol{\xi})H_{\mathbf{j}}(\boldsymbol{\xi}), \tag{4.26}$$

$$\sum_{|\mathbf{i}|=0}^{M} \frac{dv_{\mathbf{i},2}(t)}{dt} H_{\mathbf{i}}(\boldsymbol{\xi}) = \frac{1}{2}\left(\sum_{|\mathbf{i}|=0}^{M} v_{\mathbf{i},1}(t)H_{\mathbf{i}}(\boldsymbol{\xi})\right)^2$$

$$= \frac{1}{2}\sum_{|\mathbf{i}|=0}^{M}\sum_{|\mathbf{j}|=0}^{M} v_{\mathbf{i},1}(t)\, v_{\mathbf{j},1}(t) H_{\mathbf{i}}(\boldsymbol{\xi}) H_{\mathbf{j}}(\boldsymbol{\xi}). \tag{4.27}$$

Similar to the univariate case, $\mathbf{v_i}(t)$ and $H_{\mathbf{i}}(\boldsymbol{\xi})$ are simplified as $\mathbf{v_i}$ and $H_{\mathbf{i}}$, respectively in following paragraphs. Then Galerkin projection (i.e., multiply both sides of each equation by $H_{\mathbf{k}}(\boldsymbol{\xi})$, $|\mathbf{k}| = 0, \dots, M$) is applied on both equations and obtain

$$\sum_{|\mathbf{i}|=0}^{M} \frac{dv_{\mathbf{i},1}}{dt} \langle H_{\mathbf{i}}, H_{\mathbf{k}} \rangle = -\frac{1}{2}\sum_{|\mathbf{i}|=0}^{M}\sum_{|\mathbf{j}|=0}^{M} v_{\mathbf{i},1}\, v_{\mathbf{j},2} \langle H_{\mathbf{i}} H_{\mathbf{j}}, H_{\mathbf{k}} \rangle, \tag{4.28}$$

$$\sum_{|\mathbf{i}|=0}^{M} \frac{dv_{\mathbf{i},2}}{dt} \langle H_{\mathbf{i}}, H_{\mathbf{k}} \rangle = \frac{1}{2}\sum_{|\mathbf{i}|=0}^{M}\sum_{|\mathbf{j}|=0}^{M} v_{\mathbf{i},1}\, v_{\mathbf{j},1} \langle H_{\mathbf{i}} H_{\mathbf{j}}, H_{\mathbf{k}} \rangle. \tag{4.29}$$

For a degree $|\mathbf{k}|$, the selection of $\mathbf{k}$ should be all additive partitions of $|\mathbf{k}|$ (same for $\mathbf{i}$ and $\mathbf{j}$). Same as univariate case, after calculating the inner products in equations (4.28) and (4.29), a set of equations for the expansion coefficients will be obtained. And the initial values for the coefficients can be acquired by the initial condition in (4.12) as,

$$\mathbf{v}_{[0,0]}(0) = [\mu_1(0), \mu_2(0)]^{\mathrm{T}} = [1.25, -0.35]^{\mathrm{T}},$$

$$\mathbf{v}_{[1,0]}(0) = [\sqrt{\sigma_1}, 0]^{\mathrm{T}} = [0.3,0]^{\mathrm{T}},$$

$$\mathbf{v}_{[0,1]}(0) = [0, \sqrt{\sigma_2}]^{\mathrm{T}} = [0,0.3]^{\mathrm{T}},$$

$$\mathbf{v_i}(0) = [0,0]^{\mathrm{T}}, |\mathbf{i}| = 2, \dots, M. \tag{4.30}$$

Note that $[0, 0]$ is the partition of $|\mathbf{i}| = 0$, and $[1, 0]$ and $[0, 1]$ are partitions of $|\mathbf{i}| = 1$.

The remaining task is to solve the equations (4.28) and (4.29) with initial condition given in (4.30) to get expansion coefficients at different times and afterwards a surrogate of the stochastic process $\mathbf{u}(t)$ is obtained.

Tables 4.8-4.13 give the values of the inner products $\langle H_{\mathrm{i}} H_{\mathrm{j}}, H_{\mathbf{k}} \rangle$ for order up to 2.

Table 4.8 Inner products $\langle H_i H_j, H_{0,0} \rangle$

	$H_{0,0}$	$H_{1,0}$	$H_{0,1}$	$H_{2,0}$	$H_{1,1}$	$H_{0,2}$
$H_{0,0}$	1	0	0	0	0	0
$H_{1,0}$	0	1	0	0	0	0
$H_{0,1}$	0	0	1	0	0	0
$H_{2,0}$	0	0	0	1	0	0
$H_{1,1}$	0	0	0	0	1	0
$H_{0,2}$	0	0	0	0	0	1

Table 4.9 Inner products $\langle H_i H_j, H_{1,0} \rangle$

	$H_{0,0}$	$H_{1,0}$	$H_{0,1}$	$H_{2,0}$	$H_{1,1}$	$H_{0,2}$
$H_{0,0}$	0	1	0	0	0	0
$H_{1,0}$	1	0	0	1.4142	0	0
$H_{0,1}$	0	0	0	0	1	0
$H_{2,0}$	0	1.4142	0	0	0	0
$H_{1,1}$	0	0	1	0	0	0
$H_{0,2}$	0	0	0	0	0	0

Table 4.10 Inner products $\langle H_i H_j, H_{0,1} \rangle$

	$H_{0,0}$	$H_{1,0}$	$H_{0,1}$	$H_{2,0}$	$H_{1,1}$	$H_{0,2}$
$H_{0,0}$	0	0	1	0	0	0
$H_{1,0}$	0	0	0	0	1	0
$H_{0,1}$	1	0	0	0	0	1.4142
$H_{2,0}$	0	0	0	0	0	0
$H_{1,1}$	0	1	0	0	0	0
$H_{0,2}$	0	0	1.4142	0	0	0

Table 4.11 Inner products $\langle H_i H_j, H_{2,0} \rangle$

	$H_{0,0}$	$H_{1,0}$	$H_{0,1}$	$H_{2,0}$	$H_{1,1}$	$H_{0,2}$
$H_{0,0}$	0	0	0	1	0	0
$H_{1,0}$	0	1.4142	0	0	0	0
$H_{0,1}$	0	0	0	0	0	0
$H_{2,0}$	1	0	0	2.8284	0	0
$H_{1,1}$	0	0	0	0	1.4142	0
$H_{0,2}$	0	0	0	0	0	1.3671e-11

Table 4.12 Inner products $\langle H_i H_j, H_{1,1} \rangle$

	$H_{0,0}$	$H_{1,0}$	$H_{0,1}$	$H_{2,0}$	$H_{1,1}$	$H_{0,2}$
$H_{0,0}$	0	0	0	0	1	0
$H_{1,0}$	0	0	1	0	0	0
$H_{0,1}$	0	1	0	0	0	0
$H_{2,0}$	0	0	0	0	1.4142	0
$H_{1,1}$	1	0	0	1.4142	0	1.4142
$H_{0,2}$	0	0	0	0	1.4142	0

Table 4.13 Inner products $\langle H_i H_j, H_{0,2} \rangle$

	$H_{0,0}$	$H_{1,0}$	$H_{0,1}$	$H_{2,0}$	$H_{1,1}$	$H_{0,2}$
$H_{0,0}$	0	0	0	0	0	1
$H_{1,0}$	0	0	0	0	0	0
$H_{0,1}$	0	0	1.4142	0	0	0
$H_{2,0}$	0	0	0	0	0	1.3671e-11
$H_{1,1}$	0	0	0	0	1.4142	0
$H_{0,2}$	1	0	0	1.3671e-11	0	2.8284

For example, when $M = 2$, define a new vector $\mathbf{V}(t)$ by concatenating vectors $\mathbf{v}_{[0,0]}(t)$, $\mathbf{v}_{[1,0]}(t)$, …, $\mathbf{v}_{[0,2]}(t)$ as

$$\mathbf{V}(t) = \left[\mathbf{v}_{[0,0]}^{\mathrm{T}}(t), \mathbf{v}_{[1,0]}^{\mathrm{T}}(t), \mathbf{v}_{[0,1]}^{\mathrm{T}}(t), \mathbf{v}_{[2,0]}^{\mathrm{T}}(t), \mathbf{v}_{[1,1]}^{\mathrm{T}}(t), \mathbf{v}_{[0,2]}^{\mathrm{T}}(t)\right]^{\mathrm{T}} = [V_1, V_2, \ldots, V_{12}]^{\mathrm{T}}, \tag{4.31}$$

that is,

$$[V_1\, V_2] = \mathbf{v}_{[0,0]}^{\mathrm{T}},\ [V_3\, V_4] = \mathbf{v}_{[1,0]}^{\mathrm{T}},\ \ldots,\ [V_{11}\, V_{12}] = \mathbf{v}_{[0,2]}^{\mathrm{T}}. \tag{4.32}$$

Then the resulting coefficient equations are

$$\frac{dV_1}{dt} = -\frac{1}{2}(V_1V_2 + V_3V_4 + V_5V_6 + V_7V_8 + V_9V_{10} + V_{11}V_{12}),$$

$$\frac{dV_2}{dt} = \frac{1}{2}(V_1^2 + V_3^2 + V_5^2 + V_7^2 + V_9^2 + V_{11}^2),$$

$$\frac{dV_3}{dt} = -\frac{1}{2}(V_1V_4 + V_3V_2 + 1.4142V_3V_8 + 1.4142V_7V_4 + V_5V_{10} + V_9V_6),$$

$$\frac{dV_4}{dt} = \frac{1}{2}(V_1V_3 + V_3V_1 + 1.4142V_3V_7 + 1.4142V_7V_3 + V_5V_9 + V_9V_5),$$

$$\frac{dV_5}{dt} = -\frac{1}{2}(V_1V_6 + V_5V_2 + 1.4142V_5V_{12} + 1.4142V_{11}V_6 + V_3V_{10} + V_9V_4),$$

$$\frac{dV_6}{dt} = \frac{1}{2}(V_1V_5 + V_5V_1 + 1.4142V_5V_{11} + 1.4142V_{11}V_5 + V_3V_9 + V_9V_3),$$

$$\frac{dV_7}{dt} = -\frac{1}{2}[V_1V_8 + V_7V_2 + 1.4142V_3V_4 + 2.8284V_7V_8 + (1.3671\text{e}-11)V_{11}V_{12} + 1.4142V_9V_{10}],$$

$$\frac{dV_8}{dt} = \frac{1}{2}[V_1V_7 + V_7V_1 + 1.4142V_3^2 + 2.8284V_7^2 + 1.4142V_9^2 + (1.3671\text{e}-11)V_{11}^2],$$

$$\frac{dV_9}{dt} = -\frac{1}{2}(V_1V_{10} + V_9V_2 + V_3V_6 + V_5V_4 + 1.4142V_7V_{10} + 1.4142V_9V_8 + 1.4142V_9V_{12} + 1.4142V_{11}V_{10}),$$

$$\frac{dV_{10}}{dt} = \frac{1}{2}(V_1V_9 + V_9V_1 + V_3V_5 + V_5V_3 + 1.4142V_7V_9 + 1.4142V_9V_7 + 1.4142V_9V_{11} + 1.4142V_{11}V_9),$$

$$\frac{dV_{11}}{dt} = -\frac{1}{2}[V_1V_{12} + V_{11}V_2 + 1.4142V_5V_6 + 1.4142V_9V_{10} + (1.3671\text{e}-11)V_7V_{12} + (1.3671\text{e}-11)V_{11}V_8 + 2.8284V_{11}V_{12}],$$

$$\frac{dV_{12}}{dt} = \frac{1}{2}[V_1V_{11} + V_{11}V_1 + 1.4142V_5^2 + 1.4142V_9^2 + (1.3671\text{e}-11)V_7V_{11} + (1.3671\text{e}-11)V_{11}V_7 + 2.8284V_{11}^2],$$

(4.33)

and the initial condition is

$$\mathbf{V}(0) = [1.25 \quad -0.35 \quad 0.3 \quad 0 \quad 0 \quad 0.3 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^{\mathrm{T}}. \qquad (4.34)$$

For comparison, classical Monte Carlo approach is also applied in this problem. The MC ensemble prediction is achieved by first creating a set of random samples of the initial condition given in (4.12). Each of the samples is propagated through the

dynamics presented in (4.7) and (4.8). The number of the samples is denoted as m, which ranges from 8 to 800,000 in the study. The statistics obtained through PC expansion using SG method is compared with those from the MC method and the exact solution. Tables 4.14-4.16 show the values of the first three moments of the amplitude pair at time $t = 2$ acquired from three different methods.

Table 4.14 First moments of PC, Exact and MC at $t = 2$ (two-variable model)

	SG			Exact	MC
	$N = 2$	$N = 1$			m = 8/80/800/8000/80000/800000
	$M = 2$	$M = 2$	$M = 4$		
μ_1	0.7989	0.8283	0.8278	0.7988	0.8326/0.7867/0.7907/0.7980/0.7989/0.7987
μ_2	1.0188	0.9731	0.9733	1.0189	0.8999/1.0545/1.0256/1.0178/1.0160/1.0181

Table 4.15 Second moments of PC, Exact and MC at $t = 2$ (two-variable model)

	SG			Exact	MC
	$N = 2$	$N = 1$			m = 8/80/800/8000/80000/800000
	$M = 2$	$M = 2$	$M = 4$		
σ_1	0.0234	0.0292	0.0291	0.0237	0.0232/0.0287/0.0225/0.0232/0.0237/0.0237
σ_2	0.1654	0.2027	0.2032	0.1652	0.2433/0.1260/0.1686/0.1651/0.1648/0.1647
σ_{12}	-0.0144	-0.0767	-0.0747	-0.0140	-0.0643/-0.0155/-0.0132/-0.0143/-0.0141/-0.0140
ρ	-0.2314	-0.9965	-0.9722	-0.2244	-0.8554/-0.2575/-0.2140/-0.2310/-0.2259/-0.2236

Table 4.16 Third moments of PC, Exact and MC at $t = 2$ (two-variable model)

	SG			Exact	MC
	$N = 2$	$N = 1$			m = 8/80/800/8000/80000/800000
	$M = 2$	$M = 2$	$M = 4$		
τ_{111}	0.0005	-0.0001	-0.00043	0.0010	-0.0021/0.0013/0.0008/0.0011/0.0011/0.0010
τ_{112}	-0.0011	-0.0014	-0.00003	-0.0019	0.0041/-0.0008/-0.0011/-0.0016/-0.0019/-0.0019
τ_{122}	-0.0050	0.0078	0.0045	-0.0052	-0.0012/-0.0013/-0.0089/-0.0049/-0.0051/-0.0052
τ_{222}	-0.0102	-0.0313	-0.0402	-0.0134	-0.0260/-0.0076/-0.0118/-0.0112/-0.0129/-0.0134

From the tables, it can be verified that the performance of MC approach is improved when increasing the number of ensemble members. Compared to the exact values, MC ensemble forecast using 800 or fewer samples is not good. When using 8000 samples, relatively good approximations can be obtained. However, there is still some bias.

When using 80000 samples or even more, estimates that are very close to the exact values will be obtained.

For the PC (denoted as SG here) method, when using one random variable ($N = 1$) to represent the randomness of the initial condition and truncating the polynomial expression at second order ($M = 2$), the estimates are far different from the exact values, especially the high order moments, e.g., the sign of the exact τ_{122} is negative, while the estimate is positive. Even when increasing the order of the polynomial to four ($M = 4$), i.e., the PC expansion is truncated at fourth order, there is little improvement. In contrast, if two random variables ($N = 2$) are used to represent the randomness of the initial condition and the truncation of PC expression is performed at second order ($M = 2$), good estimates can be achieved even though there are still some differences from the exact values. One may improve further by increasing the order of the truncation. Here, the performance of two-variable second-order polynomial chaos expansion (denoted as SG-M2-P2) is examined further by comparing the statistics for the first three moments with those from the exact solution at different times, as shown in Tables (4.17) – (4.20).

Table 4.17 Moments of SG-M2-P2 and Exact at $t = 1$ (two-variable model)

	μ_1	μ_2	σ_1	σ_2	σ_{12}	ρ	τ_{111}	τ_{112}	τ_{122}	τ_{222}
Exact	1.1902	0.4927	0.0578	0.1478	0.0308	0.3337	-0.0026	-0.0039	-0.0026	0.0008
SG-M2-P2	1.1903	0.4927	0.0577	0.1478	0.0308	0.3332	-0.0028	-0.0037	-0.0027	0.0018

Table 4.18 Moments of SG-M2-P2 and Exact at $t = 3$ (two-variable model)

	μ_1	μ_2	σ_1	σ_2	σ_{12}	ρ	τ_{111}	τ_{112}	τ_{122}	τ_{222}
Exact	0.4618	1.2253	0.0176	0.1329	-0.0316	-0.6542	0.0015	-0.0016	0.0009	-0.0150
SG-M2-P2	0.4620	1.2252	0.0174	0.1330	-0.0323	-0.6712	0.0012	-0.0013	0.0011	-0.0099

Table 4.19 Moments of SG-M2-P2 and Exact at $t = 5$ (two-variable model)

	μ_1	μ_2	σ_1	σ_2	σ_{12}	ρ	τ_{111}	τ_{112}	τ_{122}	τ_{222}
Exact	0.1422	1.3192	0.0066	0.0978	-0.0214	-0.8377	0.0008	-0.0017	0.0031	-0.0055
SG-M2-P2	0.1425	1.3192	0.0065	0.0977	-0.0215	-0.8493	0.0007	-0.0014	0.0025	-0.0004

Table 4.20 Moments of SG-M2-P2 and Exact at $t = 10$ (two-variable model)

	μ_1	μ_2	σ_1	σ_2	σ_{12}	ρ	τ_{111}	τ_{112}	τ_{122}	τ_{222}
Exact	0.0095	1.3330	0.0003	0.0879	-0.0033	-0.6826	0.00003	-0.0002	0.0011	-0.0001
SG-M2-P2	0.0094	1.3326	0.0002	0.0888	-0.0030	0.8163	0.000004	-0.0001	0.0007	0.0034

From the tables, SG-M2-P2 has very good estimates at early times. It can be used as a surrogate of the dynamic model. When time evolves, the estimates becomes worse especially for higher order moments, for example at time $t = 10$, the sign of τ_{222} is different from the exact value. However, they still have good estimates on the first two moments. Again, one may use a higher order polynomial expansion, or use the same order but more collocation points to capture the randomness (the SC method will be discussed in next chapter).

In addition, ensemble forecast using MC approach with 800,000 or more samples has very close estimates to the exact values. The histograms using 800,000 samples from MC and PC (SG-M2-P2) approach have been examined further, as shown in Figures (4.1) to (4.4), with MC on the left and PC on the right. In consistence with the analysis from Tables (4.17) – (4.20), PC approach has good approximation at early times. There will be deviations when time evolves, for example when $t = 5$ or even larger $t = 10$, two variables u_1 and u_2 obtained from PC approach have different range and probability values from those of MC approach.

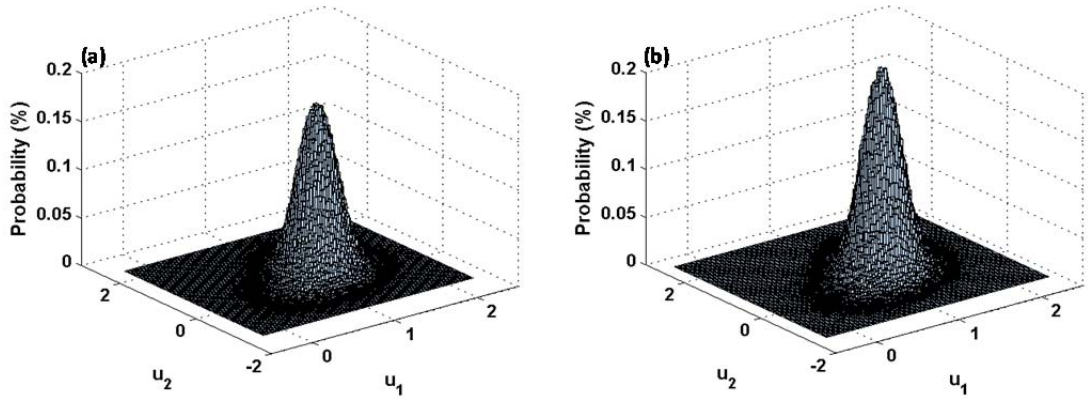


Figure 4.1 Histogram at $t = 1$ (two-variable model) (a) MC (b) PC

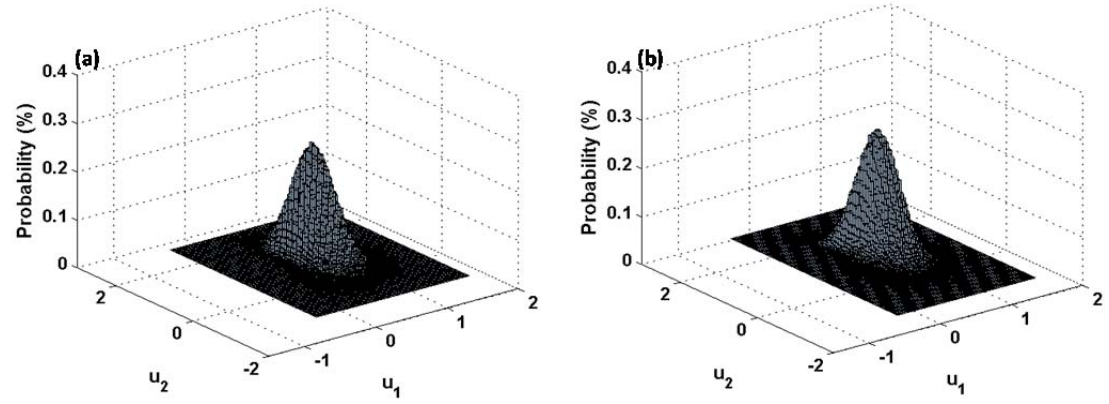


Figure 4.2 Histogram at $t = 2$ (two-variable model) (a) MC (b) PC

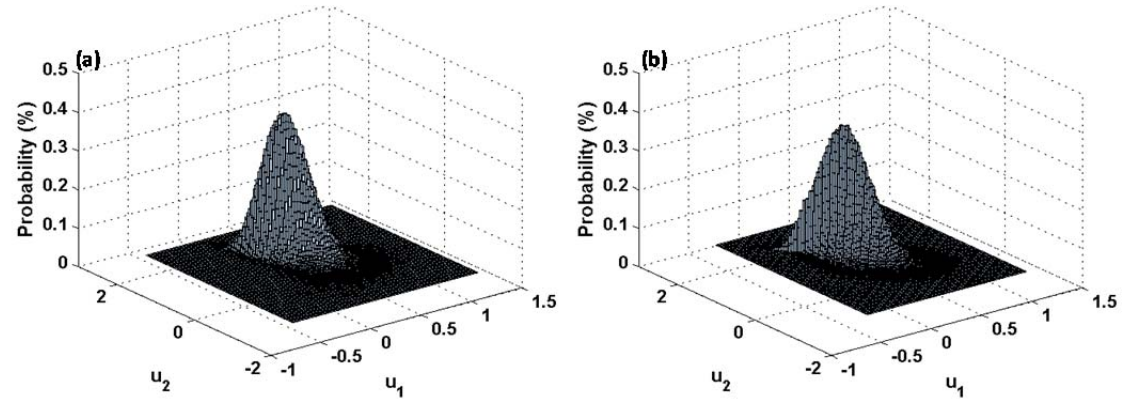


Figure 4.3 Histogram at $t = 3$ (two-variable model) (a) MC (b) PC

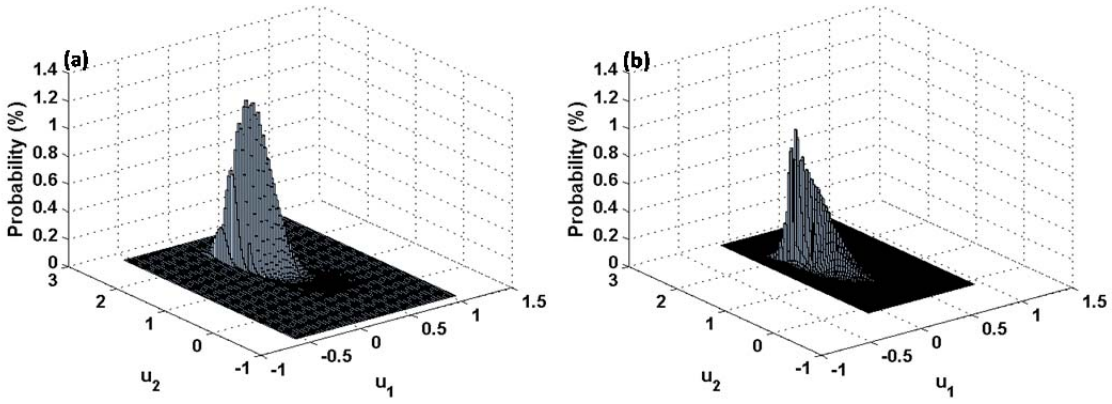


Figure 4.4 Histogram at $t = 5$ (two-variable model) (a) MC (b) PC

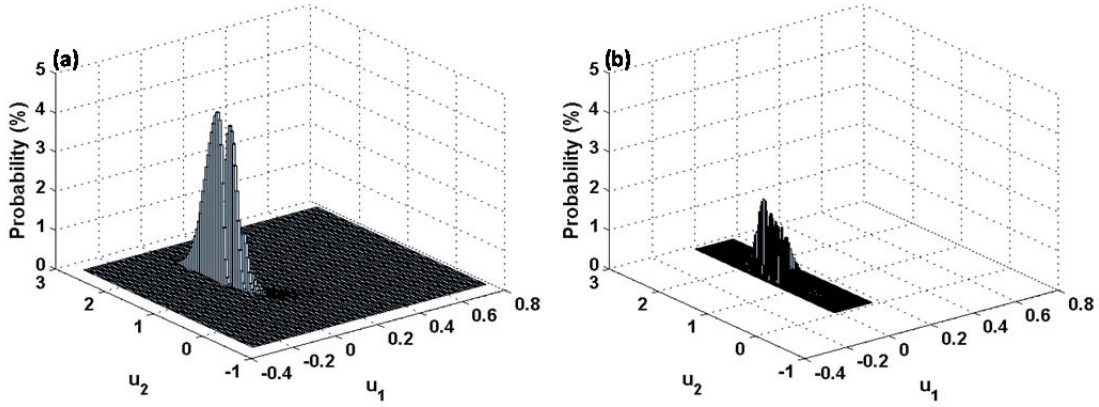


Figure 4.5 Histogram at $t = 10$ (two-variable model) (a) MC (b) PC

Figures (4.6) and (4.7) are the evolution of mean values and standard deviations of the amplitude pair derived from PC (SG-M2-P2) approach.

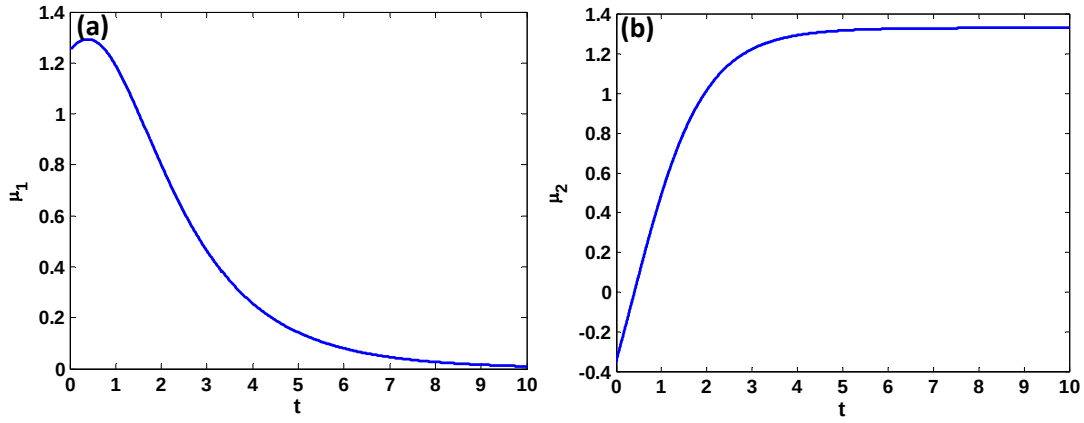


Figure 4.6 Evolution of mean values of the amplitude pair derived from PC (two-variable model) (a) μ_1 (b) μ_2

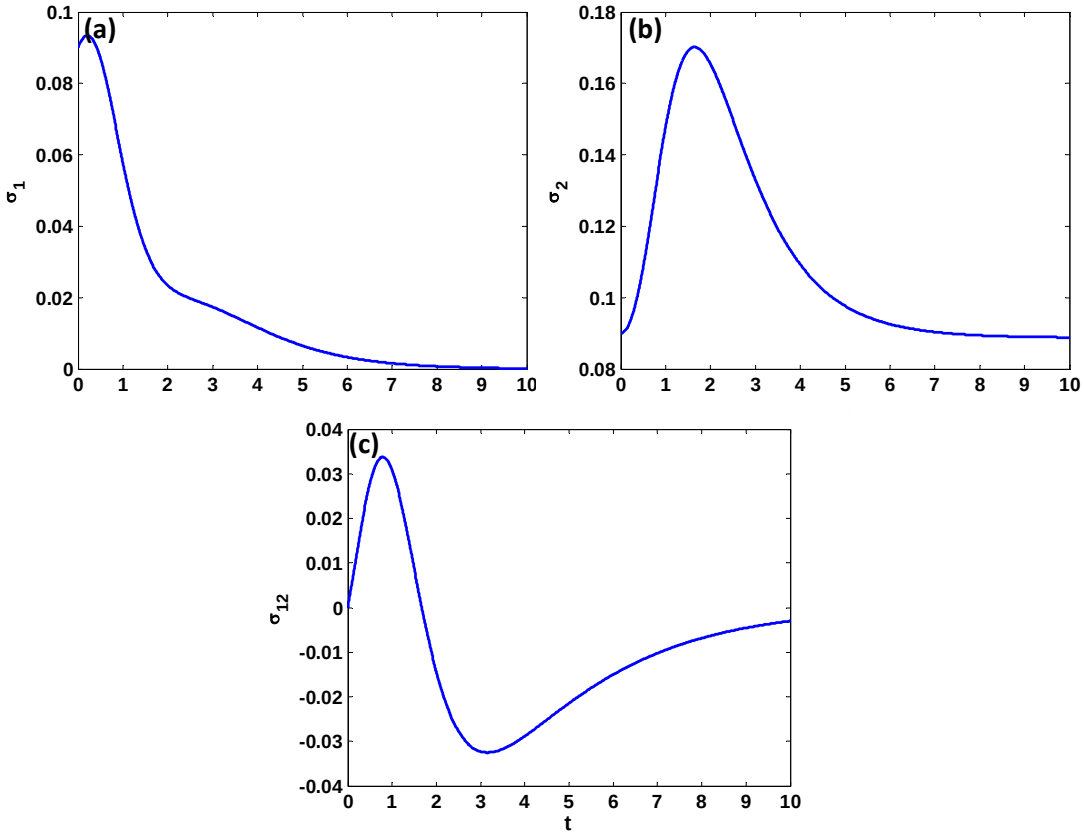


Figure 4.7 Evolution of second moments of the amplitude pair derived from PC (two-variable model) (a) σ_1 (b) σ_2 (c) σ_{12}

4.4 Discussions

Using a two-variable model with bivariate normally distributed initial condition, this chapter studied the ability of PC approach in uncertainty quantification. Specifically, stochastic Galerkin projection was examined to obtain the expansion coefficients in PC expansion. In SG, the original governing equations need to be modified to a set of equations with the expansion coefficients as unknowns. For completeness, both the scalar Hermite polynomial chaos expansion and multivariate Hermite polynomial chaos expansion were studied, and they were compared to MC simulation and the exact solution. The results show that when truncating at the same order, two-variate PC expansion outperforms scalar PC expansion. Using scalar PC expansion, the improvement is negligible when increasing the truncation order from 2 to 4 in the example discussed in this chapter. The 2-variate second order PC expansion (SG-M2-P2) was investigated further at different times in terms of first, second, third-order moments and the histograms. The results show that SG-M2-P2 gives very good estimates at early times. With time evolving, the performance becomes worse.

Chapter 5

Application of Stochastic Collocation Method

Using PC expansion to quantify the forecast uncertainty is studied further in this chapter. Other than SG method discussed in last chapter, the focus is put on using stochastic Collocation (SC) method to obtain the expansion coefficients in PC expansion. A relatively more complex model namely the five-variable mixed-layer model describing the return flow event over the Gulf of Mexico will be used in this chapter. Lewis et al. (2015) has performed a complete study on ensemble forecasting based on the classical Monte Carlo scheme. The input uncertainty includes (1) initial condition (IC) with Gaussian distribution, (2) boundary condition (BC) with Gaussian distribution and (3) parameters with uniform distribution. Four experiments focusing on different input uncertainties have been performed. These experiments are grouped as (1) IC only, which means only considering uncertainty in initial conditions, BC and parameter use their mean values separately; (2) BC only, which means only considering uncertainty in boundary conditions, IC and parameter use their mean values separately; (3) Parameter only, which means only considering uncertainty in parameters, IC and BC use their mean values separately; (4) FC (Full complement) which means considering uncertainty in all random inputs (including IC, BC and parameters). This study will show the performance of PC expansion (specifically using SC method to obtain the expansion coefficients in PC expansion) in comparison with MC using IC only and parameter only cases.

5.1 The Mixed-layer Model

In each year, a rhythmic cycle of cold air penetrates into the Gulf of Mexico (GoM) during the cool season (late fall and winter). After warmed up by the sea surface, a return of modified air generally follows the penetrations to land in response to the circulation around an eastward-moving cold anticyclone. This large-scale process is termed as "return flow" (Henry 1979a, 1979b). Typically, there are 4–5 return-flow events (RFE's) taking place over the Gulf of Mexico each month between November and March (Crisp and Lewis 1992) every year.

In 1988 and 1991 (Lewis et al. 1989), field exercises to study the RFE's have been taken during the GUFMEX (Gulf of Mexico Experiment) project. The difficulties to forecast the water-vapor mixing ratio in RFE over the GoM have been well documented in the literature over the past several decades (e.g., Janish and Lyons 1992; Weiss 1992; Thompson et al. 1994; Edwards and Weiss 1995; Manikin et al. 2000, 2001, 2002). The following factors have been conjectured to contribute to forecast errors: (1) absence of routine upper-air observations over the Gulf, (2) absence of dewpoint (moisture) measurements on tethered buoys over the Gulf's shelf water, (3) errors in sea-surface temperature (SST) due to aged data in response to cloud cover, and (4) inaccuracy in the operational model parameterizations of moisture and heat fluxes at the sea-air interface. Bias, both positive and negative, has also plagued the operational numerical prediction of the mixing ratio and this aspect of the problem has been especially problematical for forecasters. The consequence of poor guidance is extreme where forecasts can range from sea fog and stratus cloud when vapor content is low, to shallow cumulus with light

showers for intermediate values of the moisture, to cumulonimbus and associated severe weather for large-magnitude vapor mixing ratios (Lewis et al. 2015).

In this research, efforts have been made to uncover the sources of these forecast errors through use of Monte Carlo (MC), polynomial chaos (PC) expansion and unscented transformation (UT) methodologies with a dynamical model where elements of control (initial conditions, boundary conditions, and parameters) are randomly chosen. The forecast will be restricted to the "outflow phase" of the RFE where buoyancy at the sea–air interface drives the heating and moistening of the lower-tropospheric layer. The classic mixed-layer model (Ball 1960; Lilly 1968, Carson 1973; Tennekes and Driedonks 1981) has been found to faithfully describe the airmass modification in those RFE situations in (Liu et al. 1992; Burk and Thompson 1992; Lewis and Crisp 1992; Lewis 2007).

This dissertation studies a case which took place in late February 1988 during GUFMEX. It is a representative of deep penetrating RFE's and an excellent set of upper-air observations are available during this period. The detailed description of the surface, upper-air and satellite observations associated with the RFE and the analysis of the single trajectory used in the study are from (Lewis et al. 2015). The main objective of this research is to compare the ability of quantifying the forecast uncertainty using PC (in this chapter) and UT (in next chapter) with that from MC. Therefore, it is assumed that the same observations and trajectory used for the ensemble forecast with MC in (Lewis et al 2015) are adopted in this research. Hu et al. (2015) presents some results for IC only case. The observations are given in Table 5.1.

Table 5.1 Upper-air Observations at $t = 0$, $t = 6h$ and $t = 9h$

Time	θ(℃)	$h(km)$	σ(℃)	$q(g/kg)$	$\mu(g/kg)$
$t = 0$	14.50	0.90	0.50	4.50	-1.50
$t = 6h$	16.50	1.50	—	6.50	—
$t = 9h$	17.50	1.70	—	8.00	—

The upper-air observations at initial time, i.e., $t = 0$ (1800 UTC Feb 21, 1988), are from U. S. Coast Guard ship Salvia and the observations at $t = 6h$ and $t = 9h$ are from NOAA P-3 aircraft. No observations of σ and μ are obtained from the P-3 aircraft. The reason is that there is evident presence of stable and dry layers in these P-3 profiles. It is difficult to detect jumps, let alone place values on them.

Based on the observational evidences collected, a mixed-layer model which is driven by buoyancy in response to heat flux at the lower boundary can be justified to explore the development of a convective layer over the sea. The atmospheric model consists of 3 layers: (1) a near-surface layer with a thermodynamically unstable structure (~50-100 m deep), (2) a deeper convective layer with nearly uniform distributions of the variables, and (3) a non-turbulent, stably stratified layer overlying the convective layer. A schematic diagram displaying processes in the atmospheric mixed layer is shown in Figure 5.1. The tacit assumption is that the column of air remains intact as it moves over the sea surface, that is, differential speed and direction of the wind are sufficiently small such that the column remains erect. The description of the RFE and the mixed-layer model are parts from the joint paper (Lewis 2015).

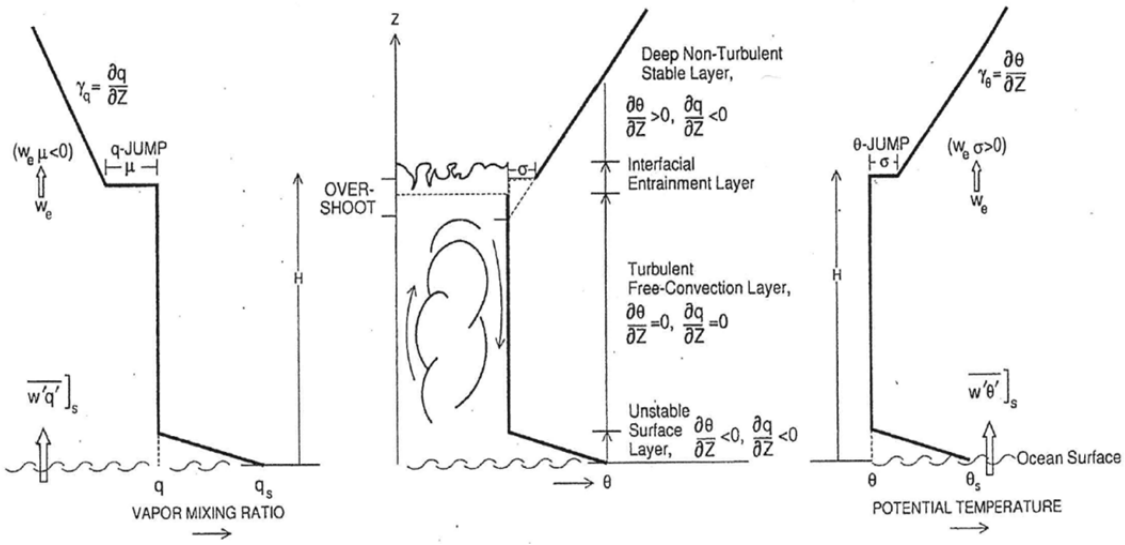


Figure 5.1 A schematic diagram of the idealized mixed layer profiles of potential temperature and mixing ratio where basic variables are identified and symbolized (Lewis et al. 2015).

The symbols in Figure 5.1 represent the following variables and parameters:

Forecast variables:

θ: Potential temperature in the mixed layer

h: Height of the mixed layer

σ: Potential-temperature jump atop the mixed layer

q: Vapor mixing ratio in the mixed layer

μ: Mixing-ratio jump atop the mixed layer

The physical and empirical parameters:

w: Large-scale subsidence

w_e: Turbulent (entrainment) velocity

C_θ: Bulk exchange coefficient for heat

C_q: Bulk exchange coefficient for moisture

V_s: Translation speed of the column

γ_θ: Lapse rate of potential temperature in the stable layer

γ_q: Lapse rate of vapor mixing ratio in the stable layer

κ: Entrainment coefficient

$\overline{w'\theta'}]_s$: Transport of heat into the mixed layer from below

$\overline{w'q'}]$: Transport of moisture into the mixed layer from below

Boundary conditions:

θ_s: Potential temperature at the air-sea interface

q_s: Saturated vapor mixing ratio at the air-sea interface

The governing equations for the mixed-layer model take the form:

$$\frac{d\theta}{dt} = C_\theta V_s(1+\kappa)(\theta_s - \theta)h^{-1}, \tag{5.1}$$

$$\frac{dh}{dt} = \kappa(C_\theta V_s)(\theta_s - \theta)\sigma^{-1} + w, \tag{5.2}$$

$$\frac{d\sigma}{dt} = \gamma_\theta \frac{dh}{dt} - \frac{d\theta}{dt} - \gamma_\theta w, \tag{5.3}$$

$$\frac{dq}{dt} = C_q V_s \left[(q_s - q) + \frac{\mu\kappa}{\sigma}(\theta_s - \theta)\right] h^{-1}, \tag{5.4}$$

$$\frac{d\mu}{dt} = \gamma_q \frac{dh}{dt} - \frac{dq}{dt} - \gamma_q w. \tag{5.5}$$

The jumps σ and μ are classified as "secondary variables" since their evolution is dependent on $\frac{d\theta}{dt}$, $\frac{dh}{dt}$ and $\frac{dq}{dt}$. A detailed development of these equations can be found in (Lewis 2007) and the pioneering work can be found in (Lilly 1968, 1987) for the case of a cloud-free mixed layer.

The control elements for the mixed-layer model consist of (1) initial condition (IC, the initial values for the forecast variables in the model, the initial time is 1800 UTC 21 Feb, 1988), (2) several parameters in the model and (3) boundary conditions (BC, the potential temperature at the air-sea interface θ_s or denoted as SST and the saturated water-vapor mixing ratio q_s at the air-sea interface). The dimension of the control vector for the mixed-layer model is 45:5 initial conditions, a total of 34 boundary conditions (17 θ_s s and 17 q_ss) and 6 parameters. Each element of the control vector is represented

by either a normal probability density function (IC's and BC's) or a uniform probability density function (parameters). The uniform distribution is used for parameters to avoid the random choice of physically unrealistic parameters, i.e., negative values for κ and γ_θ, and positive values for w and γ_q. The mean values as well as standard deviations for the IC's and BC's are found in Tables 5.2 and 5.4, respectively, and the means and ranges for the parameters are found in Table 5.3.

Table 5.2 Mean values and standard deviations for mixed-layer model initial conditions

	θ(℃)	$h(km)$	σ(℃)	$q(g/kg)$	$\mu(g/kg)$
Mean	14.5	0.90	0.50	4.50	-1.50
Standard deviation	1.0	0.075	0.20	0.50	0.50

Table 5.3 Mean values and ranges for mixed-layer model parameters

Parameter	**Mean Values**	**Range**
w *(cm* s^{-1}*)*	–0.50	(–0.10) – (– 0.90)
κ *(*non-dimensional*)*	0.25	0.20 – 0.30
$V_s C_\theta (m \cdot s^{-1})$	1.25×10^{-2}	1×10^{-2} – 1.5×10^{-2}
$V_s C_q (m \cdot s^{-1})$	1.25×10^{-2}	1×10^{-2} – 1.5×10^{-2}
γ_θ(℃ $\cdot km^{-1}$)	6.0	5.0 – 7.0
$\gamma_q (g \cdot kg^{-1} \cdot km^{-1})$	–2 .0	(–1.0) – (–3.0)

Table 5.4 Mean values and standard deviations for mixed-layer model boundary conditions

Time	t**: model time (h)**	$\theta_s(^\circ C)$	$q_s(gkg^{-1})$
1800 UTC Feb 21	0	20.8	14.92
1900 UTC Feb 21	1	21.4	15.48
2000 UTC Feb 21	2	22.0	16.06
2100 UTC Feb 21	3	23.0	17.06
0000 UTC Feb 22	6	24.0	18.12
0300 UTC Feb 22	9	25.0	19.24
0400 UTC Feb 22	10	26.0	20.42
0500 UTC Feb 22	11	26.1	20.54
0600 UTC Feb 22	12	26.1	20.54

1200 UTC Feb 22	18	24.2	18.34
1800 UTC Feb 22	24	23.5	17.59
0000 UTC Feb 23	30	24.2	18.34
0600 UTC Feb 23	36	23.1	17.17
1200 UTC Feb 23	42	23.1	17.17
1800 UTC Feb 23	48	22.7	16.76
0000 UTC Feb 24	54	22.2	16.26
0300 UTC Feb24	57	22.0	16.06

Standard deviations: at all times, θ_s +/- $1^\circ C$, q_s +/- $1\, gkg^{-1}$

5.2 Initial Condition Only

In initial condition (IC) only case, the initial conditions are assumed to follow Gaussian distribution with the mean and standard deviation values given in Table 5.2. There are five variables and they are assumed independent with each other. The parameters and boundary conditions use the mean values in Tables 5.3 and 5.4, respectively. As discussed in Chapter 2, Hermite polynomial is the best choice for Gaussian distribution.

Let $\mathbf{x} = (\theta, h, \sigma, q, \mu)^{\mathrm{T}}$, the aim is to seek an approximation $\mathbf{x}_N^M(t, \boldsymbol{\xi})$ of $\mathbf{x}(t)$ in the form of PC expression

$$\mathbf{x}_N^M(t, \boldsymbol{\xi}) = \sum_{|\mathbf{i}|=0}^{M} \mathbf{v}_{\mathbf{i}}(t)\Phi_{\mathbf{i}}(\boldsymbol{\xi}), \tag{5.6}$$

where $\{\mathbf{v}_{\mathbf{i}}(t) = [v_{\mathbf{i},1}(t), \dots, v_{\mathbf{i},5}(t)]^{\mathrm{T}}\} \in R^5$ are the expansion coefficients. N is the number of random variables used in the expansion, and M is the highest order of the polynomials. $\Phi_{\mathbf{i}}(\boldsymbol{\xi})$ will be the Hermite polynomials in this case. Appendix F shows the process of using SG projection to acquire the equations for the coefficients of PC expansion with univariate Hermite polynomials and it is seen that the process is tedious and the resulted equation is complicated especially for a complex system. Now let's turn to SC method to obtain the coefficients without altering the original equations.

In the experiment, five-dimensional variable $\boldsymbol{\xi} = (\xi_1, \dots, \xi_5)^{\mathrm{T}}$ is used in the PC expansion, i.e., $N = 5$. The elements of $\boldsymbol{\xi}$ are independent with each other. As an example, the PC expansion is truncated at degree $M = 2$, i.e., a second-order polynomial expression for $\mathbf{x}(t)$, and there are altogether $\frac{(5+2)!}{5!\times 2!} = 21$ polynomial terms (shown in Table 5.5) in the expansion. Suppose the m-th order normalized scalar Hermite polynomial at random variable ξ_i is denoted as $H_m(\xi_i)$, then $H_m(\xi_i), m = 0, 1, 2$ are $1, \xi_i$, and $\frac{(\xi_i^2-1)}{\sqrt{2}}$ separately.

Table 5.5 Five-variable normalized Hermite polynomials (order no greater than 2)

Degree m	**Multi index** $(p_1\ p_2\ p_3\ p_4\ p_5)$	$H_m(\boldsymbol{\xi})$	$H_{p_1}(\xi_1)H_{p_2}(\xi_2)H_{p_3}(\xi_3)H_{p_4}(\xi_4)H_{p_5}(\xi_5)$
0	(0 0 0 0 0)	1	$H_0(\xi_1)H_0(\xi_2)H_0(\xi_3)H_0(\xi_4)H_0(\xi_5)$
1	(1 0 0 0 0)	ξ_1	$H_1(\xi_1)H_0(\xi_2)H_0(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(0 1 0 0 0)	ξ_2	$H_0(\xi_1)H_1(\xi_2)H_0(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(0 0 1 0 0)	ξ_3	$H_0(\xi_1)H_0(\xi_2)H_1(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(0 0 0 1 0)	ξ_4	$H_0(\xi_1)H_0(\xi_2)H_0(\xi_3)H_1(\xi_4)H_0(\xi_5)$
	(0 0 0 0 1)	ξ_5	$H_0(\xi_1)H_0(\xi_2)H_0(\xi_3)H_0(\xi_4)H_1(\xi_5)$
2	(2 0 0 0 0)	$(\xi_1^2 - 1)/\sqrt{2}$	$H_2(\xi_1)H_0(\xi_2)H_0(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(1 1 0 0 0)	$\xi_1\xi_2$	$H_1(\xi_1)H_1(\xi_2)H_0(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(1 0 1 0 0)	$\xi_1\xi_3$	$H_1(\xi_1)H_0(\xi_2)H_1(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(1 0 0 1 0)	$\xi_1\xi_4$	$H_1(\xi_1)H_0(\xi_2)H_0(\xi_3)H_1(\xi_4)H_0(\xi_5)$
	(1 0 0 0 1)	$\xi_1\xi_5$	$H_1(\xi_1)H_0(\xi_2)H_0(\xi_3)H_0(\xi_4)H_1(\xi_5)$
	(0 2 0 0 0)	$(\xi_2^2 - 1)/\sqrt{2}$	$H_0(\xi_1)H_2(\xi_2)H_0(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(0 1 1 0 0)	$\xi_2\xi_3$	$H_0(\xi_1)H_1(\xi_2)H_1(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(0 1 0 1 0)	$\xi_2\xi_4$	$H_0(\xi_1)H_1(\xi_2)H_0(\xi_3)H_1(\xi_4)H_0(\xi_5)$
	(0 1 0 0 1)	$\xi_2\xi_5$	$H_0(\xi_1)H_1(\xi_2)H_0(\xi_3)H_0(\xi_4)H_1(\xi_5)$
	(0 0 2 0 0)	$(\xi_3^2 - 1)/\sqrt{2}$	$H_0(\xi_1)H_0(\xi_2)H_2(\xi_3)H_0(\xi_4)H_0(\xi_5)$
	(0 0 1 1 0)	$\xi_3\xi_4$	$H_0(\xi_1)H_0(\xi_2)H_1(\xi_3)H_1(\xi_4)H_0(\xi_5)$
	(0 0 1 0 1)	$\xi_3\xi_5$	$H_0(\xi_1)H_0(\xi_2)H_1(\xi_3)H_0(\xi_4)H_1(\xi_5)$
	(0 0 0 2 0)	$(\xi_4^2 - 1)/\sqrt{2}$	$H_0(\xi_1)H_0(\xi_2)H_0(\xi_3)H_2(\xi_4)H_0(\xi_5)$
	(0 0 0 1 1)	$\xi_4\xi_5$	$H_0(\xi_1)H_0(\xi_2)H_0(\xi_3)H_1(\xi_4)H_1(\xi_5)$
	(0 0 0 0 2)	$(\xi_5^2 - 1)/\sqrt{2}$	$H_0(\xi_1)H_0(\xi_2)H_0(\xi_3)H_0(\xi_4)H_2(\xi_5)$

Recall that the dimension of the state vector is five and the elements of the initial conditions are independent with each other. Since five independent random variables are used, each of which can represents the randomness at each dimension. According to the distribution of IC, the coefficients are easily obtained

$$\mathbf{v}_{[0,0,0,0,0]}(0) = (14.5, 0.90, 0.50, 4.50, -1.50)^T,$$

$$\mathbf{v}_{[1,0,0,0,0]}(0) = (1.0, 0, 0, 0, 0)^T,$$

$$\mathbf{v}_{[0,1,0,0,0]}(0) = (0, 0.075, 0, 0, 0)^T,$$

$$\mathbf{v}_{[0,0,1,0,0]}(0) = (0, 0, 0.20, 0, 0)^T,$$

$$\mathbf{v}_{[0,0,0,1,0]}(0) = (0, 0, 0, 0.50, 0)^T,$$

$$\mathbf{v}_{[0,0,0,0,1]}(0) = (0, 0, 0, 0, 0.50)^T,$$

$$\mathbf{v}_{\mathbf{i}}(0) = (0, 0, 0, 0, 0)^T, |\mathbf{i}| = 2. \tag{5.7}$$

Therefore the PC expansion for the initial condition is as follows:

$$\mathbf{x}(0) = \begin{pmatrix} 14.5 \\ 0.90 \\ 0.50 \\ 4.50 \\ -1.50 \end{pmatrix} + \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0 & 0.075 & 0 & 0 & 0 \\ 0 & 0 & 0.20 & 0 & 0 \\ 0 & 0 & 0 & 0.50 & 0 \\ 0 & 0 & 0 & 0 & 0.50 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix}. \tag{5.8}$$

The expansion coefficients at any time t in this chapter are obtained through SC method. There are many ways to select collocation points and estimate the coefficients. In this study, the Gaussian-Hermite quadrature rule and sparse grid scheme discussed in Chapter 2 are used. In the experiment, when approximating the equation (2.32) through Gaussian-Hermite quadrature, the random variable follows the standard Gaussian distribution and the weight function is the weighting function for standard Gaussian distribution. As an example, when the exact level is 2, i.e., $K = 2$, the 11 collocation points and corresponding weights listed in Table 5.6 are used.

Table 5.6 Sparse collocation points with weights (dimension 5, exact level 2), Gaussian-Hermite quadrature rule

No.	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	Weight
1	-1.0000	0	0	0	0	0.5000
2	0	-1.0000	0	0	0	0.5000
3	0	0	-1.0000	0	0	0.5000
4	0	0	0	-1.0000	0	0.5000
5	0	0	0	0	-1.0000	0.5000
6	0	0	0	0	0	-4.0000
7	0	0	0	0	1.0000	0.5000
8	0	0	0	1.0000	0	0.5000
9	0	0	1.0000	0	0	0.5000
10	0	1.0000	0	0	0	0.5000
11	1.0000	0	0	0	0	0.5000

When $K = 3$, there will be 61 collocation points which are listed in Table 5.7.

Table 5.7 Sparse collocation points with weights (dimension 5, exact level 3), Gaussian-Hermite quadrature rule

No.	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	Weight
1	-1.7321	0	0	0	0	0.1667
2	-1	-1	0	0	0	0.2500
3	-1	0	-1	0	0	0.2500
4	-1	0	0	-1	0	0.2500
5	-1	0	0	0	-1	0.2500
6	-1	0	0	0	0	-2.0000
7	-1	0	0	0	1	0.2500
8	-1	0	0	1	0	0.2500
9	-1	0	1	0	0	0.2500
10	-1	1	0	0	0	0.2500
11	0	-1.7321	0	0	0	0.1667
12	0	-1	-1	0	0	0.2500
13	0	-1	0	-1	0	0.2500
14	0	-1	0	0	-1	0.2500
15	0	-1	0	0	0	-2.0000
16	0	-1	0	0	1	0.2500

17	0	-1	0	1	0	0.2500
18	0	-1	1	0	0	0.2500
19	0	0	-1.7321	0	0	0.1667
20	0	0	-1	-1	0	0.2500
21	0	0	-1	0	-1	0.2500
22	0	0	-1	0	0	-2.0000
23	0	0	-1	0	1	0.2500
24	0	0	-1	1	0	0.2500
25	0	0	0	-1.7321	0	0.1667
26	0	0	0	-1	-1	0.2500
27	0	0	0	-1	0	-2.0000
28	0	0	0	-1	1	0.2500
29	0	0	0	0	-1.7321	0.1667
30	0	0	0	0	-1	-2.0000
31	0	0	0	0	0	9.3333
32	0	0	0	0	1	-2.0000
33	0	0	0	0	1.7321	0.1667
34	0	0	0	1	-1	0.2500
35	0	0	0	1	0	-2.0000
36	0	0	0	1	1	0.2500
37	0	0	0	1.7321	0	0.1667
38	0	0	1	-1	0	0.2500
39	0	0	1	0	-1	0.2500
40	0	0	1	0	0	-2.0000
41	0	0	1	0	1	0.2500
42	0	0	1	1	0	0.2500
43	0	0	1.7321	0	0	0.1667
44	0	1	-1	0	0	0.2500
45	0	1	0	-1	0	0.2500
46	0	1	0	0	-1	0.2500
47	0	1	0	0	0	-2.0000
48	0	1	0	0	1	0.2500
49	0	1	0	1	0	0.2500
50	0	1	1	0	0	0.2500
51	0	1.7321	0	0	0	0.1667

52	1	-1	0	0	0	0.2500
53	1	0	-1	0	0	0.2500
54	1	0	0	-1	0	0.2500
55	1	0	0	0	-1	0.2500
56	1	0	0	0	0	-2.0000
57	1	0	0	0	1	0.2500
58	1	0	0	1	0	0.2500
59	1	0	1	0	0	0.2500
60	1	1	0	0	0	0.2500
61	1.7321	0	0	0	0	0.1667

Using SC method, the initial values at selected collocation points need to be evaluated firstly by using expression (5.8), and then they are propagated through the dynamic model given by equations (5.1)-(5.5). Together with the corresponding weights at selected collocation points, the coefficients are then obtained through the approximation given by equation (2.32).

In order to examine the effectiveness of PC approach, the first two moments obtained through PC approach are compared with those from the classical Monte Carlo (MC) approach. The number of ensemble for MC approach is 20,000 from Lewis (2015), which is determined by the stableness of the method. Figures 5.2 and 5.3 below show the evolution of mean values and standard deviations of the five variables (here the exact level $K = 2$, 11 collocation points) by using PC and MC approach separately. The solid line represents the simulation by MC approach and the dashed line is from PC approach. The plus signs represent the observations from Table 5.1. As can be seen from the figures, the differences of the mean values for the five variables from both methods are very little and they are hard to tell. There are some differences between standard deviations, but again, they are small. One

may reduce the difference by truncating the PC expansion at a higher order or add more collocation points. Some experiments have been conducted and the improvement is not so remarkable since a relatively good estimate has already achieved. Table (5.8) further presents the covariance matrices at some time slots. From the table, it is seen that PC expansion can approximate the covariance matrix very well, which means it does not only approximate the standard deviation (or variance of each variable) well which is shown from Figure 5.3, it can also make good estimates on the covariance between different variables. It is known that the background error covariance matrix plays a very important role in data assimilation procedure and is usually approximated at some degree for large scale system because of the computational cost. Therefore, the good estimate of the forecast covariance matrix through PC approach may provide an alternative selection for data assimilation.

After solving the expansion coefficients, now a second order polynomial approximation for the stochastic state vector $\mathbf{x}(t)$ in terms of the standard Gaussian random variable $\boldsymbol{\xi}$ is obtained. By drawing the samples of the random variable $\boldsymbol{\xi}$, one can readily generate the ensemble members of $\mathbf{x}(t)$ at any time t and further the histogram can be constructed. In the experiment, 20,000 samples of $\boldsymbol{\xi}$ that correspond to the initial values in MC approach are used, i.e., both PC and MC start from the same initial conditions. Figures 5.4-5.12 show the histograms of θ, h and q at time $t = 1h, 24h, 48h$ with MC on the left and PC on the right. Overall, both PC and MC produce similar histograms especially when $t = 1h$. With the time evolving, there is small difference appearing. For example, when $t = 48h$, the peak values of q for MC

are higher than those for PC, however, the overall distributions are similar to each other and the differences of the peak values are small.

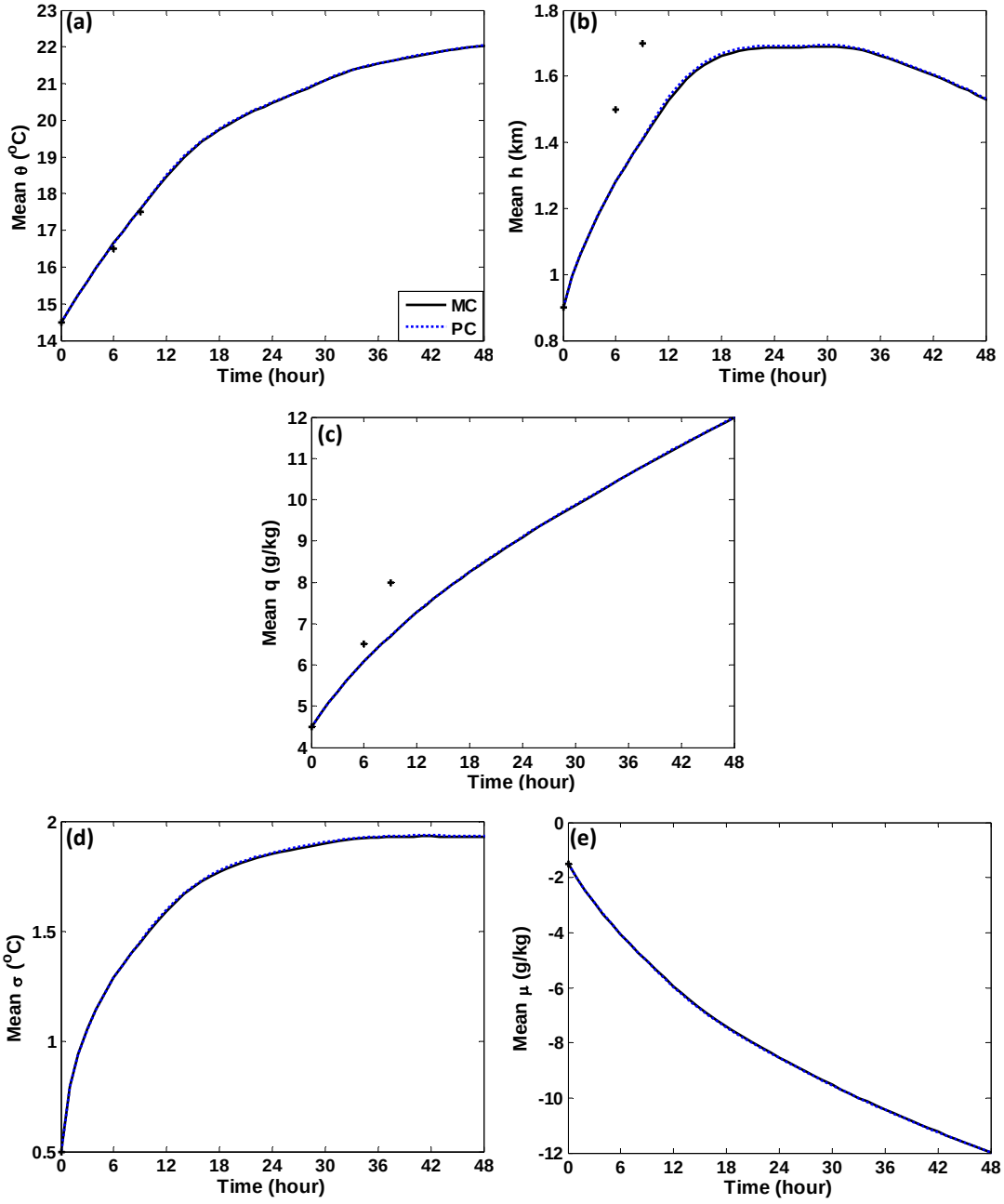


Figure 5.2 Evolution of the mean values, (mixed-layer model) IC only, PC vs. MC

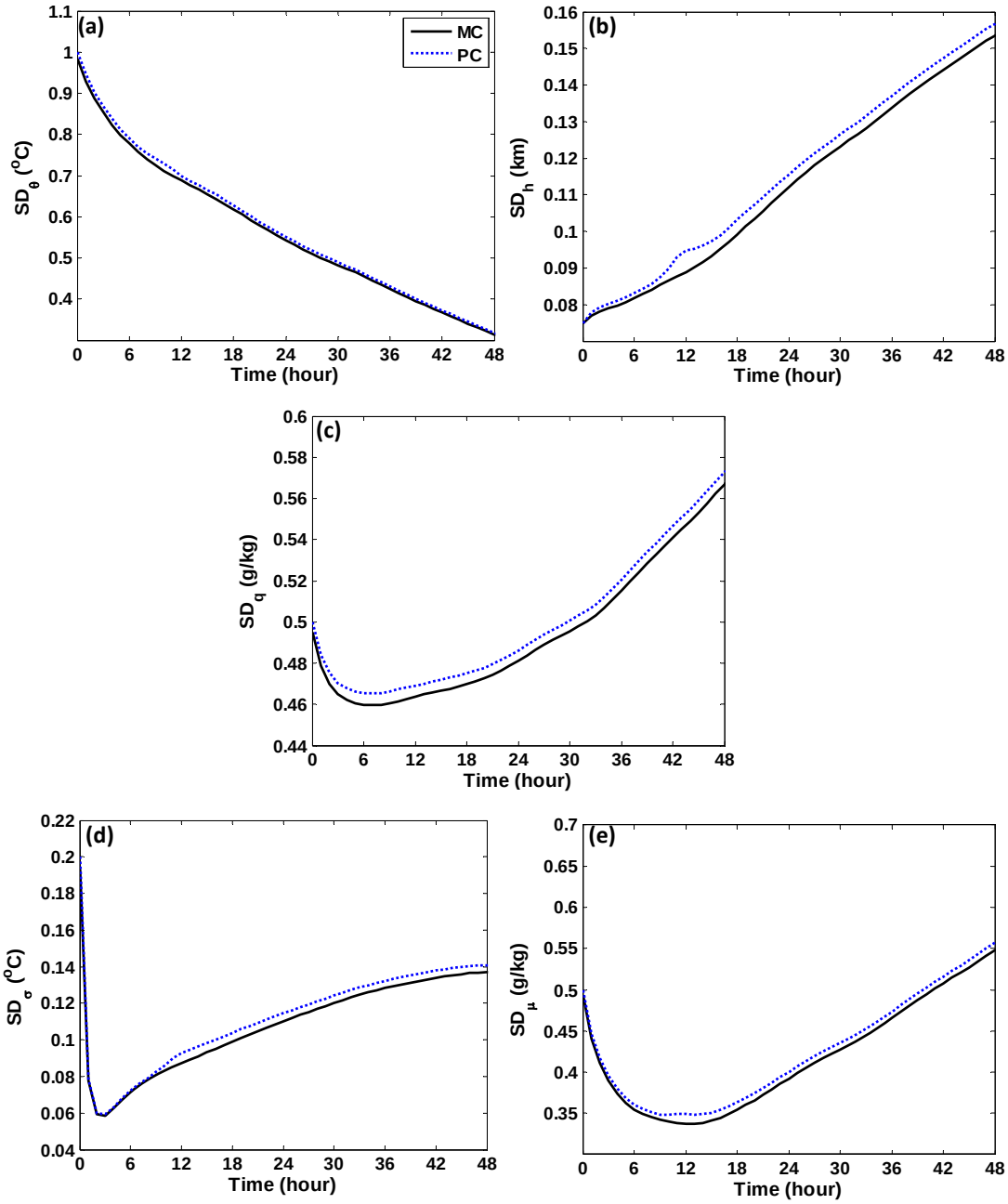


Figure 5.3 Evolution of the standard deviations, (mixed-layer model) IC only, PC vs. MC

Table 5.8 Covariance matrix, (mixed-layer model) IC only, PC vs. MC

t(h)	MC					PC				
1	0.8637	-0.0148	-0.0263	0.0280	-0.0006	0.8948	-0.0154	-0.0277	0.0267	-0.0002
	-0.0148	0.0059	0.0007	-0.0027	0.0013	-0.0154	0.0061	0.0005	-0.0027	0.0016
	-0.0263	0.0007	0.0060	0.0010	0.0005	-0.0277	0.0005	0.0061	0.0015	0.0005
	0.0280	-0.0027	0.0010	0.2293	0.0319	0.0267	-0.0027	0.0015	0.2343	0.0341
	-0.0006	0.0013	0.0005	0.0319	0.1942	-0.0002	0.0016	0.0005	0.0341	0.2001
3	0.7270	-0.0300	-0.0366	0.0647	-0.0134	0.7533	-0.0311	-0.0384	0.0646	-0.0133
	-0.0300	0.0062	0.0035	-0.0065	0.0038	-0.0311	0.0065	0.0036	-0.0065	0.0040
	-0.0366	0.0035	0.0034	-0.0037	0.0022	-0.0384	0.0036	0.0034	-0.0037	0.0023
	0.0647	-0.0065	-0.0037	0.2161	0.0615	0.0646	-0.0065	-0.0037	0.2210	0.0647
	-0.0134	0.0038	0.0022	0.0615	0.1520	-0.0133	0.0040	0.0023	0.0647	0.1563
6	0.6042	-0.0422	-0.0441	0.1026	-0.0333	0.6266	-0.0436	-0.0457	0.1039	-0.0342
	-0.0422	0.0067	0.0056	-0.0104	0.0061	-0.0436	0.0069	0.0058	-0.0105	0.0063
	-0.0441	0.0056	0.0051	-0.0093	0.0051	-0.0457	0.0058	0.0052	-0.0094	0.0053
	0.1026	-0.0104	-0.0093	0.2114	0.0794	0.1039	-0.0105	-0.0094	0.2163	0.0830
	-0.0333	0.0061	0.0051	0.0794	0.1259	-0.0342	0.0063	0.0053	0.0830	0.1295
12	0.4738	-0.0528	-0.0531	0.1413	-0.0572	0.5051	-0.0530	-0.0535	0.1473	-0.0655
	-0.0528	0.0079	0.0077	-0.0168	0.0091	-0.0530	0.0083	0.0081	-0.0168	0.0089
	-0.0531	0.0077	0.0076	-0.0167	0.0089	-0.0535	0.0081	0.0080	-0.0167	0.0088
	0.1413	-0.0168	-0.0167	0.2151	0.0881	0.1473	-0.0168	-0.0167	0.2211	0.0907
	-0.0572	0.0091	0.0089	0.0881	0.1135	-0.0655	0.0089	0.0088	0.0907	0.1188
24	0.2955	-0.0586	-0.0572	0.1762	-0.0809	0.3116	-0.0601	-0.0585	0.1826	-0.0864
	-0.0586	0.0126	0.0123	-0.0350	0.0179	-0.0601	0.0130	0.0127	-0.0356	0.0184
	-0.0572	0.0123	0.0121	-0.0342	0.0177	-0.0585	0.0127	0.0125	-0.0346	0.0181
	0.1762	-0.0350	-0.0342	0.2316	0.0690	0.1826	-0.0356	-0.0346	0.2370	0.0717
	-0.0809	0.0179	0.0177	0.0690	0.1539	-0.0864	0.0184	0.0181	0.0717	0.1590

36	0.1807	-0.0559	-0.0532	0.1790	-0.0862	0.1909	-0.0578	-0.0549	0.1857	-0.0907
	-0.0559	0.0179	0.0171	-0.0555	0.0284	-0.0578	0.0184	0.0176	-0.0566	0.0292
	-0.0532	0.0171	0.0164	-0.0527	0.0273	-0.0549	0.0176	0.0169	-0.0537	0.0281
	0.1790	-0.0555	-0.0527	0.2654	0.0301	0.1857	-0.0566	-0.0537	0.2703	0.0326
	-0.0862	0.0284	0.0273	0.0301	0.2168	-0.0907	0.0292	0.0281	0.0326	0.2229
48	0.0982	-0.0475	-0.0421	0.1609	-0.0805	0.1039	-0.0494	-0.0438	0.1671	-0.0844
	-0.0475	0.0236	0.0210	-0.0788	0.0411	-0.0494	0.0243	0.0217	-0.0806	0.0422
	-0.0421	0.0210	0.0188	-0.0695	0.0368	-0.0438	0.0217	0.0195	-0.0714	0.0380
	0.1609	-0.0788	-0.0695	0.3216	-0.0299	0.1671	-0.0806	-0.0714	0.3267	-0.0279
	-0.0805	0.0411	0.0368	-0.0299	0.3003	-0.0844	0.0422	0.0380	-0.0279	0.3078

After studying the distribution of the forecast by examining the histograms at different times built from samples obtained through PC expansion and MC approach, the performance of PC approximation of a single simulation can also be investigated. Figure 5.13 shows its behavior at base state, i.e., the IC, BC and parameters all use their mean values in Tables 5.2-5.4, therefore $\boldsymbol{\xi}$ is a zero vector in PC expresion. The dashed line represents the simulation from PC approach, the solid line is the solution by solving the governing equations (5.1)-(5.5) using Runge-Kutta method, and the plus signs are the observations. As is obvious in the figure, the PC approach provides a good approximation of the mixed-layer model at base state. Its simulation on other samples can also be investigated.

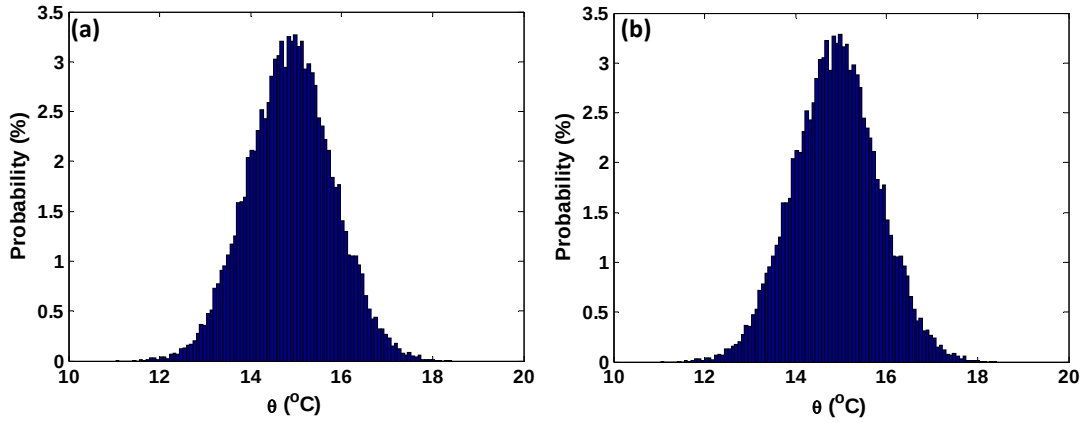


Figure 5.4 Histogram of θ at $t = 1$h, (mixed-layer model) IC only, (a) MC (b) PC

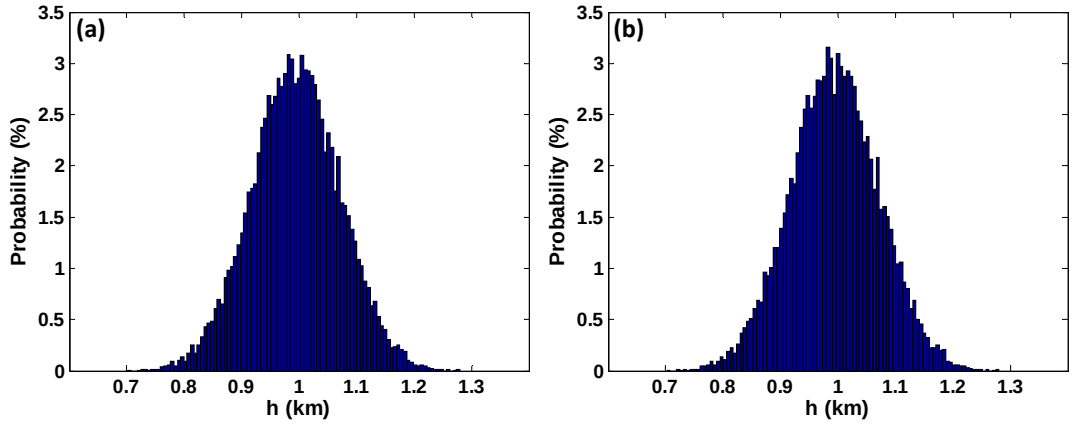


Figure 5.5 Histogram of h at $t = 1$h, (mixed-layer model) IC only, (a) MC (b) PC

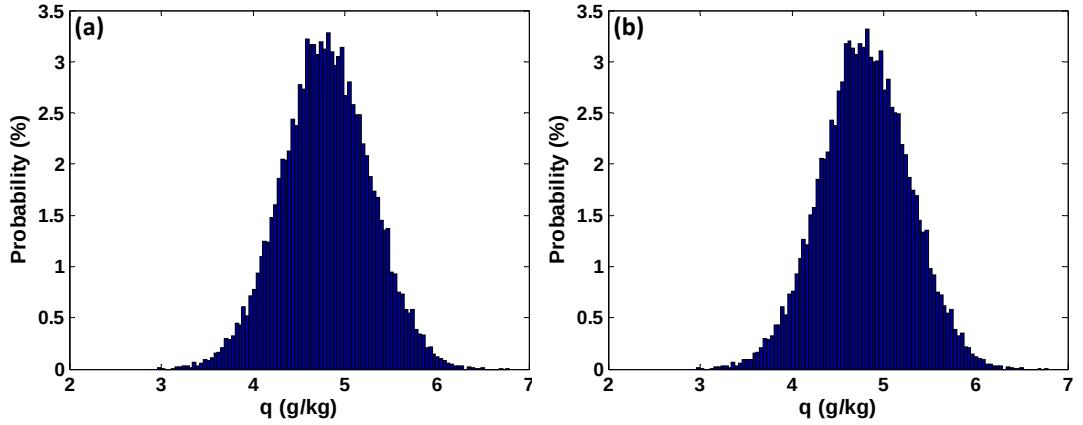


Figure 5.6 Histogram of q at $t = 1$h, (mixed-layer model) IC only, (a) MC (b) PC

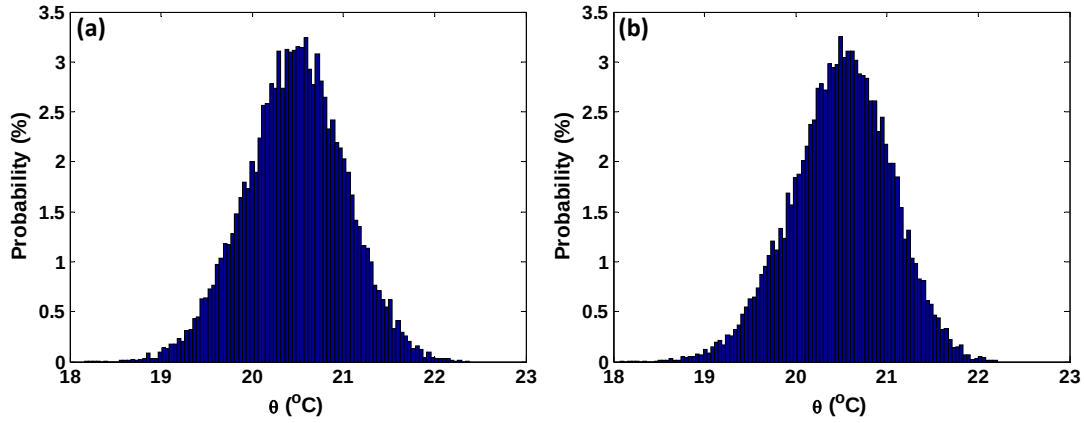


Figure 5.7 Histogram of θ at t = 24h, (mixed-layer model) IC only, (a) MC (b) PC

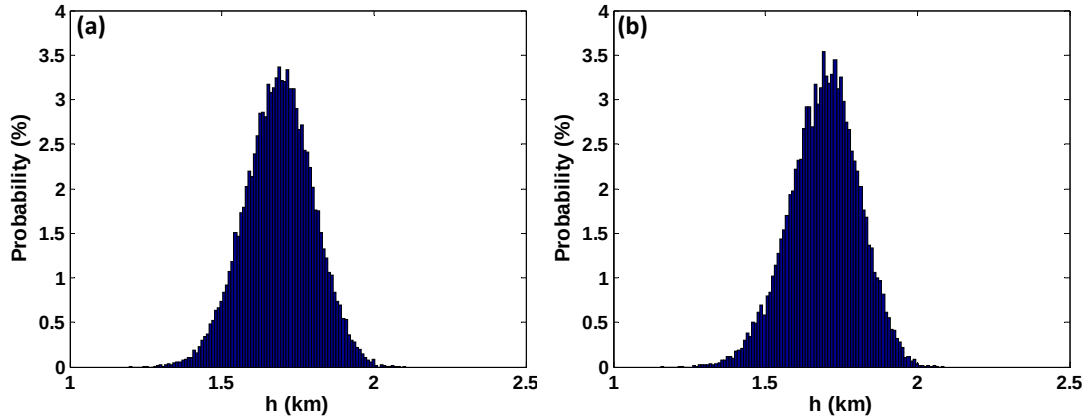


Figure 5.8 Histogram of h at t = 24h, (mixed-layer model) IC only, (a) MC (b) PC

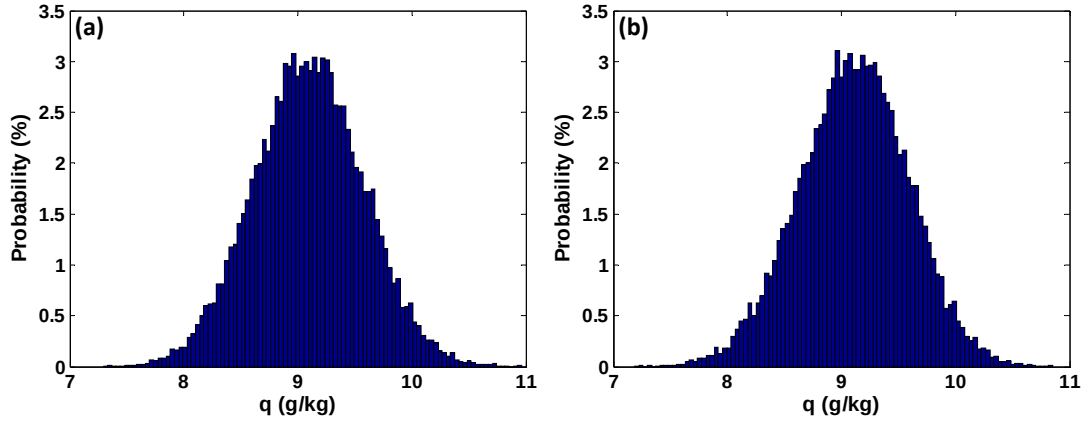


Figure 5.9 Histogram of q at t = 24h, (mixed-layer model) IC only, (a) MC (b) PC

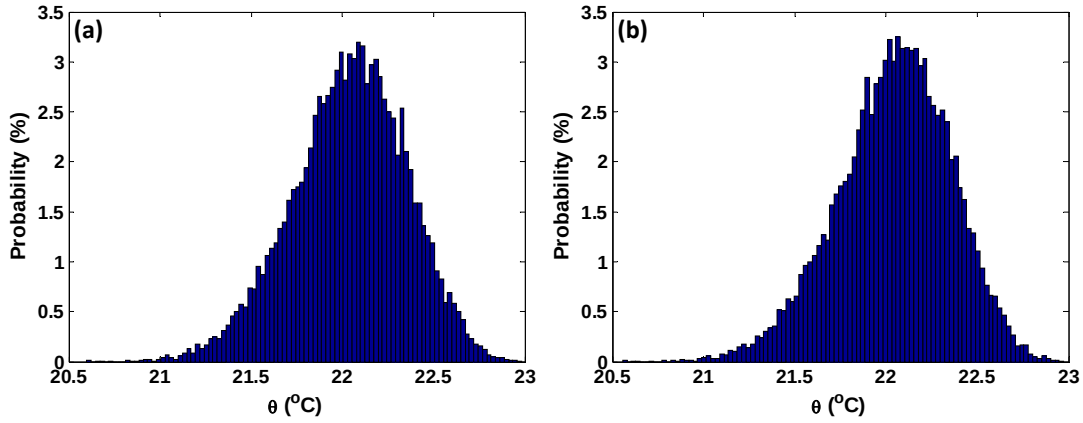


Figure 5.10 Histogram of θ at $t = 48$h, (mixed-layer model) IC only, (a) MC (b) PC

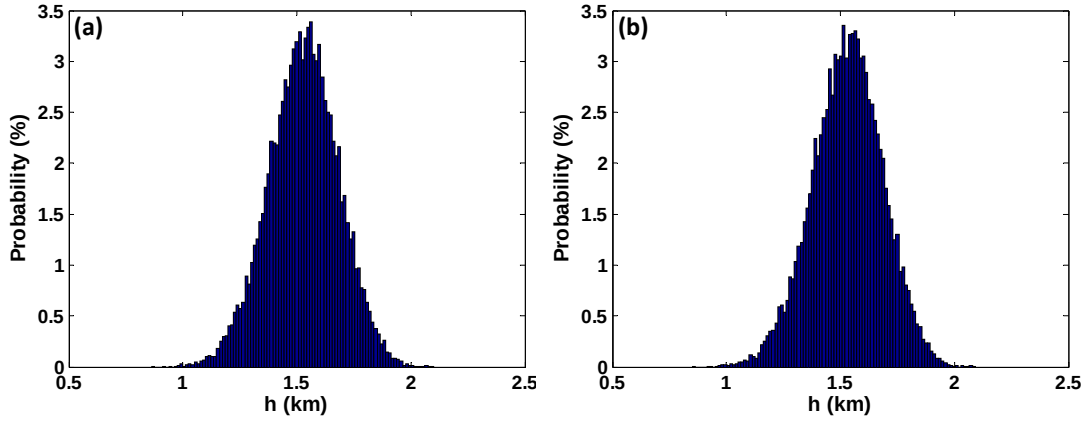


Figure 5.11 Histogram of h at $t = 48$h, (mixed-layer model) IC only, (a) MC (b) PC

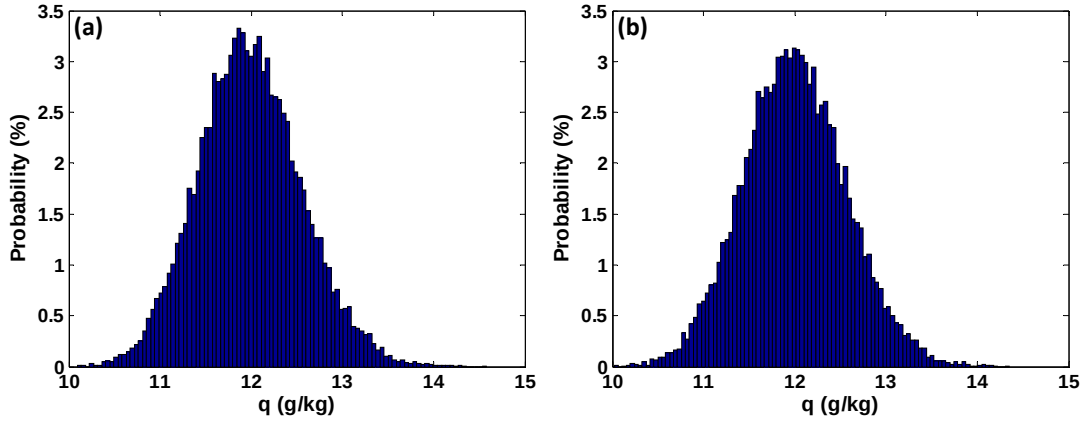


Figure 5.12 Histogram of q at $t = 48$h, (mixed-layer model) IC only, (a) MC (b) PC

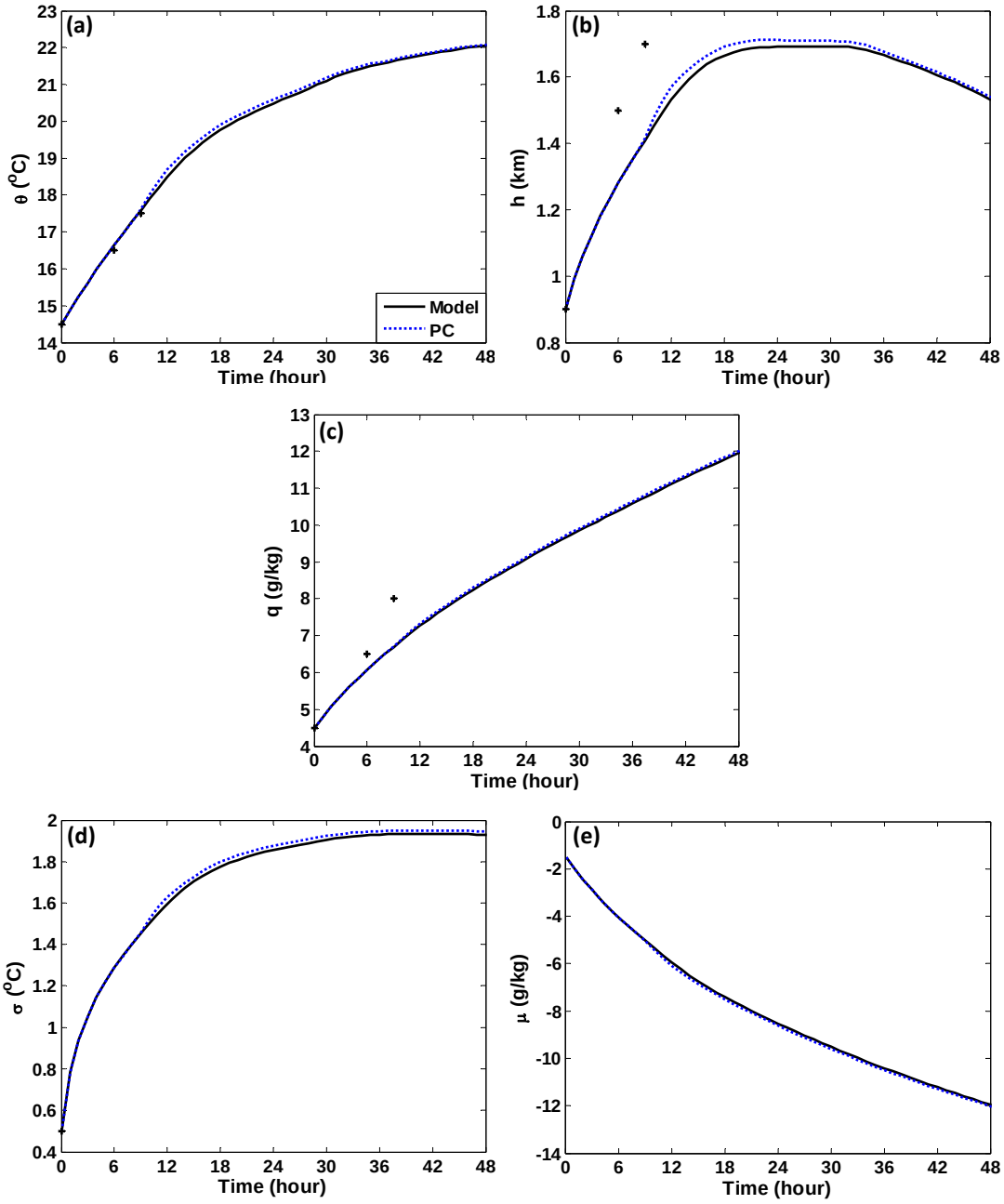


Figure 5.13 The simulation of base state by PC, (mixed-layer model) IC only

From above, polynomial chaos expansion has a good performance in IC only case, next it will be studied in parameter only case.

5.3 Parameter Only

Different from initial conditions, the parameters can only use values within some range. For example, w and γ_q can only use negative values. Therefore, they are assumed to follow uniform distributions and independent with each other. According to Table 2.1, the optimal polynomial chaos should be Legendre-chaos for uniform distribution. Refer to Appendix G for details of Legendre polynomial chaos.

Again, let $\mathbf{x} = (\theta, h, \sigma, q, \mu)^{\mathrm{T}}$, the aim is to seek an approximation $\mathbf{x}_N^M(t, \boldsymbol{\eta})$ of $\mathbf{x}(t)$ in the form of PC expression

$$\mathbf{x}_N^M(t, \boldsymbol{\eta}) = \sum_{|\mathbf{i}|=0}^{M} \mathbf{v}_{\mathbf{i}}(t)\Phi_{\mathbf{i}}(\boldsymbol{\eta}), \tag{5.9}$$

where $\{\mathbf{v}_{\mathbf{i}}(t) = [v_{\mathbf{i},1}(t), \dots, v_{\mathbf{i},5}(t)]^{\mathrm{T}}\} \in R^5$ are the expansion coefficients. In order to differentiate the randomness of the parameters from that of the IC, $\boldsymbol{\eta}$ is used instead of $\boldsymbol{\xi}$. N is the number of random variables used in the expansion and M is the highest order of the polynomials. $\Phi_{\mathbf{i}}(\boldsymbol{\eta})$ will be the Legendre polynomials in this case.

There are six independent parameters, so six dimensional random variable $\boldsymbol{\eta} = (\eta_1, \dots, \eta_6)^{\mathbf{T}}$, i.e., $N = 6$, is used in the experiment. The elements of the random variable $\boldsymbol{\eta}$ follows uniform distribution in range [-1, 1] and they are independent with each other. The PC expansion is also truncated at order 2, i.e., $M = 2$. There are altogether $\frac{(6+2)!}{6!\times 2!} = 28$ polynomial terms in the expansion. Suppose the m-th order normalized scalar Legendre polynomial at random variable η_i is denoted as $P_m(\eta_i)$, then $P_m(\eta_i), m = 0, 1, 2$ are $1, \sqrt{3}\eta_i,$ and $\frac{\sqrt{5}(3\eta_i^2-1)}{2}$ separately. Correspondingly, the 28 Legendre polynomial terms are listed in Table 5.9.

Table 5.9 Six-variable normalized Legendre polynomials (order no greater than 2)

Degree m	**Multi index** $(p_1p_2p_3p_4p_5p_6)$	$P_m(\boldsymbol{\eta})$	$P_{p_1}(\eta_1)P_{p_2}(\eta_2)P_{p_3}(\eta_3)P_{p_4}(\eta_4)P_{p_5}(\eta_5)P_{p_6}(\eta_6)$
0	(0 0 0 0 0 0)	1	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
1	(1 0 0 0 0 0)	$\sqrt{3}\eta_1$	$P_1(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 1 0 0 0 0)	$\sqrt{3}\eta_2$	$P_0(\eta_1)P_1(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 0 1 0 0 0)	$\sqrt{3}\eta_3$	$P_0(\eta_1)P_0(\eta_2)P_1(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 0 0 1 0 0)	$\sqrt{3}\eta_4$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_1(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 0 0 0 1 0)	$\sqrt{3}\eta_5$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_1(\eta_5)P_0(\eta_6)$
	(0 0 0 0 0 1)	$\sqrt{3}\eta_6$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_1(\eta_6)$
2	(2 0 0 0 0 0)	$\frac{\sqrt{5}(3\eta_1^2-1)}{2}$	$P_2(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(1 1 0 0 0 0)	$3\eta_1\eta_2$	$P_1(\eta_1)P_1(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(1 0 1 0 0 0)	$3\eta_1\eta_3$	$P_1(\eta_1)P_0(\eta_2)P_1(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(1 0 0 1 0 0)	$3\eta_1\eta_4$	$P_1(\eta_1)P_0(\eta_2)P_0(\eta_3)P_1(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(1 0 0 0 1 0)	$3\eta_1\eta_5$	$P_1(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_1(\eta_5)P_0(\eta_6)$
	(1 0 0 0 0 1)	$3\eta_1\eta_6$	$P_1(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_1(\eta_6)$
	(0 2 0 0 0 0)	$\frac{\sqrt{5}(3\eta_2^2-1)}{8}$	$P_0(\eta_1)P_2(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 1 1 0 0 0)	$3\eta_2\eta_3$	$P_0(\eta_1)P_1(\eta_2)P_1(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 1 0 1 0 0)	$3\eta_2\eta_4$	$P_0(\eta_1)P_1(\eta_2)P_0(\eta_3)P_1(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 1 0 0 1 0)	$3\eta_2\eta_5$	$P_0(\eta_1)P_1(\eta_2)P_0(\eta_3)P_0(\eta_4)P_1(\eta_5)P_0(\eta_6)$
	(0 1 0 0 0 1)	$3\eta_2\eta_6$	$P_0(\eta_1)P_1(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_1(\eta_6)$
	(0 0 2 0 0 0)	$\frac{\sqrt{5}(3\eta_3^2-1)}{2}$	$P_0(\eta_1)P_0(\eta_2)P_2(\eta_3)P_0(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 0 1 1 0 0)	$3\eta_3\eta_4$	$P_0(\eta_1)P_0(\eta_2)P_1(\eta_3)P_1(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 0 1 0 1 0)	$3\eta_3\eta_5$	$P_0(\eta_1)P_0(\eta_2)P_1(\eta_3)P_0(\eta_4)P_1(\eta_5)P_0(\eta_6)$
	(0 0 1 0 0 1)	$3\eta_3\eta_6$	$P_0(\eta_1)P_0(\eta_2)P_1(\eta_3)P_0(\eta_4)P_0(\eta_5)P_1(\eta_6)$
	(0 0 0 2 0 0)	$\frac{\sqrt{5}(3\eta_4^2-1)}{2}$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_2(\eta_4)P_0(\eta_5)P_0(\eta_6)$
	(0 0 0 1 1 0)	$3\eta_4\eta_5$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_1(\eta_4)P_1(\eta_5)P_0(\eta_6)$
	(0 0 0 1 0 1)	$3\eta_4\eta_6$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_1(\eta_4)P_0(\eta_5)P_1(\eta_6)$
	(0 0 0 0 2 0)	$\frac{\sqrt{5}(3\eta_5^2-1)}{2}$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_2(\eta_5)P_0(\eta_6)$
	(0 0 0 0 1 1)	$3\eta_5\eta_6$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_1(\eta_5)P_1(\eta_6)$
	(0 0 0 0 0 2)	$\frac{\sqrt{5}(3\eta_6^2-1)}{2}$	$P_0(\eta_1)P_0(\eta_2)P_0(\eta_3)P_0(\eta_4)P_0(\eta_5)P_2(\eta_6)$

The expansion coefficients at any time t are obtained through SC method. In parameter only case, the Gauss-Legendre quadrature rule and sparse grid scheme

discussed in Chapter 2 are used. The weight function is the weighting function for uniform distribution over [-1, 1] where the random variable is located in. Table 5.10 shows the 13 collocation points and corresponding weights when the exact level is 2, i.e., $K = 2$. Table 5.11 shows the collocation points and corresponding weights when the exact level is 3, i.e., $K = 3$.

Table 5.10 Sparse collocation points with weights (dimension 6, exact level 2), Gauss-Legendre quadrature rule

No.	η_1	η_2	η_3	η_4	η_5	η_6	Weight
1	-0.5774	0	0	0	0	0	0.5
2	0	-0.5774	0	0	0	0	0.5
3	0	0	-0.5774	0	0	0	0.5
4	0	0	0	-0.5774	0	0	0.5
5	0	0	0	0	-0.5774	0	0.5
6	0	0	0	0	0	-0.5774	0.5
7	0	0	0	0	0	0	-5.0
8	0	0	0	0	0	0.5774	0.5
9	0	0	0	0	0.5774	0	0.5
10	0	0	0	0.5774	0	0	0.5
11	0	0	0.5774	0	0	0	0.5
12	0	0.5774	0	0	0	0	0.5
13	0.5774	0	0	0	0	0	0.5

Table 5.11 Sparse collocation points with weights (dimension 6, exact level 3), Gauss-Legendre quadrature rule

No.	η_1	η_2	η_3	η_4	η_5	η_6	Weight
1	-0.7746	0	0	0	0	0	0.2778
2	-0.5774	-0.5774	0	0	0	0	0.25
3	-0.5774	0	-0.5774	0	0	0	0.25
4	-0.5774	0	0	-0.5774	0	0	0.25
5	-0.5774	0	0	0	-0.5774	0	0.25
6	-0.5774	0	0	0	0	-0.5774	0.25
7	-0.5774	0	0	0	0	0	-2.50
8	-0.5774	0	0	0	0	0.5774	0.25

9	-0.5774	0	0	0	0.5774	0	0.25
10	-0.5774	0	0	0.5774	0	0	0.25
11	-0.5774	0	0.5774	0	0	0	0.25
12	-0.5774	0.5774	0	0	0	0	0.25
13	0	-0.7746	0	0	0	0	0.2778
14	0	-0.5774	-0.5774	0	0	0	0.25
15	0	-0.5774	0	-0.5774	0	0	0.25
16	0	-0.5774	0	0	-0.5774	0	0.25
17	0	-0.5774	0	0	0	-0.5774	0.25
18	0	-0.5774	0	0	0	0	-2.50
19	0	-0.5774	0	0	0	0.5774	0.25
20	0	-0.5774	0	0	0.5774	0	0.25
21	0	-0.5774	0	0.5774	0	0	0.25
22	0	-0.5774	0.5774	0	0	0	0.25
23	0	0	-0.7746	0	0	0	0.2778
24	0	0	-0.5774	-0.5774	0	0	0.25
25	0	0	-0.5774	0	-0.5774	0	0.25
26	0	0	-0.5774	0	0	-0.5774	0.25
27	0	0	-0.5774	0	0	0	-2.50
28	0	0	-0.5774	0	0	0.5774	0.25
29	0	0	-0.5774	0	0.5774	0	0.25
30	0	0	-0.5774	0.5774	0	0	0.25
31	0	0	0	-0.7746	0	0	0.2778
32	0	0	0	-0.5774	-0.5774	0	0.25
33	0	0	0	-0.5774	0	-0.5774	0.25
34	0	0	0	-0.5774	0	0	-2.50
35	0	0	0	-0.5774	0	0.5774	0.25
36	0	0	0	-0.5774	0.5774	0	0.25
37	0	0	0	0	-0.7746	0	0.2778
38	0	0	0	0	-0.5774	-0.5774	0.25
39	0	0	0	0	-0.5774	0	-2.50
40	0	0	0	0	-0.5774	0.5774	0.25
41	0	0	0	0	0	-0.7746	0.2778
42	0	0	0	0	0	-0.5774	-2.50
43	0	0	0	0	0	0	12.6667

44	0	0	0	0	0	0.5774	-2.50
45	0	0	0	0	0	0.7746	0.2778
46	0	0	0	0	0.5774	-0.5774	0.25
47	0	0	0	0	0.5774	0	-2.50
48	0	0	0	0	0.5774	0.5774	0.25
49	0	0	0	0	0.7746	0	0.2778
50	0	0	0	0.5774	-0.5774	0	0.25
51	0	0	0	0.5774	0	-0.5774	0.25
52	0	0	0	0.5774	0	0	-2.50
53	0	0	0	0.5774	0	0.5774	0.25
54	0	0	0	0.5774	0.5774	0	0.25
55	0	0	0	0.7746	0	0	0.2778
56	0	0	0.5774	-0.5774	0	0	0.25
57	0	0	0.5774	0	-0.5774	0	0.25
58	0	0	0.5774	0	0	-0.5774	0.25
59	0	0	0.5774	0	0	0	-2.50
60	0	0	0.5774	0	0	0.5774	0.25
61	0	0	0.5774	0	0.5774	0	0.25
62	0	0	0.5774	0.5774	0	0	0.25
63	0	0	0.7746	0	0	0	0.2778
64	0	0.5774	-0.5774	0	0	0	0.25
65	0	0.5774	0	-0.5774	0	0	0.25
66	0	0.5774	0	0	-0.5774	0	0.25
67	0	0.5774	0	0	0	-0.5774	0.25
68	0	0.5774	0	0	0	0	-2.50
69	0	0.5774	0	0	0	0.5774	0.25
70	0	0.5774	0	0	0.5774	0	0.25
71	0	0.5774	0	0.5774	0	0	0.25
72	0	0.5774	0.5774	0	0	0	0.25
73	0	0.7746	0	0	0	0	0.2778
74	0.5774	-0.5774	0	0	0	0	0.25
75	0.5774	0	-0.5774	0	0	0	0.25
76	0.5774	0	0	-0.5774	0	0	0.25
77	0.5774	0	0	0	-0.5774	0	0.25
78	0.5774	0	0	0	0	-0.5774	0.25

79	0.5774	0	0	0	0	0	-2.50
80	0.5774	0	0	0	0	0.5774	0.25
81	0.5774	0	0	0	0.5774	0	0.25
82	0.5774	0	0	0.5774	0	0	0.25
83	0.5774	0	0.5774	0	0	0	0.25
84	0.5774	0.5774	0	0	0	0	0.25
85	0.7746	0	0	0	0	0	0.2778

Again, the PC approach is compared with MC approach with 20,000 samples for the parameters. Figures 5.14 and 5.15 show the evolution of the mean values and standard deviations of five variables from time $t = 0$ to $t = 48h$. The solid line represents the simulation with MC, the dashed line is the simulation by PC approach with exact level $K = 2$ and the plus signs are observations from Table 5.1. As can be seen from the figures, the PC approach can have very good estimate on mean values, which are very close to those obtained through MC ensemble approach. Though the differences on the standard deviations are larger than those for the mean values, they are within acceptable range. Using the approximation in (5.9) together with the obtained expansion coefficients, one can easily construct the ensemble members of the state vector by drawing samples of the standard random variable $\boldsymbol{\eta}$. One can further examine the distribution of each variable at time t by plotting the histograms of the ensemble samples at that time. Figures 5.16 to 5.24 are the histogram plots of θ, h and q at times $t = 1h, 24h$ and $48h$ with MC on the left and PC on the right. From the figures, overall, the histogram of PC has good resemblance with that of MC especially at the earlier times. However, with time increasing, there exhibits some differences. For example, at $t = 24h$, the range and peak value of the temperature differ from each other. The situation becomes worse when $t = 48h$.

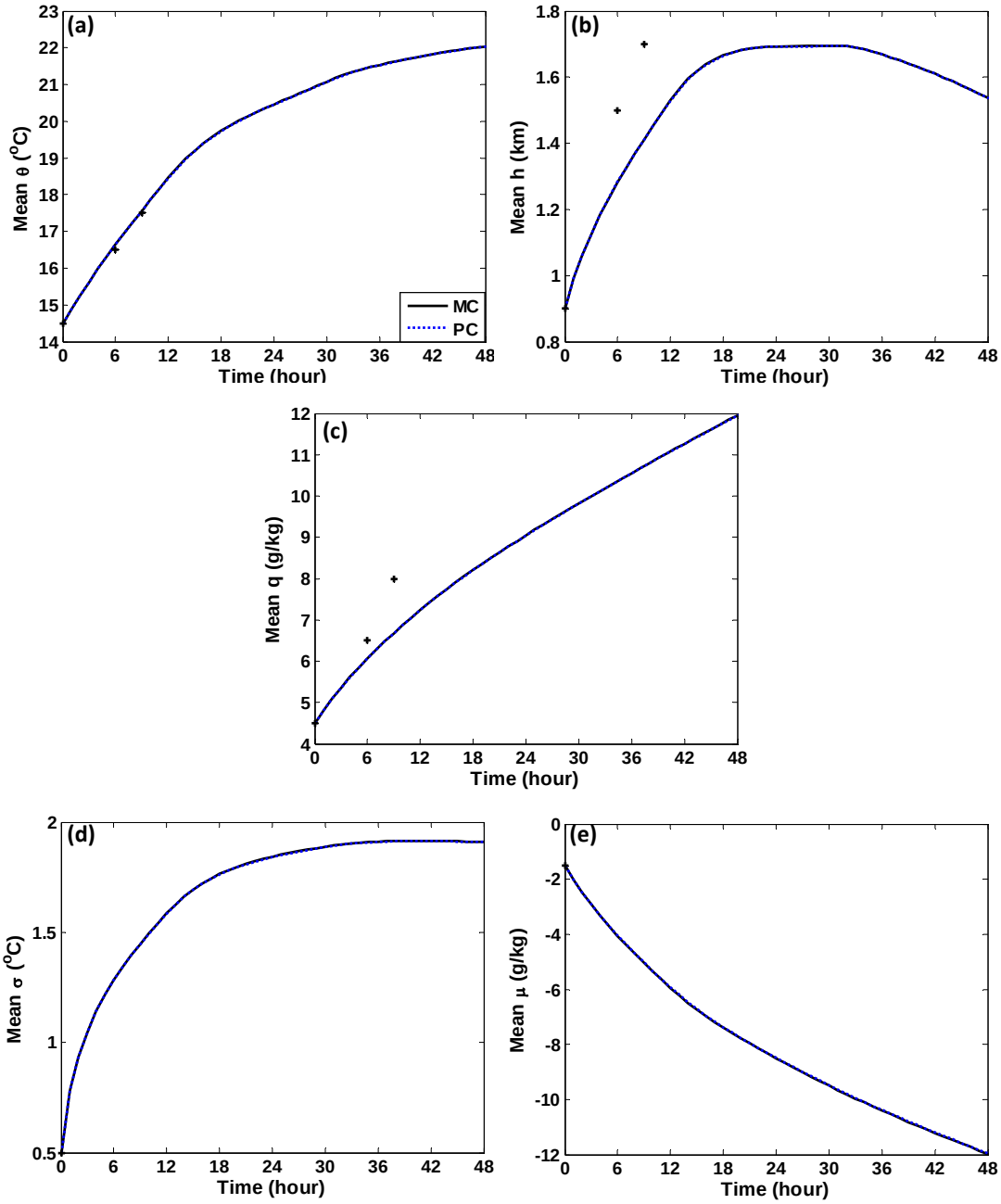


Figure 5.14 Evolution of mean values, (mixed-layer model) Parameter only, PC (exact level 2) vs. MC

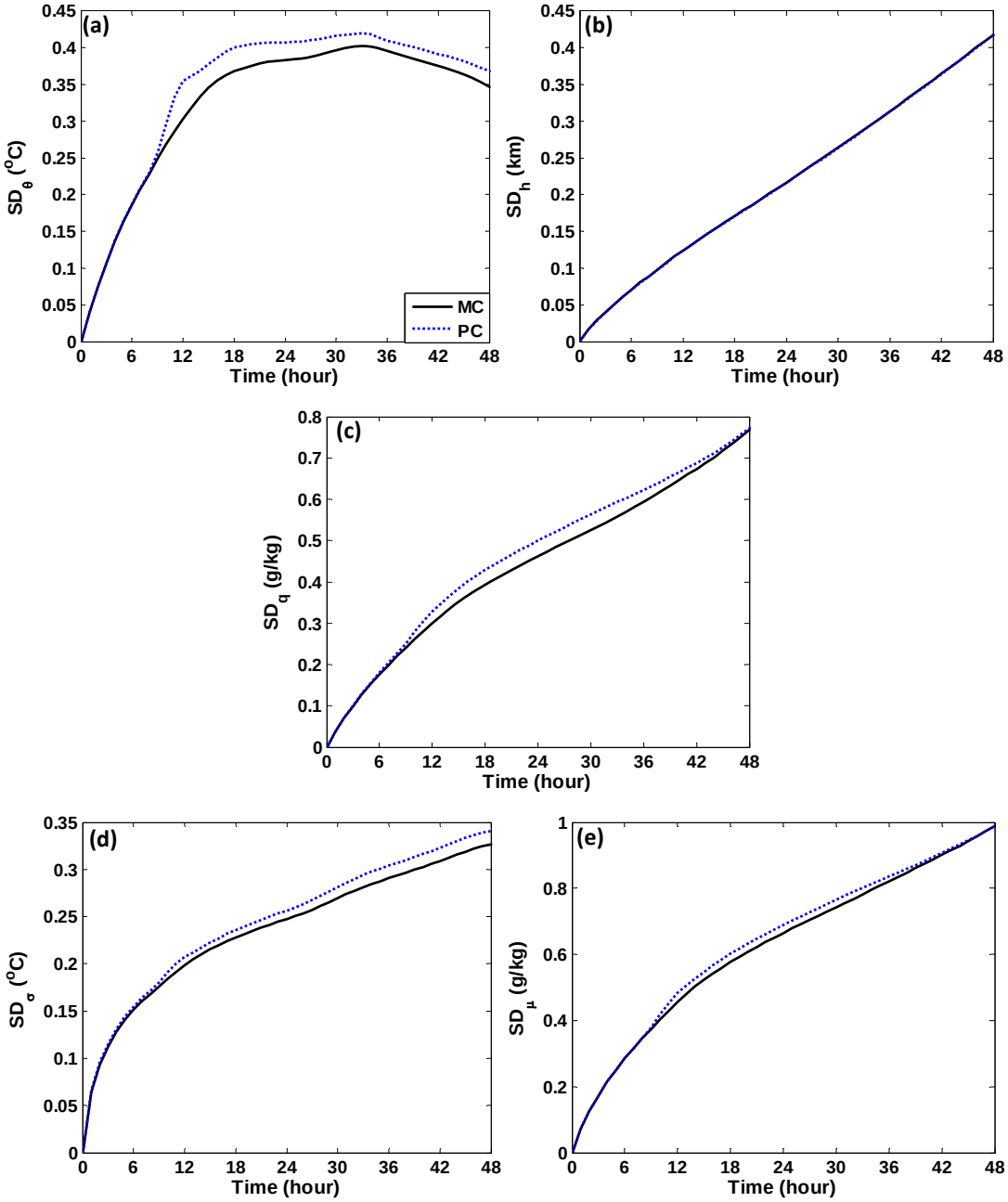


Figure 5.15 Evolution of standard deviations, (mixed-layer model) Parameter only, PC (exact level 2) vs. MC

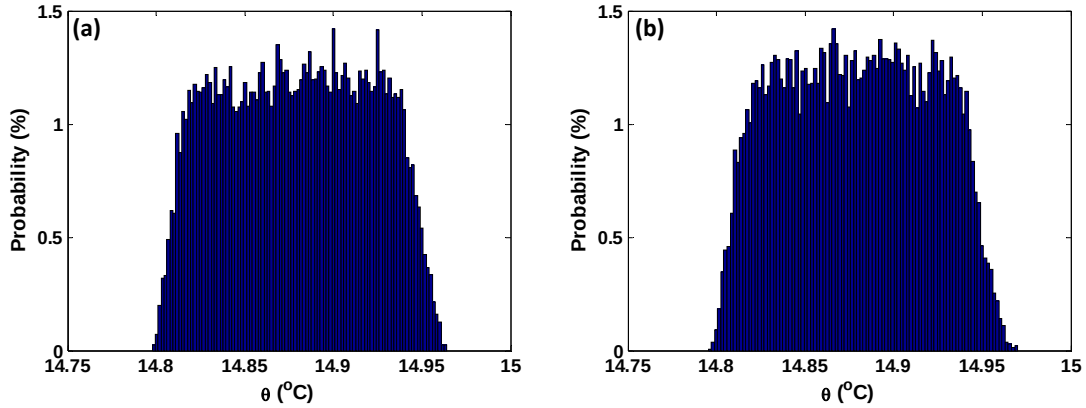


Figure 5.16 Histogram of θ at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

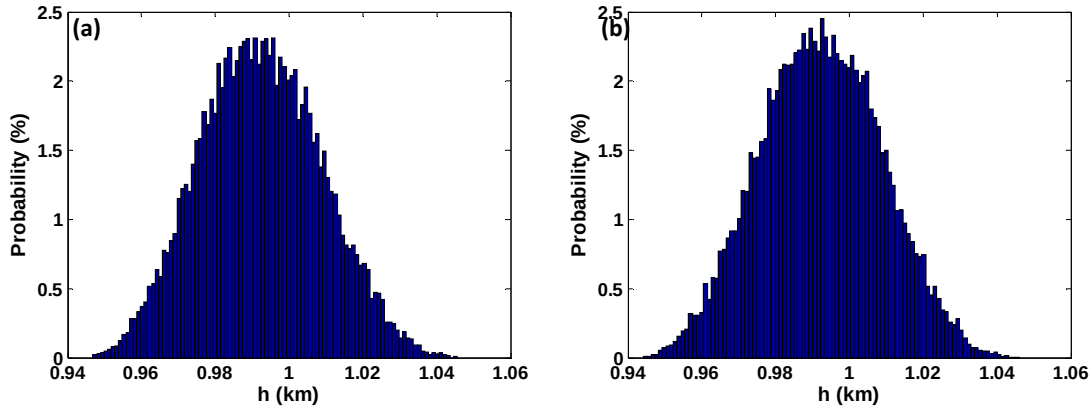


Figure 5.17 Histogram of h at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

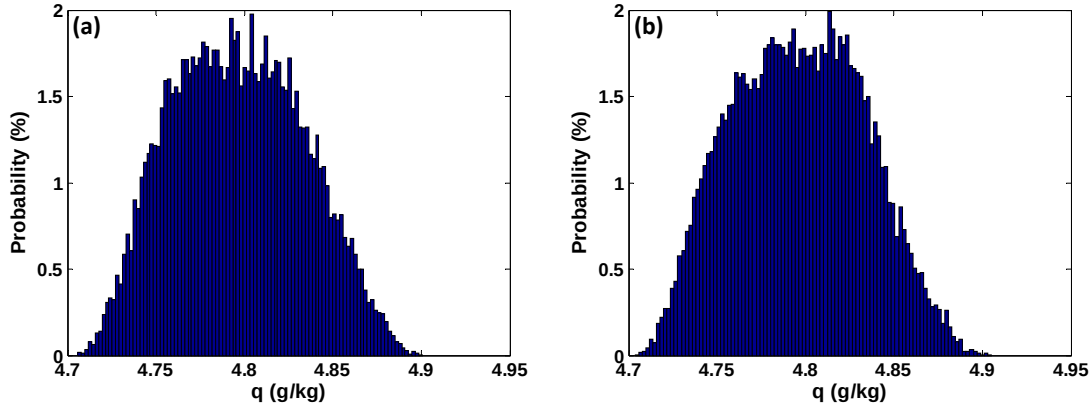


Figure 5.18 Histogram of q at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

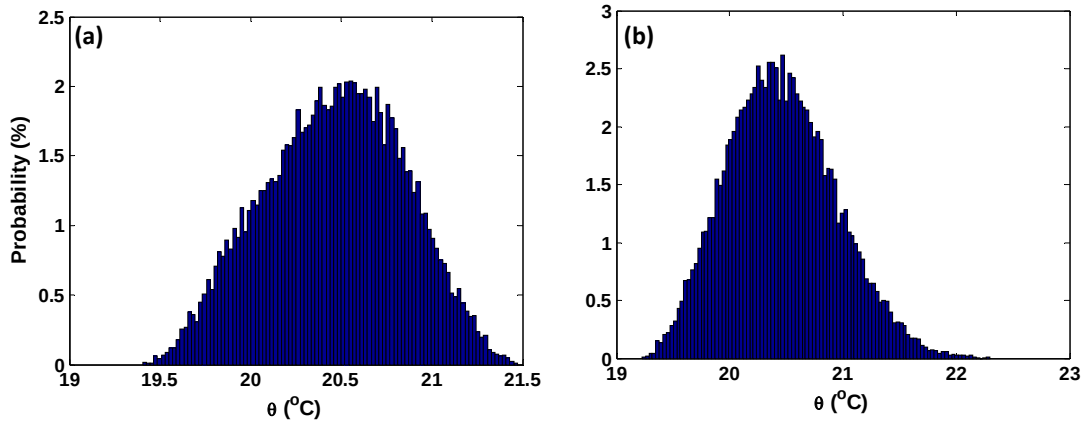


Figure 5.19 Histogram of θ at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

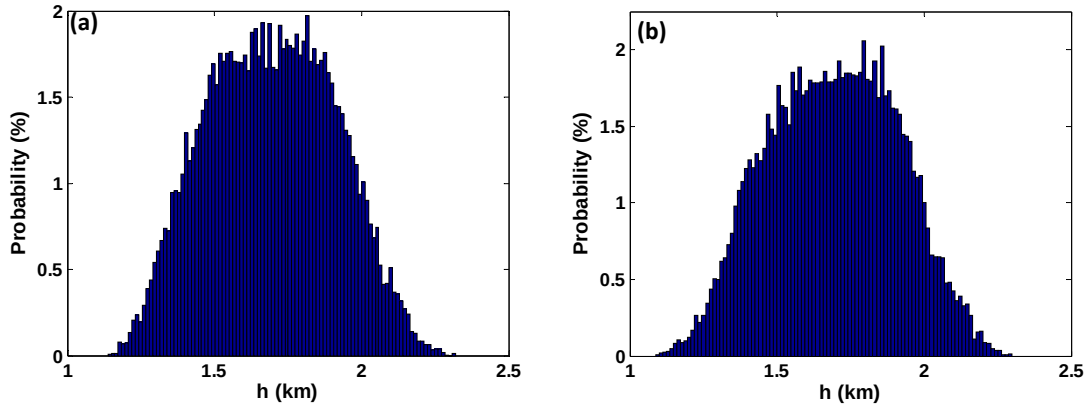


Figure 5.20 Histogram of h at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

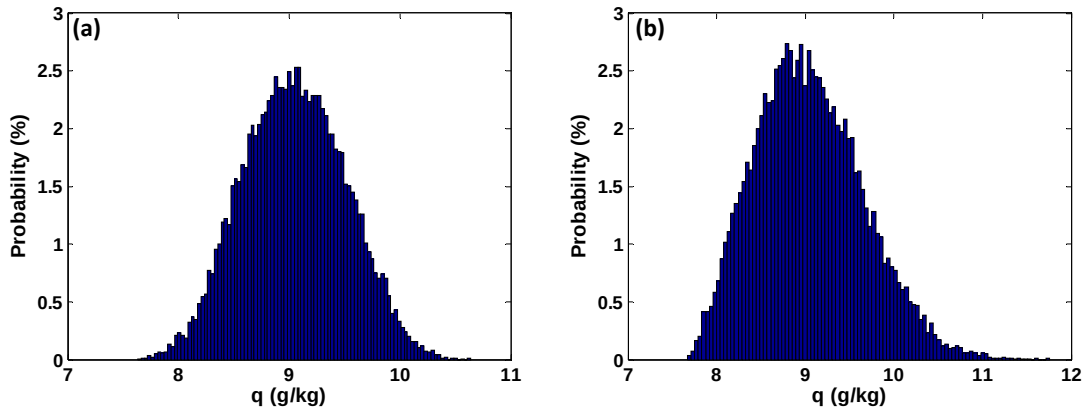


Figure 5.21 Histogram of q at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

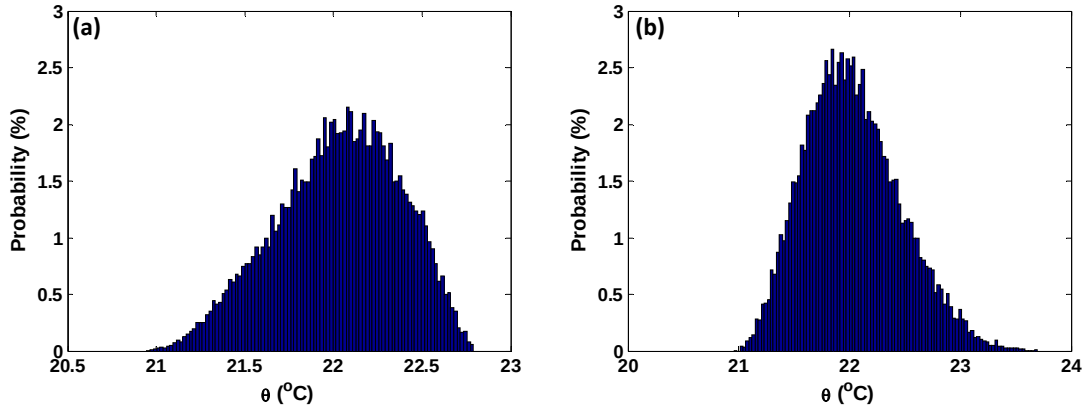


Figure 5.22 Histogram of θ at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

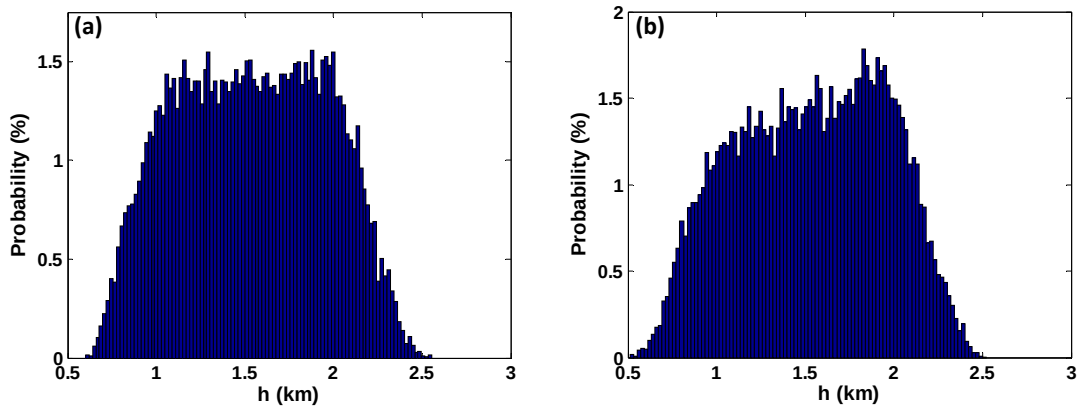


Figure 5.23 Histogram of h at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

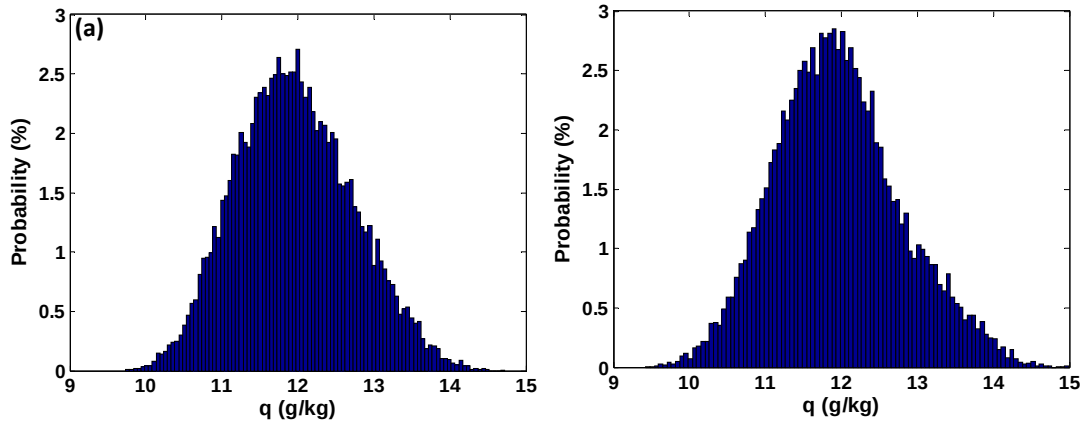


Figure 5.24 Histogram of q at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 2)

The performance of PC approach may be improved by using more collocation points. In order to see the improvement, 85 points with weights given in Table 5.11 are used. Here, another second-order polynomial approximation of the state vector $\mathbf{x}(t)$ is achieved. The evolution of mean values, standard deviations and the histograms constructed by sampling are evaluated with the comparison with those obtained using MC ensemble approach. Figures 5.25 and 5.26 are the evolution of the mean values and standard deviations. Since PC approach with $K = 2$ has already estimated the mean values very close to those from MC approach, one can hardly tell the improvement when setting $K = 3$, which is obvious excellent estimate. The improvement in estimating the standard deviation is remarkable. Except for the standard deviation of the temperature θ, all other estimates are very close to those from MC approach, one can hardly tell the difference form the figures. In addition, let's take a look at the histograms in Figures 5.27-5.35 (MC on the left and PC on the right). As can be seen from the figures, the differences between PC and MC approach become smaller. Overall, they have quite similar distributions especially for h and q, though there still exists difference. In conclusion, by introducing more collocation points, one can improve the performance of PC approach. Like everything in life, this is obtained by spending more computation time. In real applications, balance is needed to take into consideration in terms of computation and accuracy. For completeness, the approximation of PC expansion at base state is also examined. The result is shown on Figure 5.36, it does give a good estimation even though it is not as good as that for IC only case.

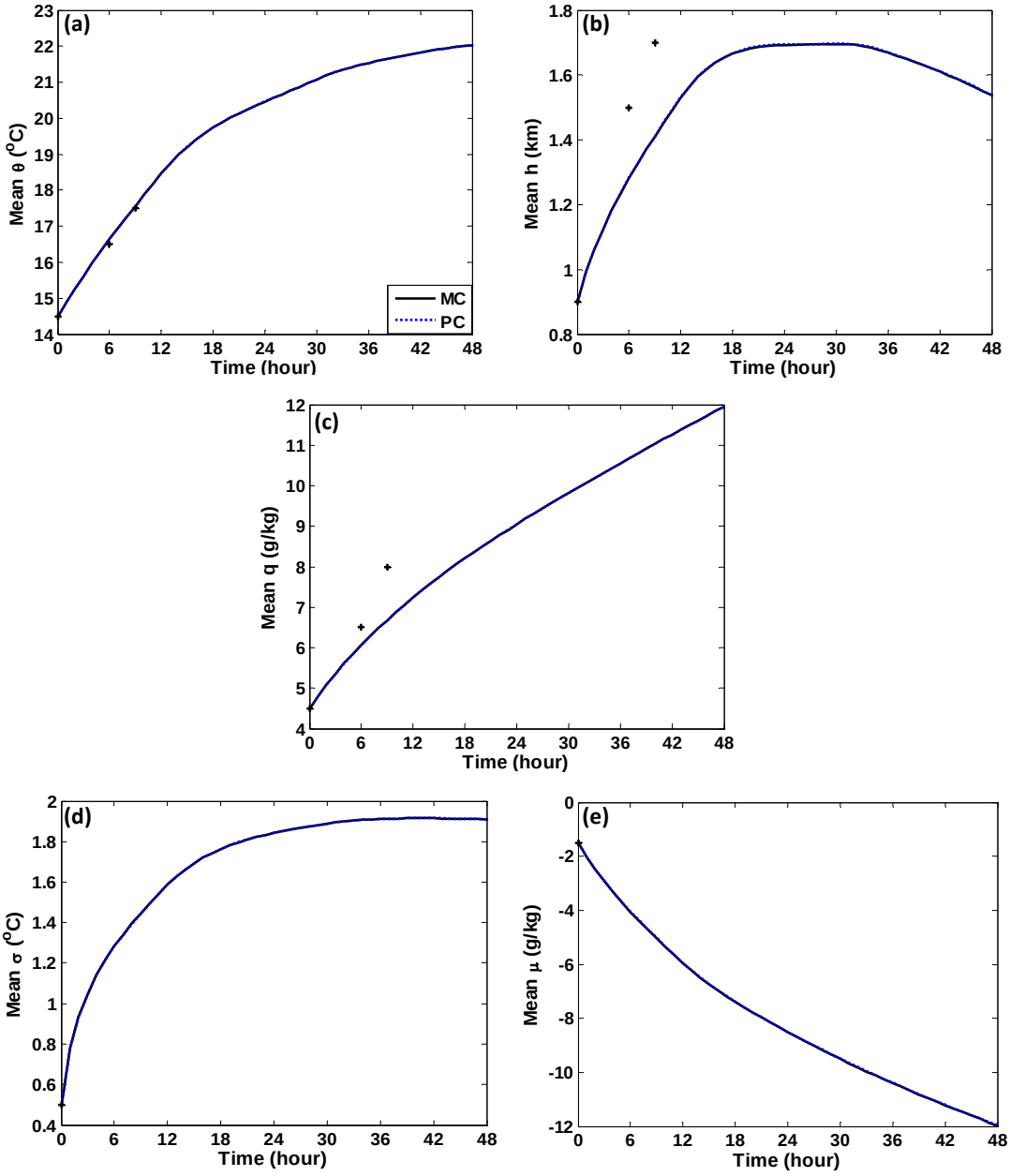


Figure 5.25 Evolution of mean values, (mixed-layer model) Parameter only, PC (exact level 3) vs. MC

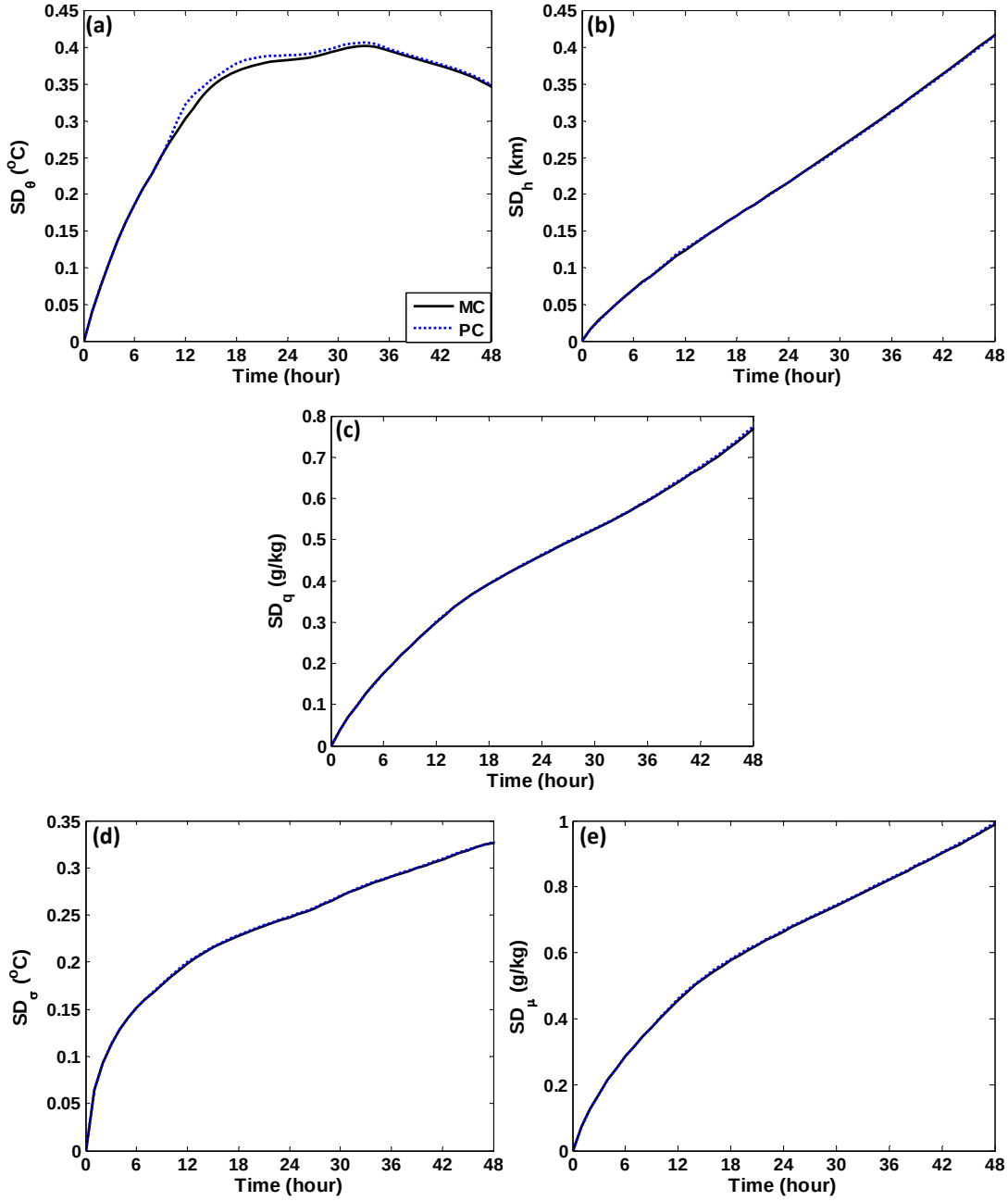


Figure 5.26 Evolution of standard deviations, (mixed-layer model) Parameter only, PC (exact level 3) vs. MC

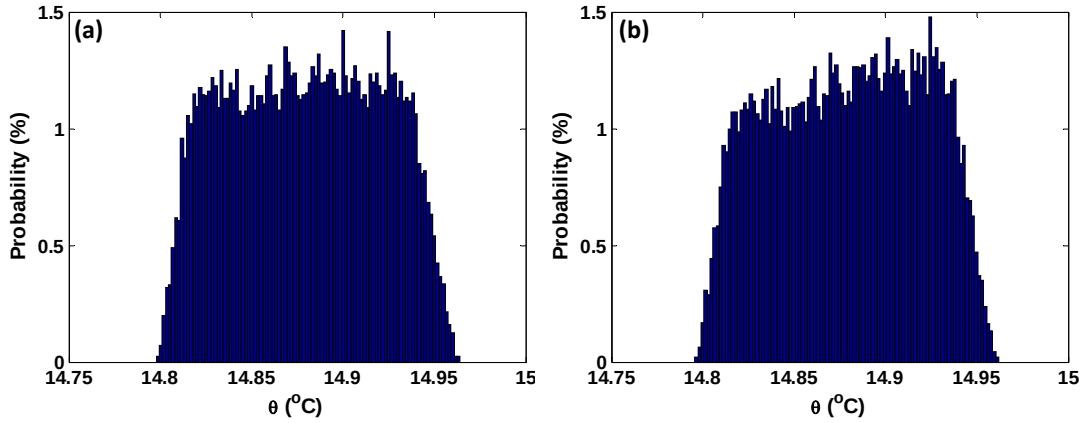


Figure 5.27 Histogram of θ at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

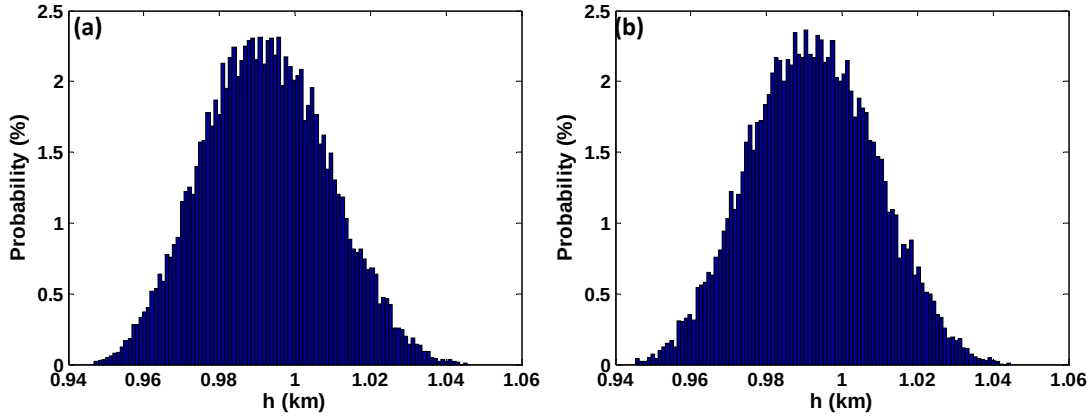


Figure 5.28 Histogram of h at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

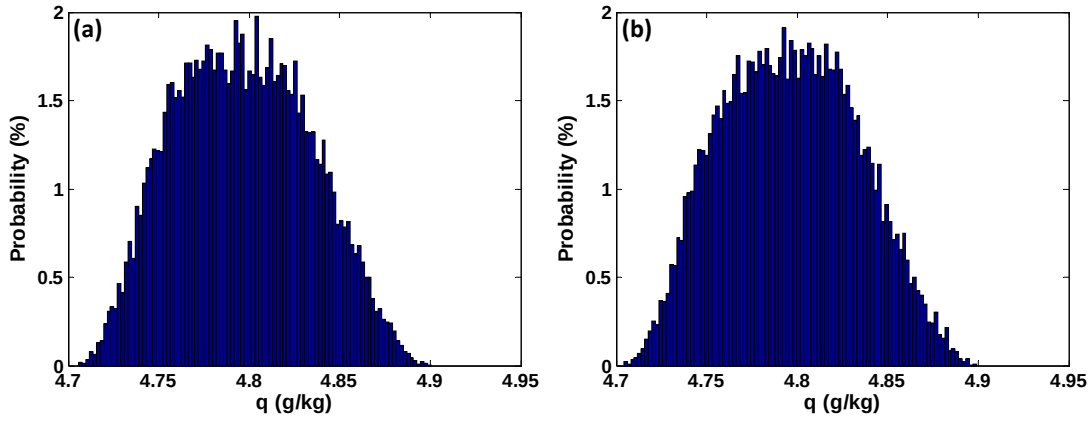


Figure 5.29 Histogram of q at $t = 1$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

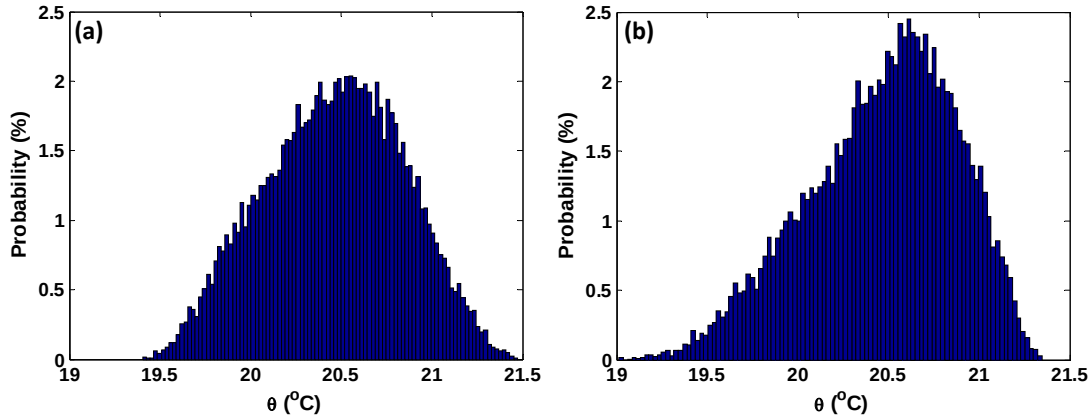


Figure 5.30 Histogram of θ at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

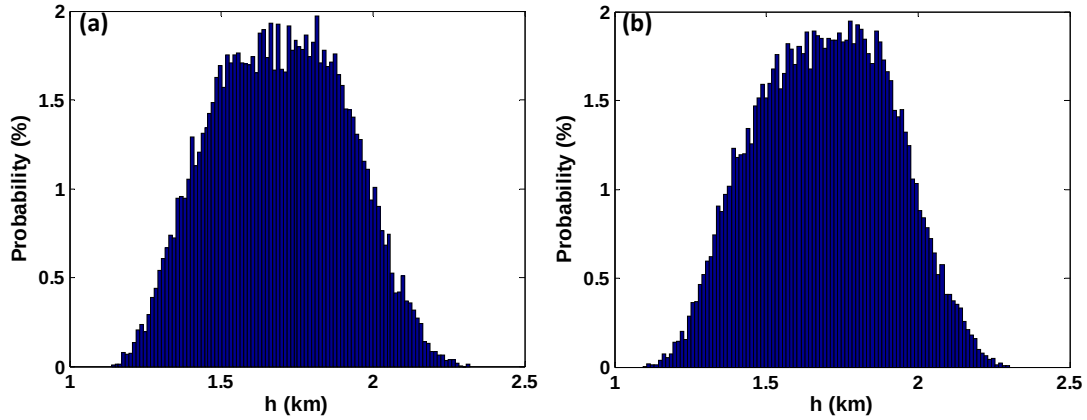


Figure 5.31 Histogram of h at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

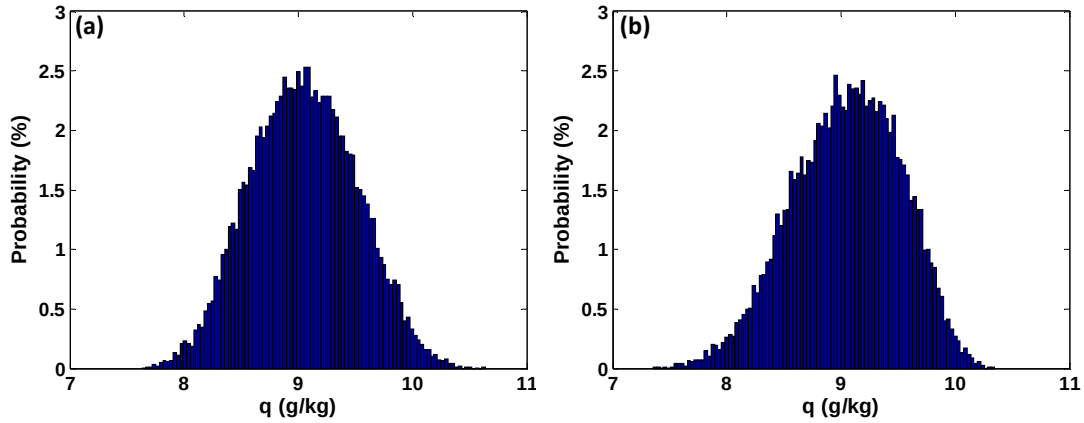


Figure 5.32 Histogram of q at $t = 24$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

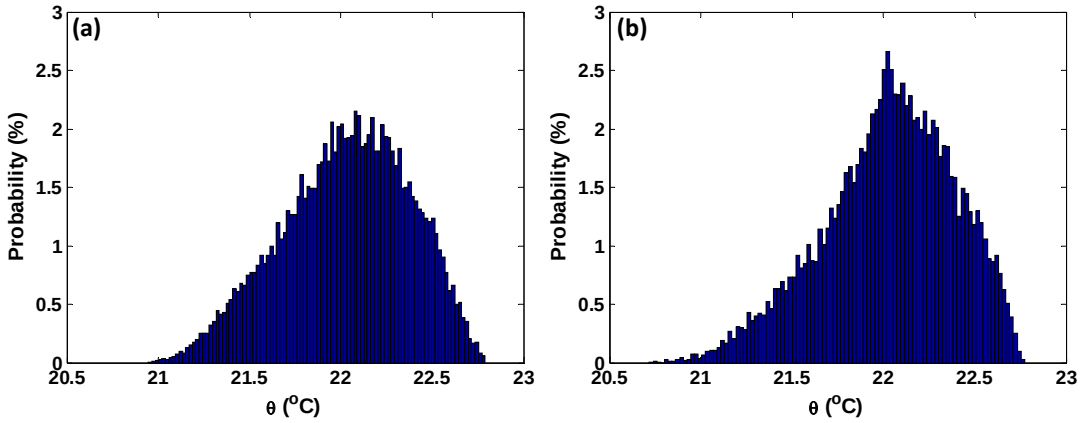


Figure 5.33 Histogram of θ at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

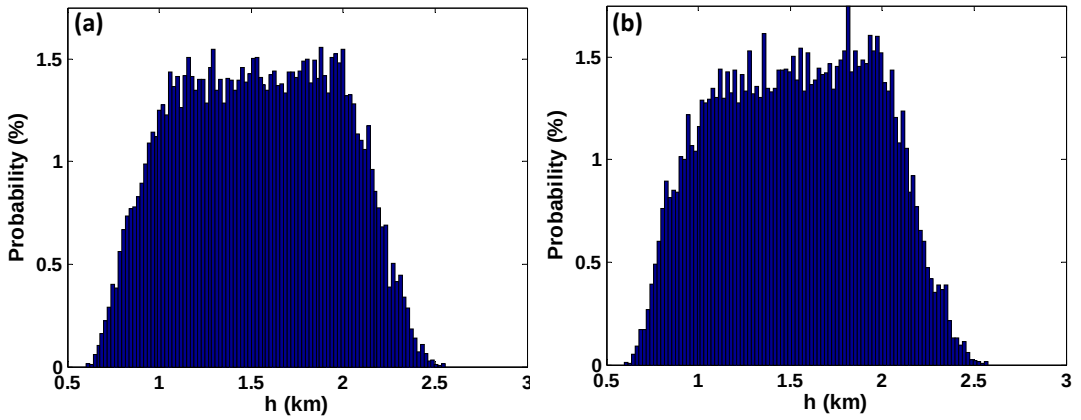


Figure 5.34 Histogram of h at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

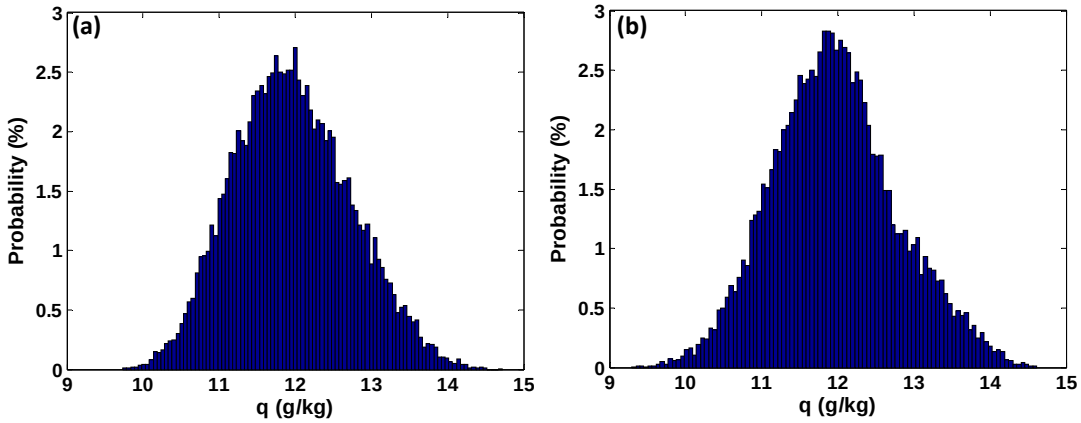


Figure 5.35 Histogram of q at $t = 48$h, (mixed-layer model) Parameter only, (a) MC (b) PC (exact level 3)

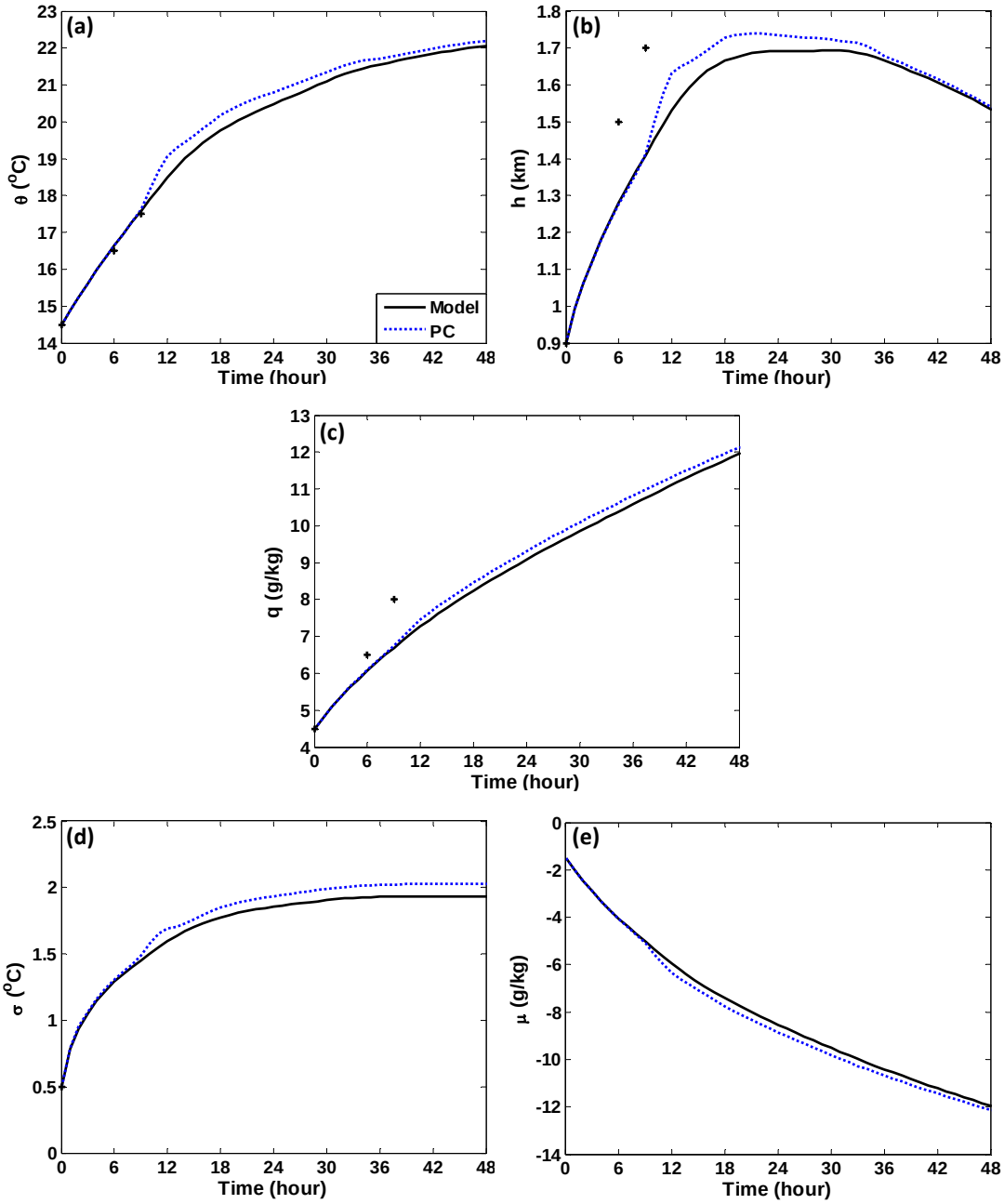


Figure 5.36 The simulation of base state by PC (exact level 3), (mixed-layer model) Parameter only

5.4 Discussions

In this chapter, a mixed-layer model was used to study the ability of uncertainty quantification by using PC expansion, specifically using stochastic Collocation (SC) approach to obtain the expansion coefficients. According to the assumption of the uncertainty in initial condition and parameters, multivariate Hermite polynomial chaos expansion was used in IC only case and multivariate Legendre polynomial chaos expansion was adopted in parameter only case. The performance of PC expansion was compared with MC in both cases. It is shown that PC can have good estimates of mean, standard deviation, and covariance. The distribution represented by the histogram was also presented. Besides, the performance of PC expansion can further be improved by increasing the number of collocation points. Different from SG, no model modification is need in SC. Instead, the simulation of the original model on selected collocation points is necessary.

Chapter 6

Application of Unscented Transformation Approach

In the last two chapters, PC expansion was studied in quantifying the uncertainty in the forecast from a dynamical system. In this chapter, using the same five-variable mixed-layer model which is used to describe the return flow event over the Gulf of Mexico, the effectiveness of unscented transformation (UT) method in uncertainty quantification will be studied. For consistency and completeness, both IC only and parameter only cases with the same distribution assumption as in Chapter 5 will be examined.

As introduced in Chapter 3, the UT method adopts a set of deterministic samples to propagate the uncertainty through a dynamic system. Specifically, the SUT scheme discussed in Chapter 3 is used in this chapter.

6.1 Initial Condition Only

In initial condition (IC) only case, the distribution of IC is described in Table 5.2. The mean values in Table 5.3-5.4 are used for parameters and BC, respectively. Therefore, the uncertainty in the forecast only comes from the randomness in the initial condition. Assume the initial vector is denoted as $\mathbf{x}_0 = (\theta_0, h_0, \sigma_0, q_0, \mu_0)^T$, according to Table 5.2, the mean and covariance matrix for IC are

$$(14.5, 0.90, 0.50, 4.50, -1.50)^T$$

and

$$\begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0 & 0.0056 & 0 & 0 & 0 \\ 0 & 0 & 0.04 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0.25 \end{pmatrix},$$

respectively. The number (dimension) of initial values is $n = 5$, therefore $2n + 1 = 11$ samples (called sigma points in UT) that capture the mean and covariance matrix of the IC are used in SUT. In the experiment, the values for the three parameters α, β, and κ are chosen as

$$\alpha = 0.5, \beta = 2, \text{and } \kappa = 0. \tag{6.1}$$

So $\lambda = -3.75$.

The 11 sigma points of the IC and corresponding weights for the mean and covariance are listed in Table 6.1.

Table 6.1 Sigma points with weights used in UT, (mixed-layer model) IC only

No.	θ_0(℃)	$h_0(km)$	σ_0(℃)	$q_0(g/kg)$	$\mu_0(g/kg)$	W^m	W^c
1	14.5	0.9	0.5	4.5	-1.5	-3.00	-0.25
2	15.618	0.9	0.5	4.5	-1.5	0.40	0.40
3	14.5	0.9839	0.5	4.5	-1.5	0.40	0.40
4	14.5	0.9	0.7236	4.5	-1.5	0.40	0.40
5	14.5	0.9	0.5	5.059	-1.5	0.40	0.40
6	14.5	0.9	0.5	4.5	-0.941	0.40	0.40
7	13.382	0.9	0.5	4.5	-1.5	0.40	0.40
8	14.5	0.8161	0.5	4.5	-1.5	0.40	0.40
9	14.5	0.9	0.2764	4.5	-1.5	0.40	0.40
10	14.5	0.9	0.5	3.941	-1.5	0.40	0.40
11	14.5	0.9	0.5	4.5	-2.059	0.40	0.40

The process of UT is to propagate each sigma point (as initial value) through the dynamics in equations (5.1)-(5.5). At each time, there are 11 propagated sigma points and the mean and covariance matrix can be computed by using formulas (3.20) and (3.21) together with the weights given in Table 6.1. Figure 6.1 shows the ensemble forecast using these 11 sigma points. The bold solid line is the ensemble mean, the dashed lines are the 11 ensemble forecasts, and the plus signs are the observations at

time 0h, 6h and 12h, respectively. As seen from the figures, the forecast of the temperature θ is very close to the observation, whereas there are some deviations for variables h and q.

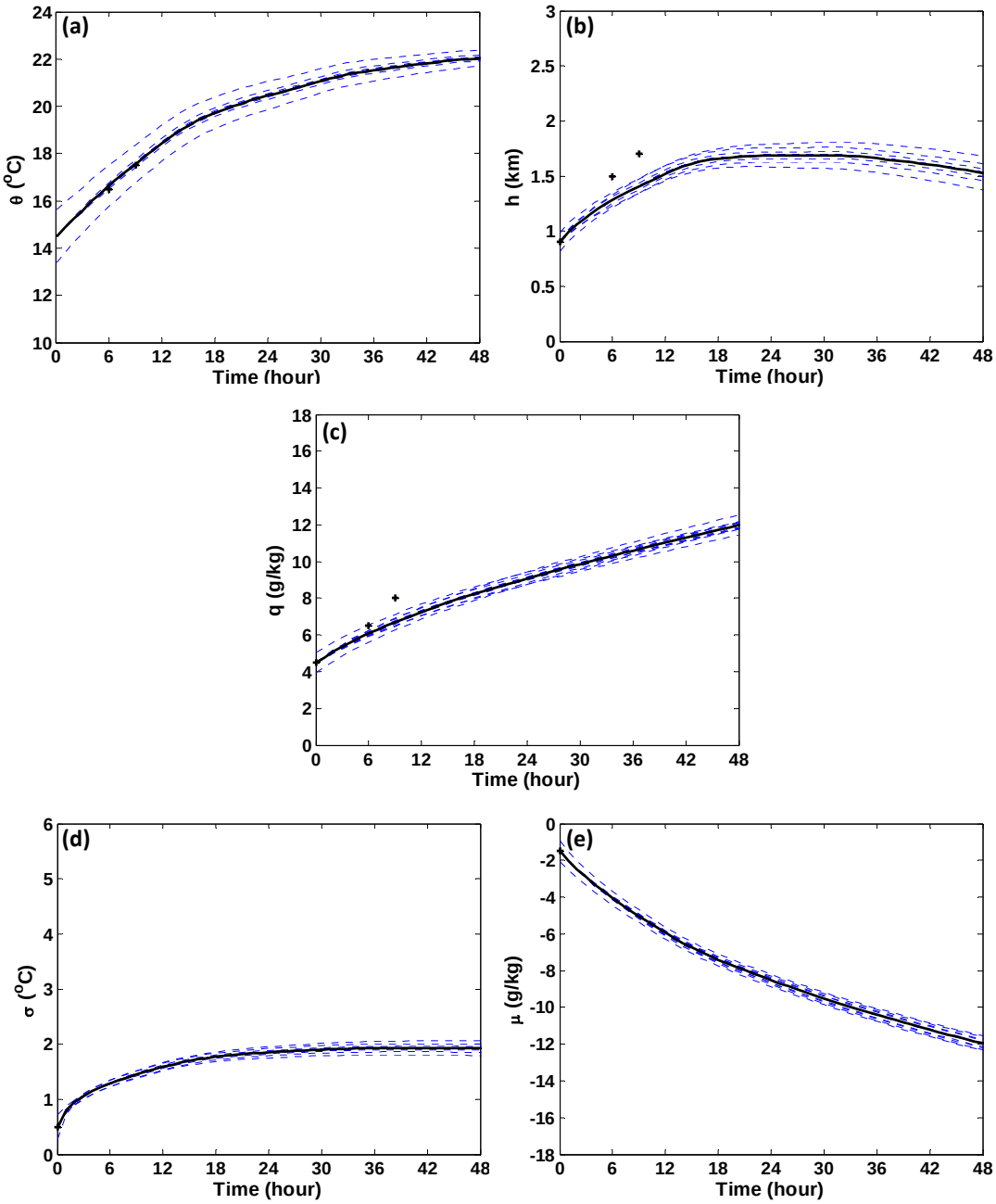


Figure 6.1 Ensemble forecast using UT, (mixed-layer model) IC only

Figures 6.2 and 6.3 further show the evolution of the ensemble means and standard deviations of the five variables compared to those obtained from MC ensemble forecast. As is clear that both methods have very close estimates for the ensemble mean, and the difference is hard to tell. The estimates for the standard deviations are not as close as those for mean values. It can be seen that UT method produces slightly higher standard deviations than MC does. However, the difference between these two methods is small, e.g., the difference of SD_θ is smaller than 0.05℃, within acceptable range. In addition, the covariance matrices from both methods at different times are given in Table 6.2. From the numbers in the table, UT can have very good estimate on forecast covariance matrix using a small set of samples, here 11 deterministic samples. In contract to MC and PC, there are only 11 samples at each time and the weights are not exactly the normal "weights" which must be in the range [0, 1]. Therefore, the histogram cannot be provided by using UT. What's more, there are no explicit expressions to compute the value of higher order moments by using the sigma points discussed in this study. Efforts can be made on developing point selection scheme to compute the statistics like higher-order moments. Some studies have been done, for example, Julier and Uhlmann (2004) presents a general sigma point selection framework to incorporate any higher order information about the moments if this information is available; Tenne and Singh (2003) consider the problem of capturing higher order moments by using augmented sigma points without the assumption of symmetry. These higher-order UTs (HUT) usually use more sigma points than the standard UT discussed in Chapter 3 and are not within the scope of this study. They will be explored in future study.

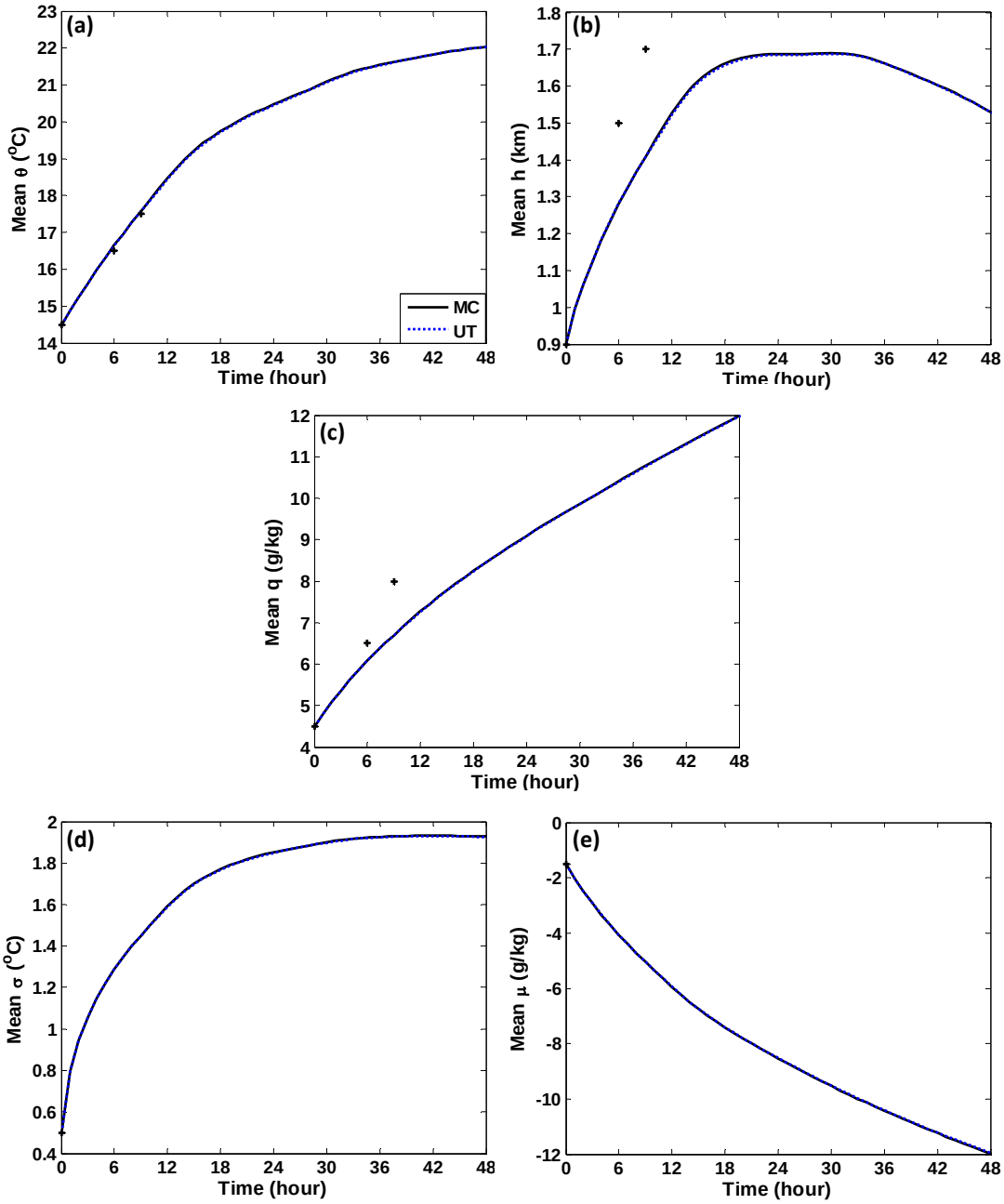


Figure 6.2 Evolution of mean values, (mixed-layer model) IC only, UT vs. MC

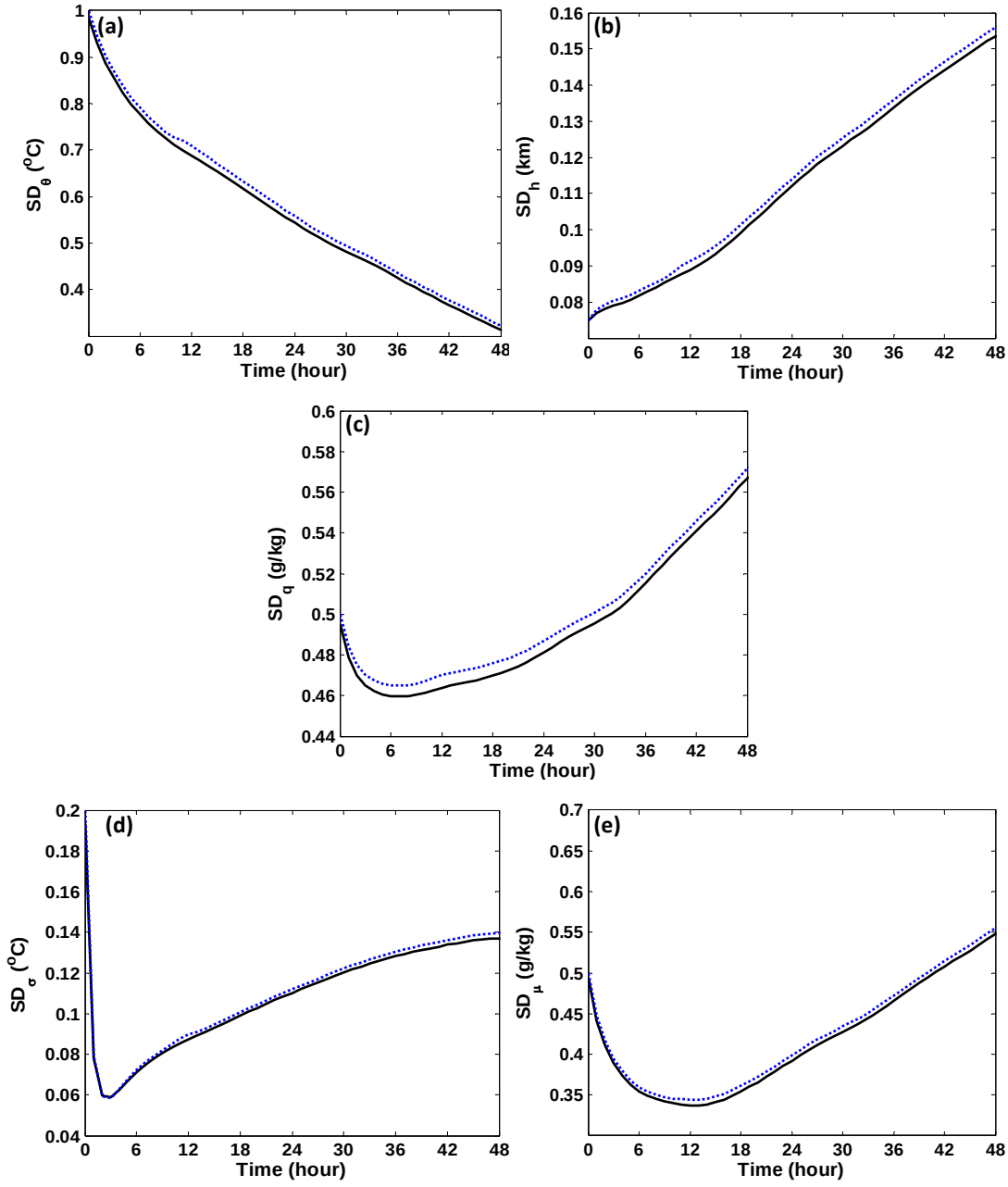


Figure 6.3 Evolution of standard deviations, (mixed-layer model) IC only, UT vs. MC

Table 6.2 Covariance matrix, (mixed-layer model) IC only, UT vs. MC

t(h)	MC					UT				
1	0.8637	-0.0148	-0.0263	0.0280	-0.0006	0.8948	-0.0154	-0.0277	0.0267	-0.0002
	-0.0148	0.0059	0.0007	-0.0027	0.0013	-0.0154	0.0061	0.0006	-0.0027	0.0016
	-0.0263	0.0007	0.0060	0.0010	0.0005	-0.0277	0.0006	0.0063	0.0015	0.0004
	0.0280	-0.0027	0.0010	0.2293	0.0319	0.0267	-0.0027	0.0015	0.2343	0.0341
	-0.0006	0.0013	0.0005	0.0319	0.1942	-0.0002	0.0016	0.0004	0.0341	0.2001
3	0.7270	-0.0300	-0.0366	0.0647	-0.0134	0.7533	-0.0311	-0.0385	0.0646	-0.0133
	-0.0300	0.0062	0.0035	-0.0065	0.0038	-0.0311	0.0065	0.0036	-0.0065	0.0040
	-0.0366	0.0035	0.0034	-0.0037	0.0022	-0.0385	0.0036	0.0034	-0.0037	0.0023
	0.0647	-0.0065	-0.0037	0.2161	0.0615	0.0646	-0.0065	-0.0037	0.2210	0.0647
	-0.0134	0.0038	0.0022	0.0615	0.1520	-0.0133	0.0040	0.0023	0.0647	0.1563
6	0.6042	-0.0422	-0.0441	0.1026	-0.0333	0.6265	-0.0436	-0.0457	0.1039	-0.0342
	-0.0422	0.0067	0.0056	-0.0104	0.0061	-0.0436	0.0069	0.0058	-0.0105	0.0063
	-0.0441	0.0056	0.0051	-0.0093	0.0051	-0.0457	0.0058	0.0052	-0.0094	0.0053
	0.1026	-0.0104	-0.0093	0.2114	0.0794	0.1039	-0.0105	-0.0094	0.2163	0.0830
	-0.0333	0.0061	0.0051	0.0794	0.1259	-0.0342	0.0063	0.0053	0.0830	0.1295
12	0.4738	-0.0528	-0.0531	0.1413	-0.0572	0.5040	-0.0529	-0.0534	0.1468	-0.0652
	-0.0528	0.0079	0.0077	-0.0168	0.0091	-0.0529	0.0083	0.0082	-0.0169	0.0088
	-0.0531	0.0077	0.0076	-0.0167	0.0089	-0.0534	0.0082	0.0080	-0.0168	0.0087
	0.1413	-0.0168	-0.0167	0.2151	0.0881	0.1468	-0.0169	-0.0168	0.2210	0.0910
	-0.0572	0.0091	0.0089	0.0881	0.1135	-0.0652	0.0088	0.0087	0.0910	0.1186
24	0.2955	-0.0586	-0.0572	0.1762	-0.0809	0.3111	-0.0601	-0.0585	0.1820	-0.0859
	-0.0586	0.0126	0.0123	-0.0350	0.0179	-0.0601	0.0130	0.0128	-0.0357	0.0184
	-0.0572	0.0123	0.0121	-0.0342	0.0177	-0.0585	0.0128	0.0125	-0.0347	0.0181
	0.1762	-0.0350	-0.0342	0.2316	0.0690	0.1820	-0.0357	-0.0347	0.2369	0.0719
	-0.0809	0.0179	0.0177	0.0690	0.1539	-0.0859	0.0184	0.0181	0.0719	0.1589

36	0.1807	-0.0559	-0.0532	0.1790	-0.0862	0.1905	-0.0577	-0.0548	0.1851	-0.0903
	-0.0559	0.0179	0.0171	-0.0555	0.0284	-0.0577	0.0185	0.0177	-0.0567	0.0292
	-0.0532	0.0171	0.0164	-0.0527	0.0273	-0.0548	0.0177	0.0170	-0.0538	0.0281
	0.1790	-0.0555	-0.0527	0.2654	0.0301	0.1851	-0.0567	-0.0538	0.2704	0.0327
	-0.0862	0.0284	0.0273	0.0301	0.2168	-0.0903	0.0292	0.0281	0.0327	0.2229
48	0.0982	-0.0475	-0.0421	0.1609	-0.0805	0.1040	-0.0493	-0.0437	0.1665	-0.0840
	-0.0475	0.0236	0.0210	-0.0788	0.0411	-0.0493	0.0243	0.0217	-0.0807	0.0423
	-0.0421	0.0210	0.0188	-0.0695	0.0368	-0.0437	0.0217	0.0195	-0.0714	0.0380
	0.1609	-0.0788	-0.0695	0.3216	-0.0299	0.1665	-0.0807	-0.0714	0.3271	-0.0281
	-0.0805	0.0411	0.0368	-0.0299	0.3003	-0.0840	0.0423	0.0380	-0.0281	0.3079

6.2 Parameter Only

In parameter only case, the uncertainty of the forecast arises from the randomness in the parameters only, the distribution of which is presented in Table 5.3. Suppose the parameter vector is denoted as $\mathbf{p} = \left(\mathrm{w}, \kappa, V_s C_\theta, V_s C_q, \gamma_\theta, \gamma_q\right)^{\mathrm{T}}$, according to the uniform distribution, the mean and covariance matrix (the units for the parameters are consistent with those from Table 5.3) of the parameter vector are

$$(-0.50, 0.25, 1.25 \times 10^{-2}, 1.25 \times 10^{-2}, 6.0, -2.0)^{\mathrm{T}}$$

and

$$\begin{pmatrix} 0.0533 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8.3333 \times 10^{-4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.0833 \times 10^{-6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.0833 \times 10^{-6} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3333 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3333 \end{pmatrix},$$

respectively. There are six parameters in total, so a number of 13 deterministic sigma points that capture the mean and covariance matrix of the parameter vector are needed

for UT method. The values for three parameters α, β, and κ in UT method are the same as those for the IC only case, i.e., the values from equation (6.1). The 13 sigma points of the parameter vector and corresponding weights for the mean and covariance are listed in Table 6.3. Note, when solving the original model numerically, the units of the variables should be consistent. Therefore, the units in Tables 6.3 for parameters *w,* $V_s C_\theta$ and $V_s C_q$ have been converted to $km\ h^{-1}$.

Table 6.3 Sigma points with weights used in UT, (mixed-layer model) Parameter only

No.	w (km/h)	κ	$V_s C_\theta$ (km/h)	$V_s C_q$ (km/h)	γ_θ (℃/km)	γ_q $(g/(kg.km))$	W^m	W^c
1	-0.0180	0.250	0.0450	0.0450	6.0000	-2.0000	-3.000	-0.250
2	-0.0078	0.250	0.0450	0.0450	6.0000	-2.0000	0.3333	0.3333
3	-0.0180	0.285	0.0450	0.0450	6.0000	-2.0000	0.3333	0.3333
4	-0.0180	0.250	0.0514	0.0450	6.0000	-2.0000	0.3333	0.3333
5	-0.0180	0.250	0.0450	0.0514	6.0000	-2.0000	0.3333	0.3333
6	-0.0180	0.250	0.0450	0.0450	6.7071	-2.0000	0.3333	0.3333
7	-0.0180	0.250	0.0450	0.0450	6.0000	-1.2929	0.3333	0.3333
8	-0.0282	0.250	0.0450	0.0450	6.0000	-2.0000	0.3333	0.3333
9	-0.0180	0.214	0.0450	0.0450	6.0000	-2.0000	0.3333	0.3333
10	-0.0180	0.250	0.0386	0.0450	6.0000	-2.0000	0.3333	0.3333
11	-0.0180	0.250	0.0450	0.0386	6.0000	-2.0000	0.3333	0.3333
12	-0.0180	0.250	0.0450	0.0450	5.2929	-2.0000	0.3333	0.3333
13	-0.0180	0.250	0.0450	0.0450	6.0000	-2.7071	0.3333	0.3333

Similar to IC only case, when the mean values of IC and BC are used, the process is a function of the parameter only. In UT, the sigma points are propagated through the dynamic system at each time and then the mean and covariance matrix are computed using the propagated sigma points. Figure 6.4 shows the propagation of the 13 sigma points (dashed lines) and their mean value (the solid line) with observations represented by plus signs. Figures 6.5 and 6.6 show the comparison of the mean values and standard

deviations with those from MC (using 20,000 samples) method. As seen from the figures, the UT approach has close approximations for the mean values and standard deviations with those from MC. Better than IC only case, the differences are hardly told from the figures.

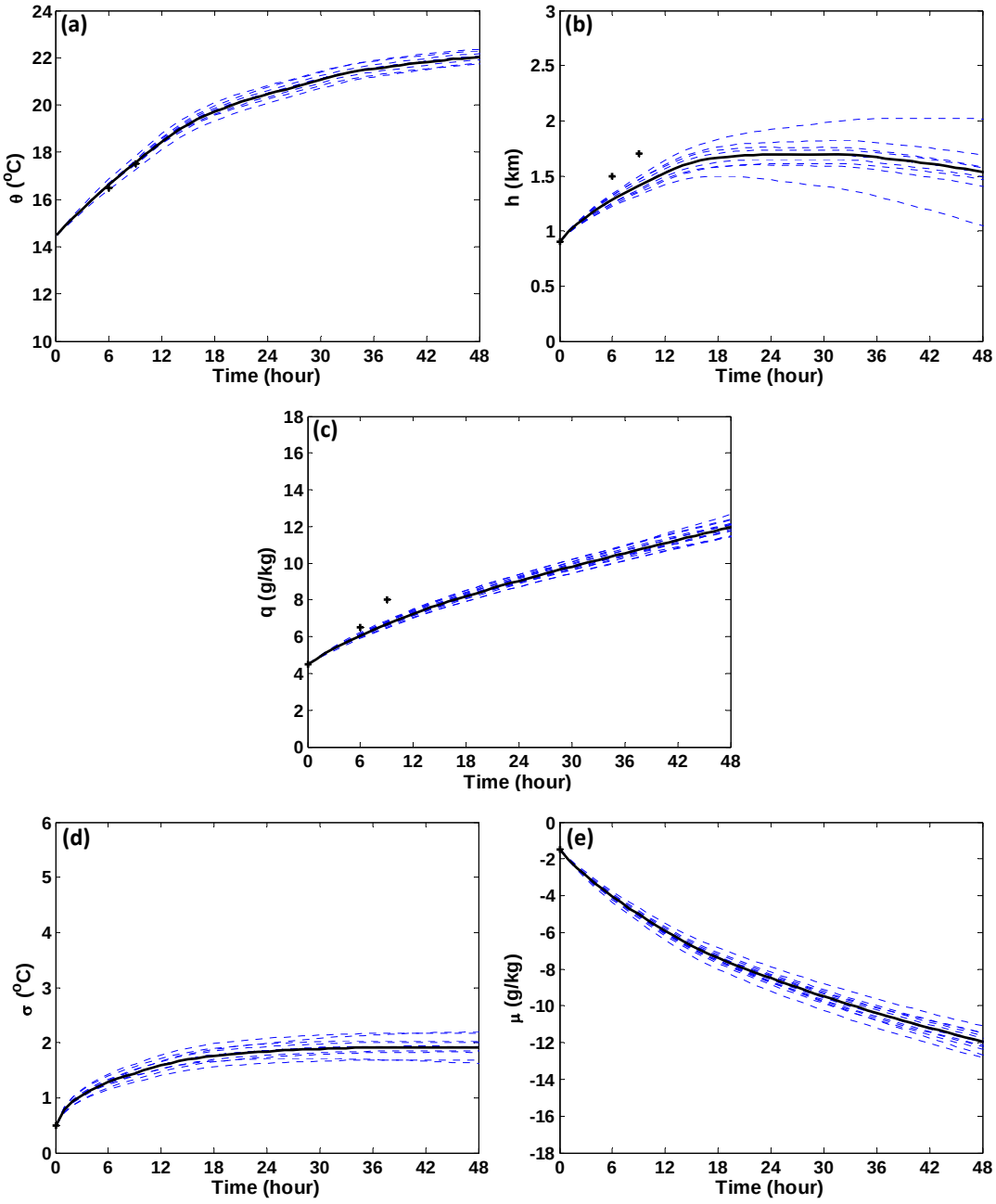


Figure 6.4 Ensemble forecast by UT, (mixed-layer model) Parameter only

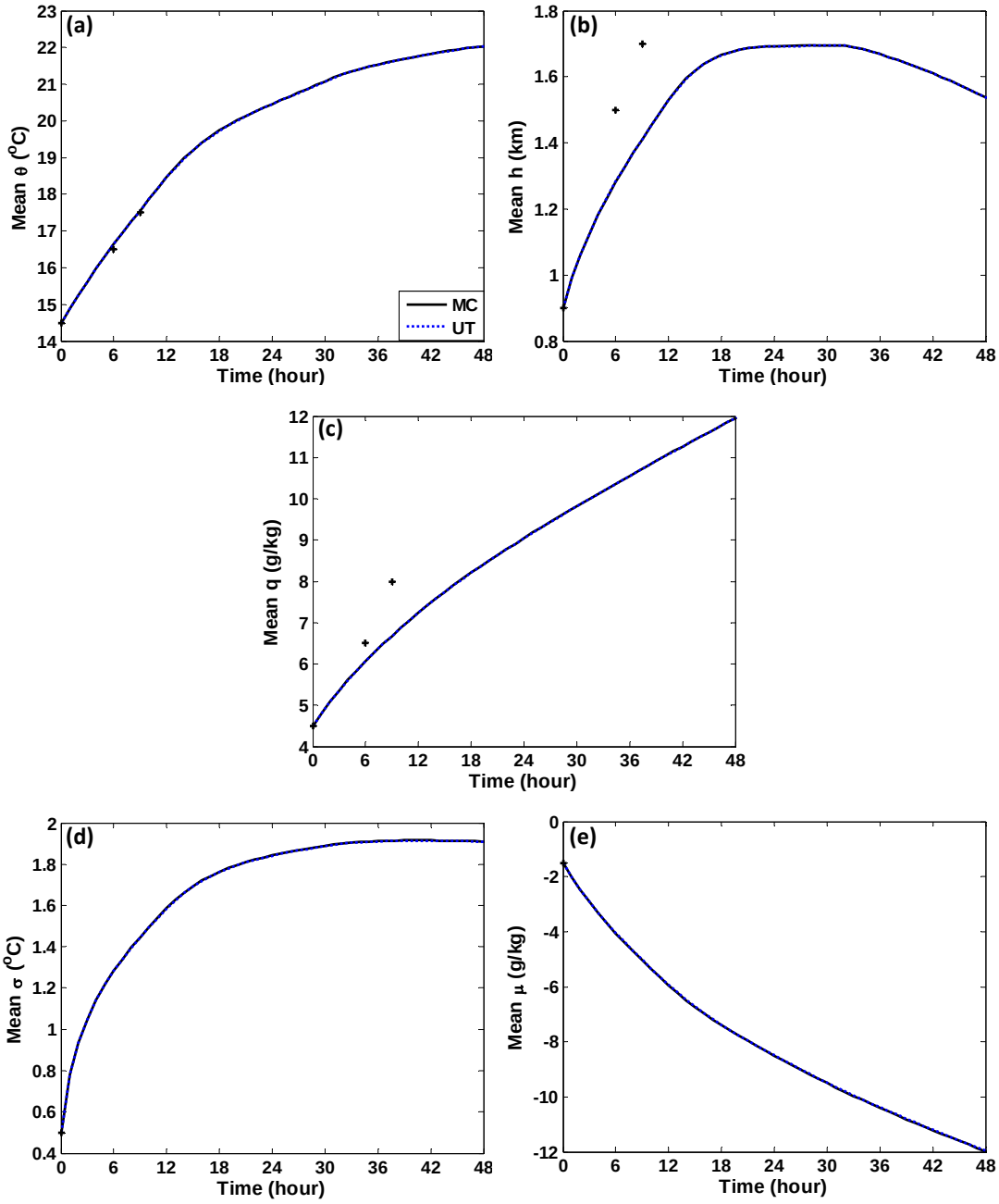


Figure 6.5 Evolution of mean values, (mixed-layer model) Parameter only, UT vs. MC

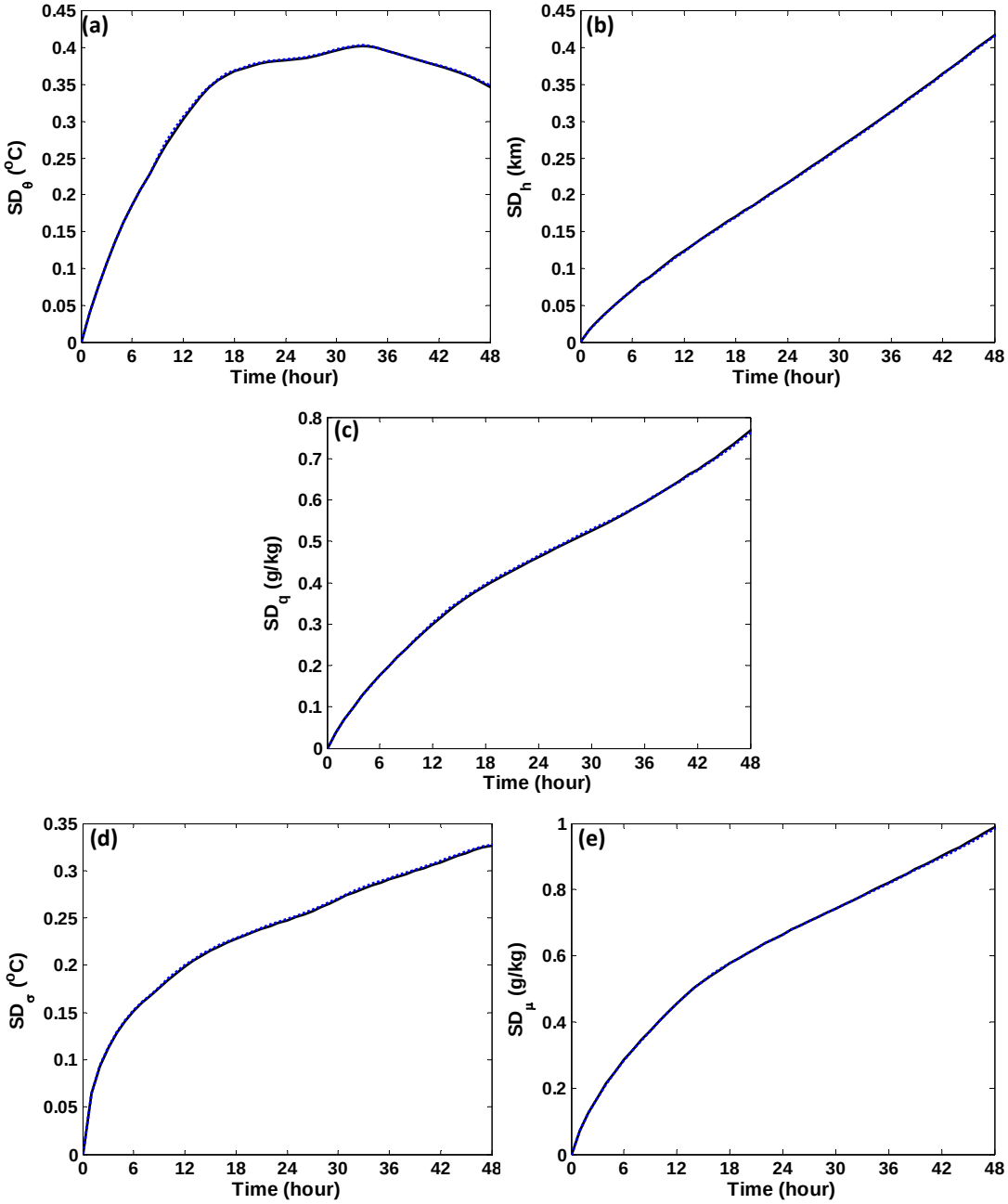


Figure 6.6 Evolution of standard deviations, (mixed-layer model) Parameter only, UT vs. MC

The estimates of the covariance matrices by UT at different times are given in Table 6.4, compared to those from MC.

Table 6.4 Covariance matrix, (mixed-layer model) Parameter only, UT vs. MC

t(h)	MC					UT				
1	0.0017	0.0005	0.0012	-0.0000	-0.0009	0.0017	0.0005	0.0012	-0.0000	-0.0009
	0.0005	0.0003	0.0005	-0.0002	-0.0002	0.0005	0.0003	0.0005	-0.0002	-0.0002
	0.0012	0.0005	0.0041	-0.0004	-0.0005	0.0012	0.0005	0.0042	-0.0004	-0.0005
	-0.0000	-0.0002	-0.0004	0.0014	-0.0008	-0.0000	-0.0002	-0.0004	0.0014	-0.0008
	-0.0009	-0.0002	-0.0005	-0.0008	0.0052	-0.0009	-0.0002	-0.0005	-0.0008	0.0051
3	0.0115	0.0023	0.0048	0.0000	-0.0052	0.0115	0.0022	0.0048	0.0001	-0.0052
	0.0023	0.0016	0.0013	-0.0012	-0.0007	0.0022	0.0016	0.0013	-0.0012	-0.0007
	0.0048	0.0013	0.0127	-0.0007	-0.0015	0.0048	0.0013	0.0128	-0.0007	-0.0014
	0.0000	-0.0012	-0.0007	0.0098	-0.0047	0.0001	-0.0012	-0.0007	0.0098	-0.0048
	-0.0052	-0.0007	-0.0015	-0.0047	0.0294	-0.0052	-0.0007	-0.0014	-0.0048	0.0292
6	0.0345	0.0046	0.0106	0.0022	-0.0152	0.0346	0.0045	0.0106	0.0023	-0.0152
	0.0046	0.0050	0.0026	-0.0044	-0.0007	0.0045	0.0049	0.0026	-0.0044	-0.0007
	0.0106	0.0026	0.0229	-0.0001	-0.0031	0.0106	0.0026	0.0231	0.0001	-0.0030
	0.0022	-0.0044	-0.0001	0.0311	-0.0103	0.0023	-0.0044	0.0001	0.0310	-0.0104
	-0.0152	-0.0007	-0.0031	-0.0103	0.0811	-0.0152	-0.0007	-0.0030	-0.0104	0.0805
12	0.0916	0.0024	0.0190	0.0181	-0.0447	0.0938	0.0020	0.0199	0.0208	-0.0467
	0.0024	0.0154	0.0076	-0.0144	0.0021	0.0020	0.0151	0.0074	-0.0143	0.0022
	0.0190	0.0076	0.0394	0.0022	-0.0062	0.0199	0.0074	0.0400	0.0032	-0.0066
	0.0181	-0.0144	0.0022	0.0903	-0.0146	0.0208	-0.0143	0.0032	0.0922	-0.0168
	-0.0447	0.0021	-0.0062	-0.0146	0.2075	-0.0467	0.0022	-0.0066	-0.0168	0.2087
24	0.1466	-0.0267	0.0109	0.0728	-0.0999	0.1476	-0.0266	0.0120	0.0750	-0.1009
	-0.0267	0.0470	0.0239	-0.0454	0.0268	-0.0266	0.0466	0.0236	-0.0458	0.0268
	0.0109	0.0239	0.0613	-0.0021	0.0003	0.0120	0.0236	0.0619	-0.0006	-0.0003
	0.0728	-0.0454	-0.0021	0.2134	-0.0517	0.0750	-0.0458	-0.0006	0.2167	-0.0553
	-0.0999	0.0268	0.0003	-0.0517	0.4419	-0.1009	0.0268	-0.0003	-0.0553	0.4420
36	0.1562	-0.0698	-0.0152	0.1395	-0.1542	0.1565	-0.0692	-0.0136	0.1407	-0.1539

	-0.0698	0.0981	0.0542	-0.1082	0.0861	-0.0692	0.0976	0.0536	-0.1085	0.0854
	-0.0152	0.0542	0.0845	-0.0304	0.0309	-0.0136	0.0536	0.0853	-0.0287	0.0302
	0.1395	-0.1082	-0.0304	0.3526	-0.1469	0.1407	-0.1085	-0.0287	0.3539	-0.1489
	-0.1542	0.0861	0.0309	-0.1469	0.6739	-0.1539	0.0854	0.0302	-0.1489	0.6711
	0.1205	-0.1012	-0.0338	0.1967	-0.1960	0.1215	-0.1006	-0.0320	0.1969	-0.1944
	-0.1012	0.1744	0.0932	-0.2378	0.2084	-0.1006	0.1736	0.0927	-0.2358	0.2049
48	-0.0338	0.0932	0.1066	-0.0932	0.0922	-0.0320	0.0927	0.1073	-0.0930	0.0928
	0.1967	-0.2378	-0.0932	0.5894	-0.3738	0.1969	-0.2358	-0.0930	0.5820	-0.3679
	-0.1960	0.2084	0.0922	-0.3738	0.9775	-0.1944	0.2049	0.0928	-0.3679	0.9670

As shown in Table 6.4, UT approach has very close approximations to those from MC. The differences between these two approaches are even smaller than those for IC only case. Same as IC only case, using the sigma point selection scheme studied in this chapter, the information of higher-order moments have not been estimated.

6.3 Discussions

Using the same model and assumption as in Chapter 5, the UT method was investigated in this chapter. It is shown from the experiments that UT can propagate the uncertainty in the input (mean and covariance in this chapter) through a nonlinear model well. However, using the sigma point selection scheme discussed in this study, neither histograms nor higher-order moments are provided by UT. Further efforts are needed on the study of HUTs in future.

Chapter 7

Discussions and Conclusions

In Chapter 2, polynomial chaos (PC) expansion approach was described from theoretical aspect. Then Chapters 4 and 5 demonstrated its application in quantifying the forecast uncertainty of a dynamical system. Once the polynomial basis is selected, the number of random variables and the highest order to truncate are decided, the remaining task is to solve for the expansion coefficients. There are two approaches to obtain the expansion coefficients, namely stochastic Galerkin (SG) and stochastic Collocation (SC) method, which are introduced in detail using an example in Chapter 4 and Chapter 5, respectively. Similarly, the theory part of unscented transformation (UT) method was described in Chapter 3 and its application in quantifying the forecast uncertainty was demonstrated using an example in Chapter 6. Both methods were compared to the well-known classical Monte Carlo (MC) approach. In this chapter, these three methods are put together and compared systematically from the following aspects.

1. Implementation

The MC approach is the most straightforward method and easy to implement. The only task needed is to randomly generate a set of samples of the random input which contributes to the forecast uncertainty. As discussed before, the random input could be initial condition, parameters and the forcing term which includes the external forcing or boundary conditions.

PC approach seems the most complex one among all three methods. The optimal polynomial basis depends on the type of the distribution for the random input and the

correspondence is given in Table (2.1). If no specific distribution is indicated, one can use the Hermite polynomial basis even though the exponential convergence rate cannot be guaranteed. Besides, the number of random variables and the truncation order are to be determined before using PC expansion. Once above elements are determined, the remaining task is to obtain the expansion coefficients using SG or SC method. If using SG, the original dynamic system needs to be transformed to a system of the expansion coefficients by Galerkin projection and then numerical methods can be used to solve the equations and obtain the expansion coefficients. While the process is sometimes tedious and it is difficult or even impossible when the system is highly complex. It is simpler to use SC, since there is no need to alter the original system. A set of collocation points are needed to run on the original system and then the coefficients are estimated by the model values on the collocation points. The key part is the selection of the collocation points according to some rules and the number of the points. This study uses the Gaussian quadrature rule together with sparse grid scheme to determine the collocation points.

UT approach is also simple. The only task needed is to generate the sigma points. Different from MC approach, these points are deterministically selected to capture the mean and covariance of the input and the number of the points is also determined. Once the three parameters α, β, and κ are chosen, the sigma points are defined.

2. Computational cost

The computation time of MC approach depends on the number of the ensemble member. If n members are used, then n model runs need to be performed on the original forecast model. Usually, a large number is needed to get a good estimate.

The computational time of PC approach depends on which method is used to solve the expansion coefficients. If using SG, the main task is to transform the original model to a system of equations for the expansion coefficients. It can be completed in advance and does not need computer time. Once the system for the expansion coefficients is obtained, the coefficients can be achieved by solving the system once, so the computer time needed is much less than MC does. Otherwise, if using SC, the computational time depends on the number of the collocation points. If the number of collocation points is p, then the original forecast model needs to run p times. In the example, less than 100 collocations were used when using multivariate polynomial basis and the performance of PC approach is good. The application of PC approach in high dimensional problems has not been tested in this study, so a conclusion for high dimensional problems cannot be reached yet. This will be one of the future research topics.

The computational time of UT approach depends on the number of the random input. If m random inputs are included, then $2m + 1$ sigma points are needed. Therefore, $2m + 1$ model runs are needed. HUT might need more sigma points and hence more time.

Table 7.1 and 7.2 show the computer execution time for each method in the mixed-layer model experiment. Table 7.1 is for initial condition (IC) only case and Table 7.2 is for parameter only case. The program was run on Matlab R2014a and the configuration of the computer is as follows:

Processor: Inter® Core(TM) i7-3770 CPU @ 3.40GHz 3.40GHz

Memory (RAM): 32.0 GB

Operation System: Windows 7, 64-bit

Table 7.1 Computing time, (mixed-layer model) IC only

Approach	MC	UT	SC (K=2)
Execution time (seconds)	1828.35	1.27	2.34

Table 7.2 Computing time, (mixed-layer model) Parameter only

Approach	MC	SC (K=2)	SC (K=3)	UT
Execution time (seconds)	1868.15	2.69	9.25	1.45

For example, in parameter only case, approximately 1868.15 seconds were needed to generate 20,000 random samples of the six parameters and simulate the model 20,000 times to get an estimate of the forecast distribution. In contrast, when using PC expansion, if the exact level K=2 (i.e., 13 collocation points), the computing time to obtain the expansion coefficients was roughly 2.69 seconds; even when the exact level is increased to K=3 (i.e., 85 collocation points), the computing time increased to 9.25 seconds, which is still much less than the time for MC approach. In the experiment, UT uses even less time (1.45 seconds) to have a good estimate on the mean and covariance of the forecast. From above discussions, compared to MC approach, both PC expansion and UT approach used much less time but gave quite similar estimates in the experiments studied in this dissertation.

3. Ability in quantifying the uncertainty

The performance of each method in quantifying the forecast uncertainty is examined based on statistical information including the mean value, covariance matrix, standard deviation, third moments, histogram, etc. Even though one of the examples studied has exact solution, which is the truth to be compared, most applications do not have exact solutions. It is widely accepted that the Monte Carlo ensemble forecast using large amount of samples can be treated as the “truth”. In the experiments, the statistics

obtained through using PC and UT approach were compared with those from MC ensemble forecast using large number of samples. Both PC and UT approaches give good estimates on the first and second order moments. Since PC approach provides a surrogate of the stochastic process in terms of polynomial expression, one can produce any number of samples by sampling the random variable used in the expansion. Therefore, a histogram can be constructed by using these samples and the probability density function can be estimated further. PC approach can also give estimates on higher-order moments. In contrast, UT with the sigma point selection scheme discussed in this study can only estimate the first and second order moments. One may estimate higher order moments by developing sigma point selection scheme in UT approach. UT cannot provide histograms.

4. Impact in data assimilation

Data assimilation is the process by which observations are incorporated into a dynamic forecast model. Applications of data assimilation arise in many fields of geosciences, perhaps most importantly in weather forecasting and hydrology. Different data assimilation techniques have been developed in the past decades. Lewis et al. (2006) provides a comprehensive summary of various approaches to dynamic data assimilation and refer to more literatures (e.g., Bishop and Toth 1999; Anderson 2001; Bishop et al. 2001; Hamill 2002; Whitaker and Hamill 2002; Evensen 2003; Wang and Bishop 2003; Hunt et al. 2004; Ott et al. 2004; Evensen 2007; Wang et al. 2008a, 2008b; Lakshmivarahan and Stensrud 2009; Wang 2011, etc.) on the recent development of data assimilation. It is known that data assimilation can benefit from a good estimation of the background error covariance matrix. From the theory and experiments in the

study, it can be seen that both PC and UT approaches perform well in quantifying the forecast uncertainty, for example, the estimates on the mean values, covariance matrices and standard deviations. So one question coming out immediately is that "can we use them in data assimilation". The answer is "Yes". PC approach with either SG or SC expansion solution has been introduced into an ensemble square root filter (EnSRF) by Li and Xiu (2009). In their application, they use PC approach to approximate the forecast model as a polynomial expression and then generate a large set of ensembles to propagate the forecast uncertainty. The observations at each data assimilation time are used to update the ensembles in which the mean and covariance matrix are updated in turn. Using some experiments, they also show that the data assimilation scheme works well.

Unscented Kalman filter (UKF) is the application of UT approach in data assimilation and many applications have shown its good performance. UKF has also been used in together with particle filter (PF) to improve the performance of PF.

To the best of my knowledge, neither PC nor UT has been tested on large scale data assimilation problems. For UT, it might be impractical since the number of sigma points depends on the dimension of the random input m. If m is large, the number of sigma points is even larger. As stated before, it will be one of the future research interests to test PC approach in large scale problems. Besides, study on the ability to capture the non-Gaussian distribution will be also interesting.

References

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. **Monthly Weather Review**, 129(12): 2884-2903.

Arnold, L., 1974: **Stochastic Differential Equations - Theory & Applications**. Wiley, New York.

Babuska, I., R. Tempone, and G. E. Zouraris, 2004: Galerkin finite element approximations of stochastic elliptic partial differential equations. **SIAM Journal on Numerical Analysis** 42(2): 800-825.

Ball, F. K., 1960: Control of inversion height by surface heating. **Quart. J. Roy. Meteor. Soc.**, 86, 483–494.

Bishop, C. H. and Z. Toth, 1999: Ensemble transformation and adaptive observations. **Journal of the atmospheric sciences**, 56(11): 1748-1765.

Bishop, C. H., B. J. Etherton and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. **Monthly Weather Review**, 129(3): 420-436.

Burk, S. D., and W. T. Thompson, 1992: Airmass modification over the Gulf of Mexico: Mesoscale model and airmass transformation model forecasts. **J. Appl. Meteor.**, 31, 925–937.

Cameron, R. H. and W. T. Martin, 1947: The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. **Annals of Mathematics**, 385-392.

Carson, D. J., 1973: The development of a dry inversion-capped convectively unstable boundary layer. **Quart. J. Roy. Meteor. Soc.**, 99, 450–467.

Crisan, D. and B. Rozovskii, 2011: **The Oxford Handbook of Nonlinear Filtering**. Oxford Handbook in Mathematics, Oxford, England.

Crisp, C. A., and J. M. Lewis, 1992: Return flow over the Gulf of Mexico. Part I: A classificatory approach with a global historical perspective. **J. Appl. Meteor.**, 31, 868–881.

Edwards, R., and S. J. Weiss, 1995: Comparisons between Gulf of Mexico sea surface temperature anomalies and southern U. S. severe thunderstorm frequency in the cool season. Preprints, 18th Conf. on Severe Local Storms, San Francisco, CA, **Amer. Meteor. Soc.**, 317–320.

Embree M., 2010: **Numerical Analysis I**. Lecture Notes. Rice University.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. **Ocean dynamics**, 53(4): 343-367.

Evensen, G., 2007: **Data assimilation: the ensemble Kalman filter**. Springer, New York.

Fox, B. L., 1999: **Strategies for Quasi-Monte Carlo**. Kluwer Academic Pub.

Frauenfelder, P., C. Schwab, and R. A. Todor, 2005: Finite elements for elliptic problems with stochastic coefficients. **Computer Methods in Applied Mechanics and Engineering** 194(2): 205-228.

Ghanem, R. G. and P. D. Spanos, 1991: **Stochastic finite elements: a spectral approach**. Springer.

Ghanem, R. G., 1999: Ingredients for a general purpose stochastic finite elements implementation. **Computer Methods in Applied Mechanics and Engineering** 168(1): 19-34.

Grigoriu, M., 2012: **Stochastic Systems: Uncertainty Quantification and Propagation**. Springer.

Hamill, T. M., 2002: **Ensemble-based data assimilation: a review**. University of Colorado and NOAA-CIRES Climate Diagnostics Center Boulder.

Heiss, F. and V. Winschel, 2008: Likelihood approximation by numerical integration on sparse grids. **Journal of Econometrics**, 144(1): 62-80.

Henry, W., 1979a: Some aspects of the fate of cold fronts in the Gulf of Mexico. **Mon. Wea. Rev.**, 107, 1078–1082.

Henry, W., 1979b: An arbitrary method of separating tropical air from “return flow” polar air. **Natl. Wea. Dig**., 4, 22–26.

Hu J., and S. Lakshmivarahan, 2015: Forecast Uncertainty Quantification of Return Flow over the Gulf of Mexico Using Monte Carlo, Generalized Polynomial Chaos and Unscented Transform Methods. **1st Annual Meeting of SIAM Central States Section**, April 11-12, 2015. Rolla, Missouri.

Hu J., S. Lakshmivarahan, and J. M. Lewis, 2015: Quantification of forecast uncertainty and data assimilation using Wiener Polynomial Chaos. **Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications**, Volume 3. (under review)

Hunt, B., E. Kalnay, E. Kostelich, E. Ott, D. Patil, T. Sauer, I. Szunyogh, J. Yorke and A. Zimin, 2004: Four-dimensional ensemble Kalman filtering. **Tellus**, 56A(4): 273-277.

Janish, P. R., and S. W. Lyons, 1992: NGM performance during cold-air outbreaks and periods of return flow over the Gulf of Mexico with emphasis on moisture-field evolution. **J. Appl. Meteor.**, 31, 995–1017.

Jazwinski, A. H., 1970: **Stochastic Processes and Filtering Theory**. Academic Press, New York

Jia, B., S. Cai, Y. Cheng, and M. Xin, 2012: Stochastic collocation method for uncertainty propagation. **AIAA/AAS Astrodynamics Specialist Conference**, Minneapolis, Minnesota.

Julier, S. J., J. K. Uhlmann, and H. F. Durrant-Whyte, 1995: A new approach for filtering nonlinear systems, **Proc. Am. Contr. Conf.**, Seattle, WA, pp. 1628–1632.

Julier, S. J., and J. K. Uhlmann, 1996: **A general method for approximating nonlinear transformations of probability distributions**. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford.

Julier, S. J., and J. K. Uhlmann, 1997a: A new extension of the Kalman filter to nonlinear systems. **Proc. AeroSense: 8th Int. Symp**. **Aerospace/Defense Sensing, Simulation and Controls**, 182-193.

Julier, S. J., and J. K. Uhlmann, 1997b: A consistent, debiased method for converting between polar and Cartesian coordinate systems. **Proc. AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls**.

Julier, S. J., and J. K. Uhlmann, 2000: A New Approach for the Nonlinear Transformation of Means and Covariances in Linear Filters. **IEEE Transactions on Automatic Control**, vol 5, no. 3, pp 477-482.

Julier, S. J., 2002: The scaled unscented transformation. **American Control Conference, Proceedings of the 2002, IEEE**.

Julier, S. J., and J. K. Uhlmann, 2004: Unscented Filtering and Nonlinear Estimation. **Proceedings of the IEEE**, VOL. 92, NO. 3.

Kallianpur, G., 1980: **Stochastic filtering theory**. Springer.

Kuo, H.-H, 2006: **Introduction to stochastic integration**. Springer.

Kushner, H. J., 1962: On the Differential Equations Satisfied by Conditional Probablitity Densities of Markov Processes, with Applications. **SIAM Journal on Control**, Vol 2, pp 106-119.

Lakshmivarahan, S. and D. Stensrud, 2009: Ensemble Kalman Filter: A innovative approach for meteorological data assimilation. **IEEE Control System Society, Special Issue**, Vol 29, pp 34-46.

Le Maıtre, O. P., O. M. Knio, H. N. Najm, and R. G. Ghanem, 2004: Uncertainty propagation using Wiener–Haar expansions. **Journal of computational physics** 197(1): 28-57.

Le Maitre, O. P. and O. M. Knio, 2010: **Spectral methods for uncertainty quantification: with applications to computational fluid dynamics**. Springer.

Lewis, J. M., C. M. Hayden, R. Merrill, and J. M. Schneider, 1989: GUFMEX: A study of return flow in the Gulf of Mexico. **Bull. Amer. Meteor. Soc.**, 70, 24–29.

Lewis, J. M., and C. A. Crisp, 1992: Return flow in the Gulf of Mexico. Part II: Variability in return-flow thermodynamics inferred from trajectories over the Gulf. ***J. Appl. Meteor.***, 31, 882–898.

Lewis, J. M., S. Lakshmivarahan and S. Dhall, 2006: **Dynamic data assimilation: a least squares approach**. Cambridge University Press.

Lewis, J. M., 2007: Use of a mixed layer model to investigate problems in operational prediction of return flow. **Mon. Wea. Rev.**, 135, 2610–2628.

Lewis J. M., 2014: Edward Epstein's Stochastic–Dynamic Approach to Ensemble Weather Prediction. **Bull. Amer. Meteor. Soc.**, 95, 99–116.

Lewis J. M., S. Lakshmivarahan, J. Hu, R. Edwards, R. A. Maddox, R. L. Thompson and S. F. Corfidi, 2015: Ensemble Forecasting of Return Flow over the Gulf of Mexico. **Electronic Journal of Severe Storms Meteorology**. (in press)

Li, J. and D. Xiu, 2009: A generalized polynomial chaos based ensemble Kalman filter with high accuracy. **Journal of computational physics**, 228(15): 5454-5469.

Lilly, D. K., 1968: Models of cloud-topped mixed layers under a strong inversion. **Quart. J. Roy. Meteor. Soc.,** 94, 292–309.

Lilly, D. K., 1987: **Mixed layers and penetrative convection** (unpublished lecture notes), School of Meteorology, Univ. of Oklahoma, Norman, OK, 15 pp. [Available from lead author.]

Liu, Q., J. M. Lewis, and J. M. Schneider, 1992: A study of cold-air modification over the Gulf of Mexico using in situ data and mixed-layer modeling. **J. Appl. Meteor.**, 31, 909–924.

Loève, M., 1977: **Probability theory**. Graduate Texts in Mathematics, 45.

Lototsky, S. and B. Rozovskii, 2006: **Stochastic differential equations: a Wiener chaos approach. From stochastic calculus to mathematical finance**. Springer, pp. 433-506.

Lototsky, S., 2011: **Chaos approach to nonlinear filtering**. Oxford University Press, pp. 231-264.

Manikin, G. S., K. E. Mitchell, and S. J. Weiss, 2000: Eta model forecasts of return flow. Preprints, **20th Conf. on Severe Local Storms**, **Amer. Meteor. Soc.**, Orlando, FL, 493–496.

Manikin, G. S., K. E. Mitchell, and S. J. Weiss, 2001: The handling of return flow in the Eta model. Preprints, **9th Conf. on Mesoscale Processes**, Amer. Meteor. Soc., Fort Lauderdale, FL, 96–99.

Manikin, G. S., K. E. Mitchell, B. S. Ferrier, and S. J. Weiss, 2002: Low level moisture in the Eta model: An update, Preprints, **21st Conf. on Severe Local Storms**, **Amer. Meteor. Soc.,** San Antonio, TX, 615–618.

Niederreiter, H., 1992: **Random Number Generation and Quasi-Monte Carlo Methods**. SIAM.

Niederreiter, H., P. Hellekalek, G. Larcher, and P. Zinterhof, 1998: **Monte Carlo and Quasi-Monte Carlo Methods**. Springer-Verlag.

Ogura, H., 1972: Orthogonal functionals of the Poisson process. **Information Theory, IEEE Transactions** on 18(4): 473-481.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. Patil and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. **Tellus**, 56A (5): 415-428.

Platzman, G. W., 1964: An exact integral of complete spectral equations for unsteady one-dimensional flow. **Tellus**, 16, 422-431.

Saaty, T. L., 1967: **Modern Nonlinear Equations**. McGraw-Hill, New York, Chapter 8.

Smith R. C., 2013: **Uncertainty Quantification, Theory, Implementation, and Applications**. SIAM: Philadelphia.

Smolyak, S. A., 1963: **Quadrature and interpolation formulas for tensor products of certain classes of functions**. Dokl. Akad. Nauk SSSR.

Soong, T. T., 1973: **Random Differential Equations in Science and Engineering**. Academic Press, New York.

Spanos, P. D. and R. Ghanem, 1989: Stochastic finite element expansion for random media. **Journal of engineering mechanics**, 115(5): 1035-1053.

Tenne D. and T. Singh, 2003: The higher order unscented filter. **Proceedings of the American Control Conference**, vol. 3, pp. 2441–2446.

Tennekes, H. and A. Driedonks, 1981: Basic entrainment equations for the atmospheric boundary layer. **Bound.-Layer Meteor.**, 20, 515–529.

Thompson, R. L., J. M. Lewis, and R. A. Maddox, 1994: Autumnal return of tropical air to the Gulf of Mexico's coastal plain. **Wea. Forecasting**, 9, 348–360.

Van Der Merwe, R., A. Doucet, N. De Freitas, and E. Wan, 2000: The unscented particle filter. paper presented at **NIPS**.

Van Der Merwe, R., 2004: **Sigma-point Kalman filters for probabilistic inference in dynamic state-space models**. Oregon Health & Science University.

Wang, X. and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. **Journal of the atmospheric sciences**, 60(9): 1140-1158.

Wang, X., D. M. Barker, C. Snyder and T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. **Monthly Weather Review**, 136(12): 5116-5131.

Wang, X., D. M. Barker, C. Snyder and T. M. Hamill, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. **Monthly Weather Review**, 136(12): 5132-5147.

Wang, X., 2011: Application of the WRF hybrid ETKF-3DVAR data assimilation system for hurricane track forecasts. **Weather and Forecasting**, 26(6): 868-884.

Wasilkowski, G. W. and H. Wozniakowski, 1995: Explicit cost bounds of algorithms for multivariate tensor product problems. **Journal of Complexity** 11(1): 1-56.

Weiss, S. J., 1992: Some aspects of forecasting severe thunderstorms during cool-season return-flow episodes. **J. Appl. Meteor.**, 31, 964–982.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. **Monthly Weather Review**, 130(7): 1913-1924.

Whittaker, E. T. and Watson, G. N., 1990: **A Course in Modern Analysis**, **4th ed**. Cambridge, England: Cambridge University Press.

Wiener, N., 1938: The homogeneous chaos. **American Journal of Mathematics**, 897-936.

Xiu, D. and G. E. Karniadakis, 2002a: The Wiener--Askey polynomial chaos for stochastic differential equations. **SIAM Journal on Scientific Computing**, 24(2): 619-644.

Xiu, D. and G. E. Karniadakis, 2002b: Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. **Computer Methods in Applied Mechanics and Engineering**, 191(43): 4927-4948.

Xiu, D. and G. E. Karniadakis, 2003: Modeling uncertainty in flow simulations via generalized polynomial chaos. **Journal of computational physics**, 187(1): 137-167.

Xiu, D. and D. M. Tartakovsky, 2006: Numerical methods for differential equations in random domains. **SIAM Journal on Scientific Computing**, 28(3): 1167-1185.

Xiu, D., 2007: Efficient collocational approach for parametric uncertainty analysis. **Communications in computational physics**, 2(2): 293-309.

Xiu, D., 2009: Fast numerical methods for stochastic computations: a review. **Communications in computational physics**, 5(2-4): 242-272.

Xiu, D., 2010. **Numerical methods for stochastic computations: a spectral method approach.** Princeton University Press.

Zakai M., 1969: On the optimal filtering of diffusion processes. **Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete**, Volume 11, pp 230-243

Appendix A

Hermite Polynomials

This appendix provides a succinct characterization of the deterministic Hermite polynomials in single and multiple variables.

1. Hermite Polynomial – Scalar Case

The Hermite polynomial $H_m(x)$ of degree m in a scalar variable x is defined by (Kuo 2006)

$$H_m(x) = (-1)^m e^{\frac{x^2}{2}} \frac{d^m}{dx^m}\left[e^{-\frac{x^2}{2}}\right]. \tag{A.1}$$

There are number of equivalent characterizations (Kuo 2006) of $H_m(x)$. In particular, the generating function for $\{H_m(x)\}_{m\geq 0}$ is given by

$$e^{tx-\frac{t^2}{2}} = \sum_{m=0}^{\infty} \frac{t^m}{m!} H_m(x). \tag{A.2}$$

In generating a polynomial for a specific degree m, the following formula is useful,

$$H_m(x) = \sum_{k=0}^{\left[\frac{m}{2}\right]} (-1)^k \binom{m}{2k} (2k-1)!!\, x^{m-2k}, \tag{A.3}$$

where $[x]$ denotes the integer part of x,

$$\binom{m}{2k} = \frac{m!}{2k!\,(m-2k)!},$$

and $m!$ is the usual factorial of m and $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2k-1)$.

2. Orthogonality Property

Let

$$w(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \tag{A.4}$$

denotes the standard Gaussian density function.

For integers m and k, define the inner product

$$\langle H_m H_k \rangle_w = \int_{-\infty}^{\infty} H_m(x) H_k(x)\, w(x) dx. \tag{A.5}$$

This inner product induces a norm $\|H_m\|_w^2$ of $H_m(x)$ defined by

$$\|H_m(x)\|_w^2 = \int_{-\infty}^{\infty} H_m^2(x)\, w(x) dx. \tag{A.6}$$

It can be verified that

$$\langle H_m H_k \rangle_w = \|H_m\|_w^2 \delta_{mk}, \tag{A.7}$$

where $\delta_{mk} = 0,$ if $m \neq k;$ and $= 1,$ if $m = k.$ That is, H_m and H_k are orthogonal for $m \neq k$. And

$$\|H_m\|_w^2 = m!. \tag{A.8}$$

Consequently, $\{H_m(x)\}_{m\geq 0}$ constitute an orthogonal system of polynomials and $\left\{\frac{H_m(x)}{\sqrt{m!}}\right\}_{m\geq 0}$ constitute an orthonormal system.

Examples of $H_m(x)$ for $0 \leq m \leq 4$ are given in Table A.1.

Table A.1 A list of $H_m(x)$, $0 \leq m \leq 4$

Degree m	$H_m(x)$	$\|H_m\|_w^2$
0	1	1
1	x	1
2	$x^2 - 1$	2
3	$x^3 - 3x$	$3! = 6$
4	$x^4 - 6x^2 + 3$	$4! = 24$

3. Hermite Polynomials – Multivariate Case

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in R^n$ and define the *n*-variate weight function

$$W(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \mathrm{e}^{-\frac{\mathbf{x}^{\mathrm{T}}\mathbf{x}}{2}} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \prod_{i=1}^{n} w(x_i), \tag{A.9}$$

where $w(x_i)$ is defined in (A.4). Let

$$m = p_1 + p_2 + \cdots + p_n \text{ with } 0 \leq p_i \leq m. \tag{A.10}$$

be an additive partition of the integer $m \geq 0$. Such a partition of m is defined as a n-turple $(p_1, p_2, \ldots, p_n)$. For example, when $n = 2$, there are 3 partitions – $\{(2,0), (1,1), (0,2)\}$ of $m = 2$.

Given m and one of its partitions $(p_1, p_2, \ldots, p_n)$, define an n-variate homogeneous Hermite polynomial of degree m,

$$H_{p_1 p_2 \ldots p_n}(\mathbf{x}) = (-1)^m e^{\frac{\mathbf{x}^{\mathrm{T}}\mathbf{x}}{2}} \frac{\partial^m}{\partial {x_1}^{p_1} \partial {x_2}^{p_2} \ldots \partial {x_n}^{p_n}} e^{-\frac{\mathbf{x}^{\mathrm{T}}\mathbf{x}}{2}}. \tag{A.11}$$

Using (A.9) in (A.11), it can be verified that

$$H_{p_1 p_2 \ldots p_n}(\mathbf{x}) = \prod_{i=1}^{n} (-1)^{p_i} e^{\frac{{x_i}^2}{2}} \left(\frac{\partial^{p_i}}{\partial {x_i}^{p_i}} e^{-\frac{{x_i}^2}{2}} \right) = \prod_{i=1}^{n} H_{p_i}(x_i). \tag{A.12}$$

By combining the multiplicative decomposition of the multi-variate Hermite polynomials in terms of the univariate Hermite polynomials given in (A.1) and the orthogonality of the latter expressed in (A.5)-(A.7), the orthogonality of the multivariate Hermite polynomials can be readily inferred.

Hence, if $m = p_1 + p_2 + \cdots + p_n$ and $k = q_1 + q_2 + \cdots + q_n$, then

$$\begin{aligned} \langle H_m(\mathbf{x}) H_k(\mathbf{x}) \rangle_{\mathrm{W}} &= 0, \text{ (if } m \neq k) \\ &= \prod_{i=1}^{n} \left\| H_{p_i}(x_i) \right\|_w^2 \delta_{p_i q_i}. \text{ (if } m = k) \end{aligned} \tag{A.13}$$

Clearly,

$$\left\| \mathrm{H}_{p_1 p_2 \ldots p_n}(\mathbf{x}) \right\|_W^2 = \prod_{i=1}^{n} p_i!. \tag{A.14}$$

While there is a unique total ordering of singly indexed scalar Hermite polynomials $H_k(x)$ as shown in Table A.1, there are many ways of ordering the multi-indexed Hermite polynomials in (A.12).

A useful ordering of this latter class of polynomials is called **graded lexicographic ordering**. In this ordering scheme, polynomials of lower total degree precede those of

higher degree. Among polynomials of same degree, the members are ordered according to the lexicographic order induced by the natural ordering of the indeterminates, that is $x_1 > x_2 > \cdots > x_n$. Thus, polynomials degree one precede those of degree two. For $m = 3, n = 2$, the lexicographic ordering of the two tuples is given by (3,0), (2,1), (1,2), (0,3).

It can be verified that there are exactly $\binom{m+n-1}{n}$ members in the lexicographic ordering of n tuples $(p_1, p_2, \ldots, p_n)$, such that $\sum_{i=1}^{n} p_i = m$. Hence, there are this many linearly independent n -variate Hermite polynomials of degree m. Further, it can be verified that the total number of linearly independent n -variate Hermite polynomials of degree less than or equal to n is given by $\binom{m+n}{n}$.

Table A.2 provides a list of the set of all 15 two variate ($n = 2$) Hermite polynomials of degree less than or equal to 4. The last column in Table A.2 gives the norm term $\left\|\mathrm{H}_{p_1 p_2}(x_1, x_2)\right\|_W^2$.

4. Hilbert Space

Let $L_2 = L_2(R^n, W)$ denote the set of all square integrable functions on R^n, that is

$$\mathrm{L}_2 = \left\{f: R^n \to R: \int_{R^n} f^2(\mathbf{x}) W(\mathbf{x}) d\mathbf{x} < \infty\right\}, \tag{A.15}$$

where W is defined in (A.9). If $f, g \in L_2$, then a natural inner product on L_2 is defined by

$$\langle f, g\rangle_W = \int_{R^n} f(\mathbf{x}) g(\mathbf{x}) W(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{A.16}$$

Hence, the norm $\|f\|_W$ is defined by

$$\|f\|_W^2 = \int_{R^n} f^2(\mathbf{x}) W(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{A.17}$$

It is well known that L_2 is a Hilbert space which is an infinite dimensional, complete, normal linear space where the norm is induced by the inner product.

Table A.2 Two-variate Hermite polynomials, degree less than or equal to 4

Degree m	**Multi index** $(p_1 p_2)$	$H_{p_1 p_2}(x_1, x_2)$	$H_{p_1}(x_1) H_{p_2}(x_2)$	$\left\|\mathrm{H}_{p_1 p_2}(x_1, x_2)\right\|_W^2$
0	0 0	1	1	1
1	1 0	x_1	$H_1(x_1)H_0(x_2)$	1
	0 1	x_2	$H_0(x_1)H_1(x_2)$	1
2	2 0	$x_1^2 - 1$	$H_2(x_1)H_0(x_2)$	2
	1 1	$x_1 x_2$	$H_1(x_1)H_1(x_2)$	1
	0 2	$x_2^2 - 1$	$H_0(x_1)H_2(x_2)$	2
3	3 0	$x_1^3 - 3x_1$	$H_3(x_1)H_0(x_2)$	6
	2 1	$x_1^2 x_2 - x_2$	$H_2(x_1)H_1(x_2)$	2
	1 2	$x_1 x_2^2 - x_1$	$H_1(x_1)H_2(x_2)$	2
	0 3	$x_2^3 - 3x_2$	$H_0(x_1)H_3(x_2)$	6
4	4 0	$x_1^4 - 6x_1^2 + 3$	$H_4(x_1)H_0(x_2)$	24
	3 1	$x_1^3 x_2 - 3x_1 x_2$	$H_3(x_1)H_1(x_2)$	6
	2 2	$x_1^2 x_2^2 - x_1^2 - x_2^2 + 1$	$H_2(x_1)H_2(x_2)$	4
	1 3	$x_1 x_2^3 - 3x_1 x_2$	$H_1(x_1)H_3(x_2)$	6
	0 4	$x_2^4 - 6x_2^2 + 3$	$H_0(x_1)H_4(x_2)$	24

5. Basis for L_2

Let $\mathrm{P_m}$ denote the linear span of set of all the *n*-variate Hermite polynomials $H_k(\mathbf{x})$ of degree $k \leq m$. That is,

$$\mathrm{P}_m = \left\{P(\mathbf{x}) | P(\mathbf{x}) = \sum_{k=0}^{m} \sum_{p_1+p_2+\cdots+p_n=k} a_{p_1,p_2,\ldots,p_n} H_{p_1,p_2,\ldots,p_n}(\mathbf{x})\right\},$$

and

$$p_1, p_2, \ldots, p_n \in R. \tag{A.18}$$

Since there are $\binom{m+n}{n}$ linearly independent *n*-variate Hermite polynomials of degree less than or equal to m, P_m is a linear vector space of finite dimension. Let $\mathrm{P}_{m-1}^{\perp}$ denote the orthogonal complement of P_{m-1}. That is, members of P_m and $\mathrm{P}_{m-1}^{\perp}$ are

mutually orthogonal. Now define the set of all homogeneous polynomials HP_m of degree exactly equal to m as

$$\mathrm{HP}_m = \mathrm{P}_m \cap \mathrm{P}_{m-1}^{\perp} . \tag{A.19}$$

It can be verified that members of HP_m are mutually orthogonal to those in P_{m-1}. Define

$$\mathrm{HP} = \oplus_{m=0}^{\infty} \mathrm{HP}_m. \tag{A.20}$$

The direct sum of homogeneous polynomials of degree $m \geq 0$. Clearly, for a fixed n,

$$|\mathrm{HP}| = \lim_{N\to\infty} \sum_{m=0}^{N} \binom{m+n-1}{n} = \infty. \tag{A.21}$$

Now a number of properties are stated without proof.

P1. Basis for L_2

$\mathrm{HP} \subset \mathrm{L}_2$ and HP constitutes a basis for L_2. Thus, any $f \in \mathrm{L}_2$ can be uniquely expressed as

$$f(\mathbf{x}) = \sum_{m=0}^{\infty} \sum_{p_1+p_2+\cdots+p_n=m} a_{p_1,p_2,\ldots,p_n} H_{p_1,p_2,\ldots,p_n}(\mathbf{x}), \tag{A.22}$$

where

$$a_{p_1,p_2,\ldots,p_n} = \frac{\langle f(\mathbf{x}), H_{p_1,p_2,\ldots,p_n}(\mathbf{x})\rangle}{\left\| H_{p_1,p_2,\ldots,p_n}(\mathbf{x})\right\|^2}. \tag{A.23}$$

P2. Orthogonal Projection

For any N finite, define

$$\Pi_N(f) = f_N(\mathbf{x}) = \sum_{m=0}^{N} \sum_{p_1+p_2+\cdots+p_n=m} a_{p_1,p_2,\ldots,p_n} H_{p_1,p_2,\ldots,p_n}(\mathbf{x}). \tag{A.24}$$

Then it can be verified that $\Pi_N(f)$ is the orthogonal projection of $f(\mathbf{x})$ onto the subspace P_N.

Define the error in the projection as

$$\varepsilon_N(\mathbf{x}) = f(\mathbf{x}) - f_N(\mathbf{x}). \tag{A.25}$$

Then, it can be verified that

$$\langle f_N(\mathbf{x}), \varepsilon_N(\mathbf{x}) \rangle = 0 \text{ for each } N > 0. \tag{A.26}$$

P3. Mean Square Convergence

It can be verified that

$$\lim_{N\to\infty} \|f(\mathbf{x}) - f_N(\mathbf{x})\|^2 = 0, \tag{A.27}$$

i.e., the quality of the projection $f_N(\mathbf{x})$ improves as N grows large.

Example A.1 From (A.2), it follows that

$$e^{tx} = e^{\frac{t^2}{2}} \sum_{m=0}^{\infty} \frac{t^m}{m!} H_m(x), \tag{A.28}$$

and

$$e^{-tx} = e^{\frac{t^2}{2}} \sum_{m=0}^{\infty} \frac{t^m}{m!} (-1)^m H_m(x). \tag{A.29}$$

Since $H_m(x) = H_m(-x)$ for m even and $H_m(-x) = -H_m(x)$ for m odd. By truncating the infinite sum in (A.28) and (A.29), a good family of approximation to e^{tx} and e^{-tx} can be obtained. Using these one can readily obtain approximation to $sin(tx)$, $cos(tx)$, etc.

It would be a good exercise to compute the quality of these approximations for varying degree of truncation and ranges of values for x.

Appendix B

Hermite Polynomial Chaos

Let (Ω, f, P) be a standard probability space where Ω represents the set of all elementary events, f denotes the $\sigma-$ algebra of subsets of simple events and P is the probability measure defined on the members of f. Let $x: \Omega \to \mathrm{R}$ be a real valued random variable. Let $P_x(x)$ be the distribution induced by x and let $p_x(x)$ be the corresponding probability density function. Then, the properties of x can be equivalently described using the triplet $\left(\mathrm{R}, \mathrm{B}, p_x(x)\right)$ where B denotes the Borel $\sigma-$ algebra over R, and $p_x(x)$ is the density of x.

Let x and y be two real valued random variables with joint density $p_{x,y}(x,y)$. Define an inner product

$$\langle x, y\rangle = E(xy) = \int_{\Omega} x(w)y(w)dp(w) = \int_{\mathrm{R}} \int_{R} xyp_{x,y}(x,y)dxdy, \qquad \text{(B.1)}$$

and the corresponding norm (second moment)

$$\|x\|^2 = E(x^2) = \int_{\Omega} x^2(w)dp(w) = \int_{R} x^2 p_x(x)dx. \qquad \text{(B.2)}$$

Let $L_2(\Omega)$ denotes the set of all random variables with finite second moment, that is,

$$L_2(\Omega) = \left\{x: \Omega \to \mathrm{R} \middle| \int_{\Omega} x^2(w)dp(w) < \infty\right\}. \qquad \text{(B.3)}$$

It can be verified $L_2(\Omega)$ is a Hilbert space.

Let x be a Gaussian random variable with mean m and variance σ^2. Then

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]. \qquad \text{(B.4)}$$

A random variable is said to be centered if expectation is zero. Let x be a centered Gaussian random variable, that is, $x \sim \mathrm{N}(0, \sigma^2)$. Then it can be verified

$$E(x^m) = 0, \text{ (if } m \text{ is odd)}$$

$$= 1 \cdot 3 \cdot 5 \cdots (m-1)\sigma^m. \text{ (if } m \text{ is even)} \tag{B.5}$$

Let G denotes the set of all centered Gaussian random variables (differing only in their variance). Clearly, G is an infinite set. Let $\overline{\mathrm{G}}$ denotes the closure of the linear span of G. It can be verified $\overline{\mathrm{G}} \subset \mathrm{L}_2(\Omega)$ and is itself a Hilbert space, called the Gaussian Hilbert space. If $\xi_1, \xi_2, \ldots, \xi_n \in \overline{\mathrm{G}}$, then it is well known that

$$E(\xi_1 \xi_2 \ldots \xi_n) = \sum \prod_{(i_1,i_2)} E\left(\xi_{i_1} \xi_{i_2}\right), \tag{B.6}$$

where (i_1, i_2) runs through the pairwise distinct partition of $\{1,2,\ldots,n\}$ and the sum is over all such partitions. For example,

$$E(\xi_1\xi_2\xi_3\xi_4) = E(\xi_1\xi_2)E(\xi_3\xi_4) + E(\xi_1\xi_3)E(\xi_2\xi_4) + E(\xi_1\xi_4)E(\xi_2\xi_3). \tag{B.7}$$

Recall, the probability density of a standard Gaussian random variable $\xi$is given by

$$p(\xi) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{\xi^2}{2}\right]. \tag{B.8}$$

Since $p(\xi)$ in (B.8) is the same as $w(x)$ in (A.4), Appendix A, it is immediate that the properties of the Hermite polynomials $H_m(x)$ given in Appendix A directly carry over to $H_m(\xi)$ where ξ is a centered standard Gaussian random variable.

The following properties of $H_m(\xi)$ can be easily verified:

(1) Examples of $H_m(\xi)$ are obtained by replacing x by ξ in Table A.1.

(2) $E[H_m(\xi)] = 0$, if $m > 0$.

(3) $\{H_m(\xi)\}_{m\geq 0}$ are orthogonal, that is,

$$\langle H_m(\xi), H_k(\xi)\rangle = E[H_m(\xi)H_k(\xi)] = 0 \text{ if } m \neq k.$$

(4) The norm of $H_m(\xi)$ is,

$$\langle H_m(\xi), H_m(\xi)\rangle = \|H_m(\xi)\|^2 = E[H_m^2(\xi)] = m!.$$

(5) $\left\{\frac{H_m(\xi)}{\sqrt{m!}}\right\}_{m\geq 0}$ form an orthonormal system of Hermite polynomials.

The above properties can be readily extended to multivariate Hermite polynomials over a set of n centered Gaussian random variables $\xi_1, \xi_2, \dots, \xi_n$:

(6) Let $p_1 + p_2 + \cdots + p_n = m$, where $0 \leq p_i \leq m$ for $1 \leq i \leq n$. Then

$$H_{p_1,p_2,\dots,p_n}(\xi_1, \xi_2, \dots, \xi_n) = H_{p_1}(\xi_1) H_{p_2}(\xi_2) \dots H_{p_n}(\xi_n).$$

(7) $\left\| H_{p_1,p_2,\dots,p_n}(\xi_1, \xi_2, \dots, \xi_n) \right\|^2 = var\left[H_{p_1,p_2,\dots,p_n}(\xi_1, \xi_2, \dots, \xi_n) \right] = p_1!\, p_2! \dots p_n!$.

(8) The definitions of the sets P_m, HP_m directly carry over to the case of Hermite polynomial over a finite set of centered Gaussian random variables.

(9) The linear space HP_m are called m^{th} order polynomial chaos in n centered Gaussian random variables.

(10) The P_m is called homogeneous chaos in n centered Gaussian random variables.

Appendix C

Gaussian Quadrature Rule

Integration calculus is widely used throughout various engineering fields. However, sometimes it is difficult or even impossible to evaluate those expressions involving integrals analytically. For this reason, many numerical methods have been developed to simplify the integral. In this appendix, Gaussian quadrature rule is presented, please refer to (Embree 2010) for more details. The aim is to approximate one dimensional integrals in the form

$$I = \int_a^b f(x)dx, \tag{C.1}$$

where $f(x)$ is called the integrand, a and b are the lower and upper limits of the integration, respectively.

1. Trapezoid Rule

It's known that the trapezoid rule

$$I(f) = \frac{b-a}{2}\big(f(a) + f(b)\big), \tag{C.2}$$

is exact for any constant and linear functions, but not all quadratics.

2. A Special Two-point Quadrature Rule

Now let's consider a more general two-point quadrature rule which is different from the trapezoid rule. Here, the two points are not predefined as a and b, but as unknowns x_1 and x_2. The integration in (C.1) is approximated as

$$I = \int_a^b f(x)dx \approx w_1 f(x_1) + w_2 f(x_2). \tag{C.3}$$

The four unknowns x_1 , x_2 , w_1 and w_2 are evaluated by assuming the above approximation gives exact value for a general three order polynomial.

In order to do this, it can be assumed that the expression is exact for $\int_a^b 1dx$, $\int_a^b xdx$, $\int_a^b x^2dx$ and $\int_a^b x^3dx$. The reason is that the linear combination of the above four integrands is a general three order polynomial. From the assumption, four equations will be given as follows

$$\int_a^b 1dx = w_1 + w_2,$$

$$\int_a^b xdx = \frac{b^2-a^2}{2} = w_1x_1 + w_2x_2,$$

$$\int_a^b x^2dx = \frac{b^3-a^3}{3} = w_1x_1^2 + w_2x_2^2,$$

$$\int_a^b x^3dx = \frac{b^4-a^4}{4} = w_1x_1^3 + w_2x_2^3. \quad \text{(C.4)}$$

These four equations can be solved to give a single acceptable solution

$$w_1 = w_2 = \frac{b-a}{2},$$

$$x_1 = \frac{b+a}{2} - \frac{\sqrt{3}}{6}(b-a),$$

$$x_2 = \frac{b+a}{2} + \frac{\sqrt{3}}{6}(b-a). \quad \text{(C.5)}$$

Hence, the two-point quadrature rule is

$$\int_a^b f(x)dx \approx \frac{b-a}{2} f\left(\frac{b+a}{2} - \frac{\sqrt{3}}{6}(b-a)\right) + \frac{b-a}{2} f\left(\frac{b+a}{2} + \frac{\sqrt{3}}{6}(b-a)\right). \quad \text{(C.6)}$$

Notice that the two points x_1 and x_2 are in the interval [*a, b*]. If it were not the case, they could not be used as the quadrature nodes for the reason that $f(x)$ might not be defined outside [*a, b*]. Since two points are used here, so it's called two-point quadrature rule. Next, the generalization in higher order case will be considered.

3. Generalization of Quadrature Rule for Higher Orders

Now, using two ($n = 2$) points, one can exactly integrate polynomials of degree up to three ($2 \times n - 1$), one might anticipate using n points to integrate polynomials of degree up to $2n - 1$.

Here, the interest is in constructing a quadrature rule of the form

$$\int_a^b f(x)w(x)dx \approx \sum_{i=1}^n f(x_i)w_i, \tag{C.7}$$

where $w(x)$ is a weight function, non-negative over [*a, b*] and takes the value of zero only on a set of measure zero. It is expected to use n points to exactly evaluate the integration in (C.7) for polynomials of degree up to $2n - 1$. That is, for arbitrary $p(x) \in \mathcal{P}_{2n-1}$,

$$\int_a^b p(x)w(x)dx = \sum_{i=1}^n p(x_i)w_i. \tag{C.7}$$

Orthogonal polynomials such as Hermite polynomials relating to Gauss distribution and other polynomials with various distributions will play a prominent role. Let $\{\phi_j\}_{j=0}^n, j$ as the degree of polynomial, be a system of orthogonal polynomials with respect to the inner product defined as

$$\langle f, g \rangle = \int_a^b f(x)g(x)w(x)dx. \tag{C.8}$$

Using polynomial division, $p(x)$ can be rewritten as

$$p(x) = \phi_n(x)q(x) + r(x), \tag{C.9}$$

for some $q(x)$ and $r(x)$ with degree less than n, which depend on $p(x)$. Then the integration of $p(x)$ becomes

$$\int_a^b p(x)w(x)dx = \int_a^b \phi_n(x)q(x)w(x)dx + \int_a^b r(x)w(x)dx. \tag{C.10}$$

According to the fact that the orthogonal polynomial ϕ_n is orthogonal to all polynomials with degree less than n, that is,

$$\int_a^b \phi_n(x)q(x)w(x)dx = 0. \tag{C.11}$$

Therefore

$$\int_a^b p(x)w(x)dx = \int_a^b r(x)w(x)dx. \tag{C.12}$$

The task now is to pick the interpolation points $\{x_i\}$ and weights $\{w_i\}$ and apply the quadrature rule to the integration given in (C.10) and according to (C.12), the following equation can be obtained

$$I(p) = \sum_{i=1}^n p(x_i)w_i = \sum_{i=1}^n \phi_n(x_i)q(x_i)w_i + \sum_{i=1}^n r(x_i)w_i = \int_a^b r(x)w(x)dx. \tag{C.13}$$

Recall that the quadrature rule is constructed so that it exactly integrates a degree $(n-1)$ polynomial interpolant to the integrand, here $r(x)$ is a degree $(n-1)$ polynomial (for example, interpolating $r(x)$ with $x_i, i = 1, \dots, n$, one can then obtain the weights w_i, it is exact for degree $(n-1)$ polynomials), i.e.,

$$\int_a^b r(x)w(x)dx = \sum_{i=1}^n r(x_i)w_i. \tag{C.14}$$

Hence, the agreement can be forced between $I(p)$ and $\int_a^b p(x)w(x)dx$ provided

$$\sum_{i=1}^n \phi_n(x_i)q(x_i)w_i = 0. \tag{C.15}$$

Assumptions about the polynomial $q(x)$ cannot be made since it changes with the polynomial $p(x)$, so the properties of $\phi_n(x)$ will be examined.

Theorem (Roots of Orthogonal polynomials): Let $\{\phi_i(x)\}_{i=1}^n$ be a system of orthogonal polynomials on [*a, b*] with weight function $w(x)$, then $\phi_i(x), i = 1, \dots, n$, has i distinct real roots $\left\{x_j^{(i)}\right\}_{j=1}^i$ with $x_j^{(i)} \in [a, b]$.

Proof. Suppose $\phi_i(x)$ changes sign at $k < i$ distinct roots $\left\{x_j^{(i)}\right\}_{j=1}^{k}$ in the interval [*a, b*]. Then define $\psi(x)$ as

$$\psi(x) = \prod_{j=1}^{k}(x - x_j) \in \mathcal{P}_k. \tag{C.16}$$

$\psi(x)$ is a function which changes sign at exactly the same points as $\phi_i(x)$ does on [*a, b*]. Therefore, $\phi_i(x)\psi(x)$ does not change sign on [*a, b*]. As is known that $w(x) \geq 0$ on [*a, b*], $\phi_i(x)\psi(x)w(x)$ does not change sign on [*a, b*]. However, as $\psi(x) \in \mathcal{P}_k$ and $k < i$, the following formula will hold

$$\int_a^b \phi_i(x)\psi(x)w(x) = 0, \tag{C.17}$$

based on the fact that the orthogonal polynomial ϕ_n is orthogonal to all polynomials with degree less than n. It's obvious a contradiction here. Thus, ϕ_n must have at least n distinct roots on [*a, b*] and since it's a polynomial with degree n, it cannot have more than n roots. Therefore, ϕ_n has n distinct roots in [*a, b*].

Based on the above theorem, $\phi_n(x)$ has n distinct real roots $\left\{x_j^{(n)}\right\}_{j=1}^{n}$ with $x_j^{(n)} \in [a,\ b]$. If these n distinct values are picked, the equation (C.15) will hold, hence a quadrature rule that is exact for all polynomials with degree up to $2n - 1$ will be achieved.

Now comes to the Gaussian quadrature rule: for a general function $f(x)$, approximate the integration in (C.7) as the integration of a polynomial which interpolates $f(x)$ at the roots of the orthogonal polynomial $\phi_n(x)$. For simplicity, the n distinct real roots $\left\{x_j^{(n)}\right\}_{j=1}^{n}$ of $\phi_n(x)$ are simplified as $\{x_i\}_{i=1}^{n}$. That is, let $p_n(x)$ be the

interpolant of $f(x)$ at the roots of the orthogonal polynomial $\phi_n(x)$, then Gaussian quadrature rule is in the form

$$\int_a^b f(x)w(x)dx \approx \int_a^b p_n(x)w(x)dx. \quad \text{(C.18)}$$

The implementation can be as follows:

Using Lagrange interpolation method, $p_n(x)$ can be written in the Lagrange basis,

$$p_n(x) = \sum_{i=1}^n f(x_i)l_i(x), \quad \text{(C.19)}$$

where the basis $l_i(x)$ is defined as

$$l_i(x) = \prod_{j=1, j\neq i}^n \frac{x-x_j}{x_i-x_j}. \quad \text{(C.20)}$$

The integration of $p_n(x)$ can be written as

$$\int_a^b p_n(x)w(x)dx = \int_a^b \sum_{i=1}^n f(x_i)l_i(x)\, w(x)dx = \sum_{i=1}^n f(x_i) \int_a^b l_i(x)w(x)dx. \quad \text{(C.21)}$$

It is obvious that we can use the formula

$$w_i = \int_a^b l_i(x)w(x)dx, \quad \text{(C.22)}$$

to obtain the weights $\{w_i\}$.

Gaussian Quadrature Rule:

$$\int_a^b f(x)w(x)dx \approx I(f) = \sum_{i=1}^n f(x_i)w_i, \quad \text{(C.23)}$$

where the points $\{x_i\}_{i=1}^n$ are the n roots of a degree n orthogonal polynomial $\phi_n(x)$ on [*a, b*] with weight function $w(x)$, and the weights are calculated using (C.22). $I(f)$ is exact for all polynomials of degree less than or equal to $2n-1$.

4. Examples of Gaussian Quadrature

Gauss-Legendre Quadrature

The best known Gaussian quadrature rule integrates functions on interval [-1, 1] with the weight function $w(x) = 1$, i.e., $\int_{-1}^1 f(x)dx$. However, $w(x) = \frac{1}{2}$ is used in the study

just to make it as the probability density function of a uniformly distributed random variable on interval [-1, 1]. As shown in Table 2.1, the corresponding orthogonal polynomials are Legendre polynomials.

For the general case $\int_a^b f(x)dx$, one can change variables. Let

$$x = \frac{b-a}{2}u + \frac{b+a}{2}. \tag{C.24}$$

Then

$$\int_a^b f(x)dx = \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{b-a}{2}u + \frac{b+a}{2}\right)du. \tag{C.25}$$

Similar ideas can be used for the general case with other Gaussian quadrature rules.

Gauss-Chebyshev Quadrature

Another popular class of Gaussian quadrature rules uses the roots of the Chebyshev polynomials as integration nodes. The degree n Chebyshev polynomial is defined as

$$C_n(x) = cos(ncos^{-1}x). \tag{C.26}$$

They are orthogonal polynomials on [-1, 1], with weight function

$$w(x) = \frac{1}{\sqrt{1-x^2}}. \tag{C.27}$$

The Chebyshev polynomials are used to approximate integrals of the form

$$\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}dx. \tag{C.28}$$

Gauss-Hermite Quadrature

The Hermite polynomials are used to approximate integrals of the form

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx. \tag{C.29}$$

Gauss-Laguerre Quadrature

The Laguerre polynomials approximates integrals of the form

$$\int_0^{\infty} f(x)e^{-x}dx. \tag{C.30}$$

Appendix D

Performance of Unscented Transformation

This appendix aims to show the performance of UT with respect to the Taylor series expansion of the nonlinear transformation. Refer to (Julier and Uhlmann 1996, 2000, 2004) for more details. For the purpose of analysis, it is assumed that the nonlinear function $\mathbf{g}(\mathbf{x})$ can be expresses as multidimensional Taylor series with an arbitrary number of terms. When the number of terms tends to infinity, the residual in the series tends to zero, i.e., the expansion converges to the true value. In fact, the implementation of UT algorithm does not need this restriction, which is set only for the purpose of examining the performance of UT especially when compared with the linearization.

Let the prior variable $\mathbf{x}$ be expressed as the mean $\bar{\mathbf{x}}$ plus a zero mean perturbation $\mathbf{e}$ with covariance $\mathbf{P}_{\mathbf{x}}$, i.e.,

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{e}. \tag{D.1}$$

The Taylor series expansion of the nonlinear function $\mathbf{g}(\mathbf{x})$ is

$$\begin{aligned} \mathbf{g}(\mathbf{x}) &= \mathbf{g}(\bar{\mathbf{x}} + \mathbf{e}) \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \mathbf{D}_{\mathbf{e}}\mathbf{g} + \frac{\mathbf{D}_{\mathbf{e}}^2\mathbf{g}}{2!} + \frac{\mathbf{D}_{\mathbf{e}}^3\mathbf{g}}{3!} + \frac{\mathbf{D}_{\mathbf{e}}^4\mathbf{g}}{4!} + \cdots, \end{aligned} \tag{D.2}$$

where $\mathbf{D}_{\mathbf{e}}\mathbf{g}$ operator evaluates the total differential of $\mathbf{g}(\cdot)$ when perturbed around a nominal value $\bar{\mathbf{x}}$ by $\mathbf{e}$. There are two ways to arrange the operator $\mathbf{D}_{\mathbf{e}}\mathbf{g}$. First, it can be expressed as $\mathcal{J}_g \boldsymbol{e}$, where $\mathcal{J}_g$ is the Jacobian matrix of $\mathbf{g}(\cdot)$ evaluated at $\bar{\mathbf{x}}$. Second, it can be written as the scalar operator as

$$\mathbf{D}_{\mathbf{e}}\mathbf{g} = \left(\sum_{j=1}^{n} e_j \frac{\partial}{\partial x_j}\right) \mathbf{g}(\mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}} . \tag{D.3}$$

The *i*th term in the Taylor series for $\mathbf{g}(\cdot)$ is given by

$$\frac{\mathbf{D}_{\mathbf{e}}^{i}\mathbf{g}}{i!} = \frac{1}{i!}\left(\sum_{j=1}^{n} e_j \frac{\partial}{\partial x_j}\right)^{i} \mathbf{g}(\mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}} \tag{D.4}$$

where e_j is the *j*th component of **e**. Therefore, the *i*th term in the Taylor series is a sum of *i*th-order polynomials, each of which is the *i*th-order product of the components of **e** with the coefficient given by an *i*th-order partial derivatives of $\mathbf{g}(\cdot)$ with respect to **x** and evaluated at $\mathbf{x} = \bar{\mathbf{x}}$.

Hence, the mean $\bar{\mathbf{y}}$ of the posterior variable $\mathbf{y} = \mathbf{g}(\mathbf{x})$ can be evaluated as

$$\begin{aligned}\bar{\mathbf{y}} &= \mathrm{E}[\mathbf{g}(\bar{\mathbf{x}} + \mathbf{e})] \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \mathrm{E}\left[\mathbf{D}_{\mathbf{e}}\mathbf{g} + \frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}}{2!} + \frac{\mathbf{D}_{\mathbf{e}}^{3}\mathbf{g}}{3!} + \frac{\mathbf{D}_{\mathbf{e}}^{4}\mathbf{g}}{4!} + \cdots\right].\end{aligned} \tag{D.5}$$

The term $\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{i}\mathbf{g}}{i!}\right]$ is expanded as

$$\begin{aligned}\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{i}\mathbf{g}}{i!}\right] &= \mathrm{E}\left[\frac{1}{i!}\left(\sum_{j=1}^{n} e_j \frac{\partial}{\partial x_j}\right)^{i}\right] \mathbf{g}(\mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}} \\ &= \frac{1}{i!}\left(m_{i1}\frac{\partial^{i}\mathbf{g}}{\partial {x_1}^{i}} + m_{i2}\frac{\partial^{i}\mathbf{g}}{\partial {x_1}^{i-1}\partial x_2} + \cdots\right).\end{aligned} \tag{D.6}$$

From above discussions, m_{ij} is the expectation of the *i*th-order product of the components of **e,** i.e., the *i*th-order moment of **e**. Therefore, if the mean is correctly estimated to the *i*th-order, both the derivatives of the function $\mathbf{g}(\cdot)$ and the moments of **e** must be known up to the *i*th order.

Let's turn to the covariance $\mathbf{P}_{\boldsymbol{y}}$ of the posterior variable **y**. $\mathbf{P}_{\boldsymbol{y}}$ is calculated as

$$\mathbf{P}_{\boldsymbol{y}} = \mathrm{E}[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^{\mathrm{T}}]. \tag{D.7}$$

From expansions for **y** and $\bar{\mathbf{y}}$ in (D.2) and (D.5)

$$\mathbf{y} - \bar{\mathbf{y}} = \mathbf{D}_{\mathbf{e}}\mathbf{g} + \frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}}{2!} + \frac{\mathbf{D}_{\mathbf{e}}^{3}\mathbf{g}}{3!} + \frac{\mathbf{D}_{\mathbf{e}}^{4}\mathbf{g}}{4!} + \cdots - \mathrm{E}\left[\mathbf{D}_{\mathbf{e}}\mathbf{g} + \frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}}{2!} + \frac{\mathbf{D}_{\mathbf{e}}^{3}\mathbf{g}}{3!} + \frac{\mathbf{D}_{\mathbf{e}}^{4}\mathbf{g}}{4!} + \cdots\right]. \tag{D.8}$$

Substituting (D.8) in (D.7), taking outer products and expectations and exploiting the symmetry of $\mathbf{e}$ which makes the odd terms all evaluate to zero, the covariance $\mathbf{P}_{\boldsymbol{y}}$ becomes

$$\begin{aligned}\mathbf{P}_{\boldsymbol{y}} = {} & \mathrm{E}[\mathbf{D}_{\mathbf{e}}\mathbf{g}(\mathbf{D}_{\mathbf{e}}\mathbf{g})^{\mathrm{T}}] \\ & + \mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}\mathbf{g}(\mathbf{D}_{\mathbf{e}}^{3}\mathbf{g})^{\mathrm{T}}}{3!} + \frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}(\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g})^{\mathrm{T}}}{2!\times 2!} + \frac{\mathbf{D}_{\mathbf{e}}^{3}\mathbf{g}(\mathbf{D}_{\mathbf{e}}\mathbf{g})^{\mathrm{T}}}{3!}\right] \\ & - \mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}}{2!}\right]\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}}{2!}\right]^{\mathrm{T}} + \cdots . \end{aligned} \tag{D.9}$$

The first term $\mathrm{E}[\mathbf{D}_{\mathbf{e}}\mathbf{g}(\mathbf{D}_{\mathbf{e}}\mathbf{g})^{\mathrm{T}}]$ in (D.9) can be written as

$$\mathrm{E}[\mathbf{D}_{\mathbf{e}}\mathbf{g}(\mathbf{D}_{\mathbf{e}}\mathbf{g})^{\mathrm{T}}] = \mathrm{E}\left[\mathcal{J}_{g}\boldsymbol{e}\left(\mathcal{J}_{g}\boldsymbol{e}\right)^{\mathrm{T}}\right] = \mathcal{J}_{g}\mathbf{P}_{\mathbf{x}}\left(\mathcal{J}_{g}\right)^{\mathrm{T}}. \tag{D.10}$$

According to (D.9) and (D.10), the *i*th-order term in the covariance series is calculated correctly only if both the derivatives of the function $\mathbf{g}(\cdot)$ and the moments of $\mathbf{e}$ are known up to the 2*i*th order.

From the analysis above, if the derivatives and the moments are known to a given order, the order of the accuracy of the mean estimate is higher than that of the covariance estimate.

Let's go back to (D.5), the odd terms $\mathrm{E}[\mathbf{D}_{\mathbf{e}}\mathbf{g}]$ and $\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{3}\mathbf{g}}{3!}\right]$ are zero and the term $\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}}{2!}\right]$ can be written as

$$\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{2}\mathbf{g}}{2!}\right] = \mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}(\mathbf{D}_{\mathbf{e}}\mathbf{g})}{2!}\right] = \mathrm{E}\left[\left(\frac{\mathbf{e}^{\mathrm{T}}\nabla\mathbf{e}^{\mathrm{T}}\nabla}{2!}\right)\mathbf{g}\right] = \mathrm{E}\left[\left(\frac{\nabla^{\mathrm{T}}\mathbf{e}\mathbf{e}^{\mathrm{T}}\nabla}{2!}\right)\mathbf{g}\right] = \left(\frac{\nabla^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\nabla}{2!}\right)\mathbf{g}. \tag{D.11}$$

The expansion for the mean becomes

$$\begin{aligned}\bar{\mathbf{y}} &= \mathrm{E}[\mathbf{g}(\bar{\mathbf{x}} + \mathbf{e})] \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \left(\frac{\nabla^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\nabla}{2!}\right)\mathbf{g} + \mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^{4}\mathbf{g}}{4!} + \cdots\right]. \end{aligned} \tag{D.12}$$

Remark: Linearization truncates the Taylor series expansion at the first order and estimates the mean and covariance as

$$\bar{\mathbf{y}}_{\boldsymbol{lin}} = \mathbf{g}(\bar{\mathbf{x}}), \tag{D.13}$$

and

$$\mathbf{P}_{\boldsymbol{y_{lin}}} = \mathcal{J}_g \mathbf{P}_{\mathbf{x}} \left(\mathcal{J}_g\right)^{\mathrm{T}}, \tag{D.14}$$

respectively.

The estimate in (D.13) is accurate only if the expected values of the second and higher order terms in the series are zero. For a linear system, it is always true. However, for a general nonlinear system, the condition does not hold. Therefore the errors are introduced at the second order. Similarly, errors for the covariance estimate in (D.14) are introduced at the fourth order.

1. Performance in Predicting the Mean of a Continuous Function

Consider the sigma points with weights provided in (3.6), the Taylor series expansion for each transformed point is

$$\begin{aligned} \mathbf{g}(\boldsymbol{\chi}_i) &= \mathbf{g}(\bar{\mathbf{x}} + \boldsymbol{\sigma}_i) \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g} + \frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}}{3!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots, \end{aligned} \tag{D.15}$$

where $\boldsymbol{\sigma}_i = \boldsymbol{\chi}_i - \bar{\mathbf{x}}$**.** According to (3.9), the estimate for the mean $\bar{\mathbf{y}}$ is

$$\begin{aligned} \bar{\mathbf{y}} &= \frac{\kappa}{n+\kappa}\mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}\left[\mathbf{g}(\bar{\mathbf{x}}) + \mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g} + \frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}}{3!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots\right] \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \tfrac{1}{2(n+\kappa)}\textstyle\sum_{i=1}^{2n}\left[\mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g} + \frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}}{3!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots\right]. \end{aligned} \tag{D.16}$$

Comparing (D.16) with the true series in (D.5), it is seen that different values will be obtained only if the moments of $\mathbf{e}$ and $\boldsymbol{\sigma}_i$ are different. Here moment is a generalized

conception, which means the true moment only if the sigma points are for a probability function, i.e., $\kappa \geq 0$. If $\kappa < 0$, it is actually the weighted average of components raised to a particular power. As is known from the sigma points, the distribution of $\boldsymbol{\sigma}_i$ is symmetric. So the odd terms are summed to zero. Recalling that $\boldsymbol{\sigma}_i$ are columns of the matrix square root $\sqrt{(n+\kappa)\mathbf{P_x}}$, and the second order even terms can be written as

$$\frac{\mathbf{D}_{\boldsymbol{\sigma}_i}^2 \mathbf{g}}{2!} = \left(\frac{\nabla^{\mathrm{T}} \boldsymbol{\sigma}_i (\boldsymbol{\sigma}_i)^{\mathrm{T}} \nabla}{2!}\right) \mathbf{g}, \tag{D.17}$$

Then the estimation of the mean becomes

$$\bar{\mathbf{y}} = \mathbf{g}(\bar{\mathbf{x}}) + \left(\frac{\nabla^{\mathrm{T}} \mathbf{P_x} \nabla}{2!}\right) \mathbf{g} + \frac{1}{2(n+\kappa)} \sum_{i=1}^{2n} \left[\frac{\mathbf{D}_{\boldsymbol{\sigma}_i}^4 \mathbf{g}}{4!} + \cdots\right]. \tag{D.18}$$

Comparing the estimation in (D.18) with the theoretical value in (D.12), the estimation agrees with the true mean up to the third order and the errors are introduced in the fourth and higher order terms. One cannot say the estimation is more accurate than the linearization before examining the higher order terms in the series. Now the behavior of the higher order terms will be considered.

In order to examine the high order errors, the random variable **e** is decoupled in terms of an uncorrelated random variable $\mathbf{e}'$ with covariance **I (**where **I** is the identity matrix**).** The decoupling process is achieved by a linear transformation

$$\mathbf{e} = \mathbf{A}\mathbf{e}', \tag{D.19}$$

where $\mathbf{A} = \left(a_{ij}\right)_{n \times n}$ is a matrix square root of $\mathbf{P_x}$. Then the D operator in the series can be expressed as

$$\mathbf{D_e} = \sum_{i=1}^{n} e_i' \left(\sum_{j=1}^{n} a_{ij} \frac{\partial}{\partial x_j}\right). \tag{D.20}$$

Similarly, $\boldsymbol{\sigma}_i$ can be decoupled in terms of $\boldsymbol{\sigma}_i'$ where $\boldsymbol{\sigma}_i'$ is an uncorrelated random variable with covariance **I**. $\boldsymbol{\sigma}_i'$ is related to $\boldsymbol{\sigma}_i$ by

$$\boldsymbol{\sigma}_i = \mathbf{A}\boldsymbol{\sigma}_i'. \tag{D.21}$$

Again, A is any matrix square root of $\mathbf{P_x}$. Rather than handling a correlated random vector, the only thing needed now is to consider the uncorrelated random vectors $\mathbf{e}'$ and $\boldsymbol{\sigma}_i'$.

For Gaussian distribution, it can be verified that the fourth order moments (or kurtosis) are given as

$$\begin{aligned} \mathrm{E}\left[e_i'^4\right] &= 3, \forall i \\ \mathrm{E}\left[e_i'^2 e_j'^2\right] &= 1, \forall i \neq j \end{aligned} \tag{D.22}$$

and all other fourth order moments are zero. For the sigma points, the kurtosis of the *j*th components are

$$\frac{1}{2(n+\kappa)} \sum_{i=1}^{2n} \sigma_{ij}'^4 = n + \kappa, \forall j. \tag{D.23}$$

And all other fourth order products are zero.

The analysis above shows us the effect of parameter κ. Although the first three moments does not change with the selection of κ, it does affect the fourth and higher order moments of $\boldsymbol{\sigma}_i$. If information about the predicted distribution is known, the information can be incorporated to choose proper value for parameter κ to minimize the estimation error. However, if no information about the higher order terms of $\mathbf{g}(\cdot)$ is known, the choice of κ is usually made to ensure that the errors are smaller than those committed by linearization (Julier and Uhlmann 1996).

Comparing the kurtosis for the true distribution in (D.22) and that for the sigma points in (D.23), two differences can be seen. First, the kurtosis of a single state for the Gaussian distribution is 3 but $n + \kappa$ for the sigma point distribution. Second, the cross kurtosis is zero for sigma point distribution (actually all higher order moments are zero)

but nonzero for the Gaussian distribution. Therefore, except for the one dimensional case in which there is indeed no cross kurtosis, the "shape" of the moments are different. If the value of κ is chosen to satisfy the condition $n + \kappa = 3$, then the kurtosis of the single states will be the same for both Gaussian distribution and the sigma point distribution. There will be no difference for the fourth order moments for single dimensional state space, which indicates that the errors are introduced in the sixth and higher order moments. For multidimensional state space, fourth order errors are introduced by the cross kurtosis terms. As seen from (D.13), the fourth order moments are assumed to be zero in linearization. Therefore, linearization introduces bigger absolute errors in the fourth order moments than UT does.

Let's consider the sixth and higher order moments. The values of the higher order moments for a Gaussian distribution grow factorially while the higher order moments for the sigma point distribution grow geometrically with common factor $n + \kappa$. Therefore, for any choice of parameter κ, it's possible to select a large order such that the moments of the true series exceed those for the sigma point distribution. When $n + \kappa = 3$, the moments coincide at the fourth order. For higher orders, the moments of Gaussian distribution are larger in magnitude than those of sigma point distribution. As is known that all the higher order terms are enforced to be zero in linearization which means more errors are introduced by linearization.

It can be seen that when $n + \kappa$ tends to zero, the kurtosis and higher order moments for the sigma points will tend to zero. As a result, the mean estimate will converge to

$$\lim_{n+\kappa \to 0} \bar{\mathbf{y}} = \mathbf{g}(\bar{\mathbf{x}}) + \left(\frac{\nabla^{\mathbf{T}} \mathbf{P}_{\mathbf{x}} \nabla}{2!}\right) \mathbf{g}\,, \tag{D.24}$$

which is equivalent to the well-known truncated second order filter. However, one thing needs to be pointed out is that no Jacobian and Hessian calculations are needed in this UT approach.

From above, a simple conclusion that for all continuous nonlinear transformations, UT method can give more accurate estimates than linearization does, can be conducted. The performance is determined by the choice of parameter κ as it scales the fourth and higher order moments of the distribution. When the information about the true conditional mean (e.g., through Monte Carlo simulation) is known, the value of κ can be chosen to minimize the error. For most filtering applications, the first two terms (the first and second term) are dominant and κ has a minimal effect on estimation performance.

2. Performance in Predicting the Covariance of a Continuous Function

According to (D.15) and (D.16), the estimation

$$\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}} = \mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g} + \frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}}{3!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots$$
$$- \frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}\left[\mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g} + \frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}}{3!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots\right], i = 0,1,\ldots,2n. \quad \text{(D.25)}$$

As discussed above, $\boldsymbol{\sigma}_i, i = 1,2,\ldots,2n$ are symmetric, the odd terms in the summation of (D.16) are zero. Besides,$\boldsymbol{\sigma}_0 = \mathbf{0}$, then

$$\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}} = \mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g} + \frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}}{3!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots$$
$$- \frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}\left[\frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots\right], i = 1,\ldots,2n, \quad \text{(D.26)}$$

$$\boldsymbol{\mathcal{Y}}_0 - \bar{\mathbf{y}} = -\frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}\left[\frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}}{2!} + \frac{\mathbf{D}^4_{\boldsymbol{\sigma}_i}\mathbf{g}}{4!} + \cdots\right]. \quad \text{(D.27)}$$

And

$$\frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}\mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g}\left(\mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g}\right)^{\mathrm{T}}=\frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}\mathcal{J}_g\boldsymbol{\sigma}_i(\boldsymbol{\sigma}_i)^{\mathrm{T}}\left(\mathcal{J}_g\right)^{\mathrm{T}}=\mathcal{J}_g\mathbf{P}_{\mathbf{x}}\left(\mathcal{J}_g\right)^{\mathrm{T}}. \quad \text{(D.28)}$$

According to (3.9) and (D.16), the estimate for the covariance $\mathbf{P}_{\mathbf{y}}$ is

$$\begin{aligned}\mathbf{P}_{\mathbf{y}}&=\frac{\kappa}{n+\kappa}(\boldsymbol{\mathcal{Y}}_0-\bar{\mathbf{y}})(\boldsymbol{\mathcal{Y}}_0-\bar{\mathbf{y}})^T+\frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}(\boldsymbol{\mathcal{Y}}_i-\bar{\mathbf{y}})(\boldsymbol{\mathcal{Y}}_i-\bar{\mathbf{y}})^T\\&=\mathcal{J}_g\mathbf{P}_{\mathbf{x}}\left(\mathcal{J}_g\right)^{\mathrm{T}}+\frac{1}{2(n+\kappa)}\sum_{i=1}^{2n}\left[\frac{\mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g}\left(\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}\right)^{\mathrm{T}}}{3!}+\frac{\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}\left(\mathbf{D}^2_{\boldsymbol{\sigma}_i}\mathbf{g}\right)^{\mathrm{T}}}{2!\times 2!}+\frac{\mathbf{D}^3_{\boldsymbol{\sigma}_i}\mathbf{g}\left(\mathbf{D}_{\boldsymbol{\sigma}_i}\mathbf{g}\right)^{\mathrm{T}}}{3!}\right]\\&\quad-\left[\left(\frac{\nabla^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\nabla}{2!}\right)\mathbf{g}\right]\left[\left(\frac{\nabla^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\nabla}{2!}\right)\mathbf{g}\right]^{\mathrm{T}}+\cdots. \quad \text{(D.29)}\end{aligned}$$

Comparing (D.29) with the theoretical value (D.9), and according to (D.10) and (D.11), the estimation in (D.29) agrees with the series up to the second order terms. As discussed in the mean estimation, the kurtoses of the true distribution do not agree with those of the sigma point distribution, errors are introduced in the fourth and higher order terms. And similarly, the absolute error in the covariance estimation obtained by UT is smaller than that of linearization. However, if $\kappa < 0$, the approximated covariance is not ensured to be positive semidefinite.

The estimation of the covariance for a continuous function can be simply concluded. Though both UT and linearization approaches can estimate the covariance up to the second order, the absolute errors in the fourth and higher order terms for UT are smaller. However, UT cannot ensure the positive semi-definiteness of the covariance. An alternative method for covariance calculation has been proposed which ensures the semi-definiteness of the covariance in (Julier and Uhlmann 1996) and is still more accurate than linearization.

3. Estimation of the Mean and Covariance for a Discontinuous Function

The analysis ahead is valid only if the transformation function is continuous across all possible values of the state estimates. However, in many practical applications, the model functions are discontinuous.

Let's consider the linearization approach first. Since the transformation function has a finite expected value, all discontinuities must include finite discontinuities in behavior of the function. Besides, if the transformation is piecewise by a number of continuous functions, each of which has its own Taylor series. The linearization exhibits two types of behaviors. If the estimate does not lie at a discontinuity, the estimates for the mean and covariance do not reflect the discontinuity at all. However, if it lies at a discontinuity, there are difficulties in applying linearization. Specifically, if the function is non-differentiable at that point, then it's not possible to use linearization approach to predict the covariance.

Contrast to linearization, UT approach does not require the condition that the transformation should be differentiable. However, the performance will be worse if the function is discontinuous. If the discontinuity does not occur within the covariance ellipse formed by the sigma points then the estimates of the mean and covariance do not acknowledge its existence (Julier and Uhlmann 1996). However, if it lies outside, it is unlikely to affect a significant proportion of the distribution. If it lies at a sigma point, generally it is not possible to make any comments about the performance. If a discontinuity lies within the sigma points, generally the odd terms in the summations of the Taylor series do not cancel out and first order errors will be introduced into the mean and covariance estimation. Although, the sigma points can represent the first moment correctly, it is scaled by $1/\sqrt{n+\kappa}$. The error can be reduced by reducing the

value of κ at the cost of distorting the higher order terms in the series. Further, when $n + \kappa$ tends to zero, the sigma points converge to one another and might miss the discontinuity. The error is significant only if the distribution is largely affected by the discontinuity.

In general, let's conclude UT with comparison to linearization as follows. When the transformation is continuous, the estimation form UT is accurate to the third order and the errors are introduced at the fourth order. Linearization can only give estimate which is accurate to the second order. In practice, lower terms are significant, so UT is more accurate. When the function is discontinuous, linearization only incorporates it when the discontinuity lies on the current state estimate. If the function is not differentiable at that point, the covariance cannot be calculated. UT uses a distribution of points and captures the discontinuity if the discontinuity affects a large portion of the distribution.

Appendix E

Performance of Scaled Unscented Transformation

1. The Auxillary Random Variable

In this part, the aim is to show the Taylor series expansion for the mean and covariance of the auxillary random variable **z** in (3.11) agree with the expansion for the mean and covariance of random variable **y** in (3.1) up to the second order.

The Taylor series expansion of the auxillary transformation in (3.11) about $\bar{\mathbf{x}}$ is

$$\begin{aligned}\mathbf{z} &= \mathbf{h}(\bar{\mathbf{x}} + \mathbf{e}) \\ &= \mathbf{h}(\bar{\mathbf{x}}) + \mathbf{D_e h} + \frac{\mathbf{D_e^2 h}}{2!} + \frac{\mathbf{D_e^3 h}}{3!} + \frac{\mathbf{D_e^4 h}}{4!} + \cdots \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{\alpha^2}\left(\alpha \mathbf{D_e g} + \frac{\alpha^2 \mathbf{D_e^2 g}}{2!} + \frac{\alpha^3 \mathbf{D_e^3 g}}{3!} + \frac{\alpha^4 \mathbf{D_e^4 g}}{4!} + \cdots\right) \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{\alpha}\mathbf{D_e g} + \frac{\mathbf{D_e^2 g}}{2!} + \frac{\alpha \mathbf{D_e^3 g}}{3!} + \frac{\alpha^2 \mathbf{D_e^4 g}}{4!} + \cdots. \end{aligned} \tag{E.1}$$

Taking expectations, the estimate for the mean is

$$\bar{\mathbf{z}} = \mathbf{g}(\bar{\mathbf{x}}) + \mathrm{E}\left(\frac{1}{\alpha}\mathbf{D_e g} + \frac{\mathbf{D_e^2 g}}{2!} + \frac{\alpha \mathbf{D_e^3 g}}{3!} + \frac{\alpha^2 \mathbf{D_e^4 g}}{4!} + \cdots\right). \tag{E.2}$$

The odd terms $\mathrm{E}[\mathbf{D_e g}]$ and $\mathrm{E}\left[\frac{\mathbf{D_e^3 g}}{3!}\right]$ are zero due to the symmetry of $\mathbf{e}$, and the term $\mathrm{E}\left[\frac{\mathbf{D_e^2 g}}{2!}\right]$ is equal to $\left(\frac{\nabla^{\mathrm{T}}\mathbf{P_x}\nabla}{2!}\right)\mathbf{g}$, the estimation (E.2) becomes

$$\bar{\mathbf{z}} = \mathbf{g}(\bar{\mathbf{x}}) + \left(\frac{\nabla^{\mathrm{T}}\mathbf{P_x}\nabla}{2!}\right)\mathbf{g} + \mathrm{E}\left(\frac{\alpha^2 \mathbf{D_e^4 g}}{4!} + \cdots\right). \tag{E.3}$$

From (E.3) and (D.12), it is clear that $\bar{\mathbf{z}}$**,** the expansion of the mean for the auxillary variable **z** agrees with that of **y (**i.e, $\bar{\mathbf{y}}$) up to the second order. The parameter α only affects the higher orders and its values can be chosen so that the scaling effects in the higher order terms are minimized.

Let's consider the covariance now. Suppose the estimation of $\mathbf{z}$ is calculated as

$$\mathbf{P}_{\boldsymbol{z}} = \alpha^2 \mathrm{E}[(\mathbf{z}-\bar{\mathbf{z}})(\mathbf{z}-\bar{\mathbf{z}})^{\mathrm{T}}]. \tag{E.4}$$

From expansions for $\mathbf{z}$ and $\bar{\mathbf{z}}$ in (E.1) and (E.2),

$$\mathbf{z}-\bar{\mathbf{z}} = \left(\frac{1}{\alpha}\mathbf{D}_{\mathbf{e}}\mathbf{g} + \frac{\mathbf{D}_{\mathbf{e}}^2\mathbf{g}}{2!} + \frac{\alpha\mathbf{D}_{\mathbf{e}}^3\mathbf{g}}{3!} + \frac{\alpha^2\mathbf{D}_{\mathbf{e}}^4\mathbf{g}}{4!} + \cdots\right)$$

$$-\mathrm{E}\left(\frac{1}{\alpha}\mathbf{D}_{\mathbf{e}}\mathbf{g} + \frac{\mathbf{D}_{\mathbf{e}}^2\mathbf{g}}{2!} + \frac{\alpha\mathbf{D}_{\mathbf{e}}^3\mathbf{g}}{3!} + \frac{\alpha^2\mathbf{D}_{\mathbf{e}}^4\mathbf{g}}{4!} + \cdots\right). \tag{E.5}$$

Taking outer products and expectations and exploiting the symmetry of $\mathbf{e}$ which makes the odd terms all evaluate to zero, the covariance $\mathbf{P}_{\boldsymbol{z}}$ will be expressed as

$$\mathbf{P}_{\boldsymbol{z}} = \mathrm{E}[\mathbf{D}_{\mathbf{e}}\mathbf{g}(\mathbf{D}_{\mathbf{e}}\mathbf{g})^{\mathrm{T}}]$$

$$+\alpha^2\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}\mathbf{g}(\mathbf{D}_{\mathbf{e}}^3\mathbf{g})^{\mathrm{T}}}{3!} + \frac{\mathbf{D}_{\mathbf{e}}^2\mathbf{g}(\mathbf{D}_{\mathbf{e}}^2\mathbf{g})^{\mathrm{T}}}{2!\times 2!} + \frac{\mathbf{D}_{\mathbf{e}}^3\mathbf{g}(\mathbf{D}_{\mathbf{e}}\mathbf{g})^{\mathrm{T}}}{3!}\right]$$

$$-\alpha^2\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^2\mathbf{g}}{2!}\right]\mathrm{E}\left[\frac{\mathbf{D}_{\mathbf{e}}^2\mathbf{g}}{2!}\right]^{\mathrm{T}} + \cdots. \tag{E.6}$$

From (D.9) and (E.6), the expansion of $\mathbf{P}_{\boldsymbol{z}}$ agrees with the expansion of $\mathbf{P}_{\boldsymbol{y}}$ up to the second order. The higher order terms are scaled with parameter α.

2. The Scaled Unscented Transformation

This part of the appendix will show the estimates of the mean and covariance from the SUT scheme agree with those from the auxillary form of the unscented transform for any sigma point distribution.

2.1 The Weight Selection

First, let's consider how the weights associated with the scaled sigma points are allocated as (3.14). The weights $\mathrm{W}_i', i = 0,1,\ldots,p$ are chosen to satisfy the following conditions:

$$\sum_{i=0}^{p} \mathrm{W}_i' = 1,$$

$$\sum_{i=0}^{p} \mathrm{W}_i' \boldsymbol{\chi}_i' = \bar{\mathbf{x}},$$

$$\sum_{i=0}^{p} \mathrm{W}_i' (\boldsymbol{\chi}_i' - \bar{\mathbf{x}})(\boldsymbol{\chi}_i' - \bar{\mathbf{x}})^{\mathrm{T}} = \mathbf{P}_{\mathbf{x}}. \quad \text{(E.7)}$$

From (3.10),

$$\boldsymbol{\chi}_i' - \boldsymbol{\chi}_0 = \alpha(\boldsymbol{\chi}_i - \boldsymbol{\chi}_0). \quad \text{(E.8)}$$

Comparing the expressions of $\mathbf{P}_{\mathbf{x}}$ in (E.7) and (3.3), substituting (E.8) in (E.7) and using the fact that $\boldsymbol{\chi}_0 = \bar{\mathbf{x}}$**,** it is easily to set

$$\mathrm{W}_i' = \frac{1}{\alpha^2} W_i, i = 1, \dots, p. \quad \text{(E.9)}$$

Since $\boldsymbol{\chi}_0 - \bar{\mathbf{x}} = \mathbf{0}$, let's examine the weight on the zeroth sigma point from

$$\begin{aligned} \mathrm{W}_0' &= 1 - \sum_{i=1}^{p} \mathrm{W}_i' \\ &= 1 - \frac{1}{\alpha^2} \sum_{i=1}^{p} W_i \\ &= 1 - \frac{1}{\alpha^2}(1 - W_0) \\ &= \frac{\mathrm{W}_0}{\alpha^2} + (1 - \frac{1}{\alpha^2}). \end{aligned} \quad \text{(E.10)}$$

The weight values (E.9) and (E.10) are exactly those given in (3.14) for SUT.

2.2 The Estimation of the Mean

From (3.12) and using the fact $\boldsymbol{\chi}_0' = \boldsymbol{\chi}_0 = \bar{\mathbf{x}}$,

$$\begin{aligned} \boldsymbol{z}_i &= \frac{\mathbf{g}[\bar{\mathbf{x}} + \alpha(\boldsymbol{\chi}_i - \bar{\mathbf{x}})] - \mathbf{g}(\bar{\mathbf{x}})}{\alpha^2} + \mathbf{g}(\bar{\mathbf{x}}) \\ &= \frac{\mathbf{g}[\boldsymbol{\chi}_i'] - \mathbf{g}(\boldsymbol{\chi}_0')}{\alpha^2} + \mathbf{g}(\boldsymbol{\chi}_0') \\ &= \frac{\mathbf{g}[\boldsymbol{\chi}_i']}{\alpha^2} + \left(1 - \frac{1}{\alpha^2}\right) \mathbf{g}(\boldsymbol{\chi}_0') \\ &= \frac{\boldsymbol{y}_i'}{\alpha^2} + \left(1 - \frac{1}{\alpha^2}\right) \boldsymbol{y}_0'. \end{aligned} \quad \text{(E.11)}$$

After the substitution of (E.11) in (3.13) and using the fact in (3.2),

$$\bar{\mathbf{z}} = \sum_{i=0}^{p} \mathrm{W}_i \, \boldsymbol{z}_i$$

$$= \sum_{i=0}^{p} \mathrm{W}_i \left[\frac{\boldsymbol{y}_i'}{\alpha^2} + \left(1 - \frac{1}{\alpha^2}\right) \boldsymbol{y}_0' \right]$$

$$= \sum_{i=1}^{p} \frac{\mathrm{W}_i}{\alpha^2} \boldsymbol{y}_i' + \left[\frac{\mathrm{W}_0}{\alpha^2} + (1 - \tfrac{1}{\alpha^2}) \right] \boldsymbol{y}_0'. \qquad \text{(E.12)}$$

After substitution of (E.9), (E.10) in (E.12), $\bar{\mathbf{z}}$ becomes

$$\bar{\mathbf{z}} = \sum_{i=0}^{p} \mathrm{W}_i' \, \boldsymbol{y}_i'$$

$$= \bar{\mathbf{y}}'. \qquad \text{(E.13)}$$

Now it has been shown that the estimation of the mean from SUT is the same as that from the auxillary of the UT. It holds for the covariance also, i.e., $\mathbf{P}_{\mathbf{z}} = \mathbf{P}_{\boldsymbol{y}}'$, which will be shown next.

2.3 The Estimation of the Covariance

From (E.11) and (E.13),

$$\boldsymbol{z}_i - \bar{\mathbf{z}} = \frac{\boldsymbol{y}_i'}{\alpha^2} + \left(1 - \frac{1}{\alpha^2}\right) \boldsymbol{y}_0' - \bar{\mathbf{y}}'$$

$$= \frac{\boldsymbol{y}_i'}{\alpha^2} + \left(1 - \tfrac{1}{\alpha^2}\right) \boldsymbol{y}_0' - \left[\tfrac{1}{\alpha^2} + \left(1 - \tfrac{1}{\alpha^2}\right) \right] \bar{\mathbf{y}}'$$

$$= \tfrac{1}{\alpha^2} (\boldsymbol{y}_i' - \bar{\mathbf{y}}') + \left(1 - \tfrac{1}{\alpha^2}\right) (\boldsymbol{y}_0' - \bar{\mathbf{y}}'). \qquad \text{(E.14)}$$

Then from (3.13),

$$\mathbf{P}_{\mathbf{z}} = \alpha^2 \sum_{i=0}^{p} \mathrm{W}_i \, (\boldsymbol{z}_i - \bar{\mathbf{z}})(\boldsymbol{z}_i - \bar{\mathbf{z}})^{\mathrm{T}}$$

$$= \alpha^2 \sum_{i=0}^{p} \mathrm{W}_i \left[\frac{1}{\alpha^2} (\boldsymbol{y}_i' - \bar{\mathbf{y}}') + \left(1 - \frac{1}{\alpha^2}\right) (\boldsymbol{y}_0' - \bar{\mathbf{y}}') \right]$$

$$\times \left[\frac{1}{\alpha^2}(\boldsymbol{y}_i' - \bar{\mathbf{y}}') + \left(1 - \frac{1}{\alpha^2}\right)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')\right]^{\mathrm{T}}$$

$$= \frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i \begin{Bmatrix} (\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} + (\alpha^2 - 1)(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}} \\ +(\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} + (\alpha^2 - 1)^2(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}} \end{Bmatrix}. \quad \text{(E.15)}$$

From (E.9) and (E.10),

$$\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \sum_{i=0}^{p} \mathrm{W}_i'(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} - \left(1 - \frac{1}{\alpha^2}\right)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}. \quad \text{(E.16)}$$

Together with (3.2),

$$\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\alpha^2 - 1)(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \left[\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\alpha^2 - 1)(\boldsymbol{y}_i' - \bar{\mathbf{y}}')\right](\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \left[\frac{(\alpha^2-1)}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\boldsymbol{y}_i' - \bar{\mathbf{y}}')\right](\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= (\alpha^2 - 1)\left\{\left[\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i\boldsymbol{y}_i'\right] - \left[\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i\bar{\mathbf{y}}'\right]\right\}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= (\alpha^2 - 1)\left\{\sum_{i=0}^{p} \mathrm{W}_i'\boldsymbol{y}_i' - (1 - \frac{1}{\alpha^2})\boldsymbol{y}_0' - \frac{1}{\alpha^2}\bar{\mathbf{y}}'\right\}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= (\alpha^2 - 1)\left[\bar{\mathbf{y}}' - (1 - \frac{1}{\alpha^2})\boldsymbol{y}_0' - \frac{1}{\alpha^2}\bar{\mathbf{y}}'\right](\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= (\alpha^2 - 1)\left[\left(1 - \frac{1}{\alpha^2}\right)\bar{\mathbf{y}}' - \left(1 - \frac{1}{\alpha^2}\right)\boldsymbol{y}_0'\right](\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \frac{-(\alpha^2-1)^2}{\alpha^2}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}. \quad \text{(E.17)}$$

Similarly,

$$\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= (\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')\left[\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}}\right]$$

$$= (\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}') \left[\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\boldsymbol{y}_i')^{\mathrm{T}} - \frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\bar{\mathbf{y}}')^{\mathrm{T}}\right]$$

$$= (\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}') \left[\sum_{i=0}^{p} \mathrm{W}_i'(\boldsymbol{y}_i')^{\mathrm{T}} - (1 - \frac{1}{\alpha^2})(\boldsymbol{y}_0')^{\mathrm{T}} - \frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\bar{\mathbf{y}}')^{\mathrm{T}}\right]$$

$$= (\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}') \left[(\bar{\mathbf{y}}')^{\mathrm{T}} - \left(1 - \frac{1}{\alpha^2}\right)(\boldsymbol{y}_0')^{\mathrm{T}} - \frac{1}{\alpha^2}(\bar{\mathbf{y}}')^{\mathrm{T}}\right]$$

$$= (\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}') \left[\left(1 - \frac{1}{\alpha^2}\right)(\bar{\mathbf{y}}')^{\mathrm{T}} - \left(1 - \frac{1}{\alpha^2}\right)(\boldsymbol{y}_0')^{\mathrm{T}}\right]$$

$$= -\frac{(\alpha^2-1)^2}{\alpha^2}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}. \tag{E.18}$$

$$\frac{1}{\alpha^2}\sum_{i=0}^{p} \mathrm{W}_i(\alpha^2 - 1)^2(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \frac{(\alpha^2-1)^2}{\alpha^2}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}. \tag{E.19}$$

Substitute (E.16) – (E.19) in (E.15) and obtain

$$\mathbf{P_z} = \sum_{i=0}^{p} \mathrm{W}_i'(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} - \left(1 - \frac{1}{\alpha^2}\right)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$-\frac{2(\alpha^2-1)^2}{\alpha^2}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}} + \frac{(\alpha^2-1)^2}{\alpha^2}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \sum_{i=0}^{p} \mathrm{W}_i'(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} - \left[1 - \frac{1}{\alpha^2} + \frac{(\alpha^2-1)^2}{\alpha^2}\right](\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \sum_{i=0}^{p} W_i'(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} - \frac{\alpha^2-1+(\alpha^2-1)^2}{\alpha^2}(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \sum_{i=0}^{p} W_i'(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} - (\alpha^2 - 1)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}$$

$$= \sum_{i=0}^{p} W_i'(\boldsymbol{y}_i' - \bar{\mathbf{y}}')(\boldsymbol{y}_i' - \bar{\mathbf{y}}')^{\mathrm{T}} + (1 - \alpha^2)(\boldsymbol{y}_0' - \bar{\mathbf{y}}')(\boldsymbol{y}_0' - \bar{\mathbf{y}}')^{\mathrm{T}}. \tag{E.20}$$

which is the same as the formula for $\mathbf{P}_{\boldsymbol{y}}'$ in (3.15), i.e., $\mathbf{P_z} = \mathbf{P}_{\boldsymbol{y}}'$.

Appendix F

Stochastic Galerkin for Mixed-layer Model

Let $\mathbf{x} = (\theta, h, \sigma, q, \mu)^{\mathrm{T}}$, the aim is to seek a PC expansion of $\mathbf{x}(t, \xi)$ in the following form

$$\mathbf{x}(t, \xi) = \sum_{i=0}^{N} \mathbf{v}_i(t)\phi_i(\xi). \tag{F.1}$$

Here ξ is a scalar random variable with standard Gaussian distribution. $\mathbf{v}_i(t) \epsilon R^{5\times 1}$, $\xi \sim \mathrm{N}(0,1)$, $\phi_i(\xi)$ are normalized Hermite polynomials. For simplicity, $\mathbf{v}_i(t)$ is denoted as $\mathbf{v}_i$ and $\phi_i(\xi)$ is denoted as ϕ_i in the following paragraphs. Denote

$$\mathbf{v}_i = (v_{i1}, v_{i2}, v_{i3}, v_{i4}, v_{i5})^T, \tag{F.2}$$

$$\dot{\mathbf{x}} = \sum_{i=0}^{N} \dot{\mathbf{v}}_\iota \phi_i. \tag{F.3}$$

Let $C_\theta V_s(1+\kappa) = \alpha$, $C_\theta V_s \kappa = \beta$ and $C_q V_s = \gamma$.

Equation 1:

Equation (5.1) can be written as

$$\dot{\theta} h = \alpha(\theta_s - \theta). \tag{F.4}$$

Using PC expansion, the following can be obtained

$$(\sum_{i=0}^{N} \dot{v_{\iota 1}}\phi_i)(\sum_{i=0}^{N} v_{i2}\phi_i) = \alpha(\theta_s - \sum_{i=0}^{N} v_{i1}\phi_i), \tag{F.5}$$

$$\sum_{i,j=0}^{N} \dot{v_{\iota 1}} v_{j2}\phi_i\,\phi_j = \alpha(\theta_s - \sum_{i=0}^{N} v_{i1}\phi_i). \tag{F.6}$$

Taking inner product with both sides by $\phi_k (k = 0,1,\dots,N)$,

$$\sum_{i,j=0}^{N} \dot{v_{\iota 1}} v_{j2} < \phi_i\,\phi_j, \phi_k > = \alpha(\theta_s \langle \phi_k \rangle - \sum_{i=0}^{N} v_{i1} < \phi_i, \phi_k >). \tag{F.7}$$

According to the orthogonality of the Hermite polynomials, the following equation can be obtained

$$\sum_{i,j=0}^{N} \dot{v_{\iota 1}} v_{j2} < \phi_i\,\phi_j, \phi_k >$$

$$= \alpha(\theta_s - v_{k1}), (\text{if } k = 0)$$

$$-\alpha v_{k1}. (\text{if } k = 1,2, \dots, N) \quad \text{(F.8)}$$

Equation 2:

Equation (5.2) can be written as

$$(\dot{h} - w)\sigma = \beta(\theta_s - \theta). \quad \text{(F.9)}$$

Using PC expansion, the following will be obtained

$$(\sum_{i=0}^{N} v_{i2}^{\cdot}\phi_i - w)(\sum_{i=0}^{N} v_{i3}\phi_i) = \beta(\theta_s - \sum_{i=0}^{N} v_{i1}\phi_i), \quad \text{(F.10)}$$

$$\sum_{i,j=0}^{N} v_{i2}^{\cdot} v_{j3}\phi_i\,\phi_j - w\sum_{i=0}^{N} v_{i3}\phi_i = \beta(\theta_s - \sum_{i=0}^{N} v_{i1}\phi_i). \quad \text{(F.11)}$$

Taking inner product with both sides by $\phi_k(k = 0,1, \dots, N)$,

$$\sum_{i,j=0}^{N} v_{i2}^{\cdot} v_{j3} < \phi_i\,\phi_j, \phi_k > -w\sum_{i=0}^{N} v_{i3}\langle\phi_i, \phi_k\rangle = \beta(\theta_s\langle\phi_k\rangle - \sum_{i=0}^{N} v_{i1} < \phi_i, \phi_k >). \quad \text{(F.12)}$$

According to the orthogonality of the Hermite polynomials, the following equation can be obtained

$$\sum_{i,j=0}^{N} v_{i2}^{\cdot} v_{j3} < \phi_i\,\phi_j, \phi_k >$$

$$= \beta(\theta_s - v_{k1}) + wv_{k3}, (\text{if } k = 0)$$

$$-\beta v_{k1} + wv_{k3}. (\text{if } k = 1,2, \dots, N) \quad \text{(F.13)}$$

Equation 3:

Equation (5.3) can be written as

$$\dot{\sigma} = \gamma_\theta \dot{h} - \dot{\theta} - w\gamma_\theta. \quad \text{(F.14)}$$

Using PC expansion, the following will be obtained

$$\sum_{i=0}^{N} v_{i3}^{\cdot}\phi_i = \gamma_\theta \sum_{i=0}^{N} v_{i2}^{\cdot}\phi_i - \sum_{i=0}^{N} v_{i1}^{\cdot}\phi_i - w\gamma_\theta. \quad \text{(F.15)}$$

Taking inner product with both sides by $\phi_k(k = 0,1, \dots, N)$ and using the orthogonality of the Hermite polynomials, the following equation can be obtained

$$\dot{v_{k3}}$$

$$\dot{v_{k3}} = \gamma_\theta \dot{v_{k2}} - \dot{v_{k1}} - w\gamma_\theta, (\text{if } k = 0)$$

$$\gamma_\theta \dot{v_{k2}} - \dot{v_{k1}}. (\text{if } k = 1,2, \dots, N) \quad \text{(F.16)}$$

Equation 4:

Equation (5.4) can be written as

$$\dot{q}h\sigma = \gamma\sigma(q_s - q) + \gamma\kappa\mu(\theta_s - \theta). \quad \text{(F.17)}$$

Using PC expansion,

$$(\sum_{i=0}^{N} \dot{v_{i4}}\phi_i)(\sum_{i=0}^{N} v_{i2}\phi_i)(\sum_{i=0}^{N} v_{i3}\phi_i) = \gamma(\sum_{i=0}^{N} v_{i3}\phi_i)(q_s - \sum_{i=0}^{N} v_{i4}\phi_i) +$$

$$\gamma\kappa(\sum_{i=0}^{N} v_{i5}\phi_i)(\theta_s - \sum_{i=0}^{N} v_{i1}\phi_i), \quad \text{(F.18)}$$

$$\sum_{i,j,m=0}^{N} \dot{v_{i4}} v_{j2} v_{m3} \phi_i \phi_j \phi_m = \gamma q_s(\sum_{i=0}^{N} v_{i3}\phi_i) - \gamma \sum_{i,j=0}^{N} v_{i3} v_{j4} \phi_i \phi_j$$

$$+\gamma\kappa\theta_s(\sum_{i=0}^{N} v_{i5}\phi_i) - \gamma\kappa \sum_{i,j=0}^{N} v_{i1} v_{j5} \phi_i \phi_j. \quad \text{(F.19)}$$

Taking inner product with both sides by $\phi_k (k = 0,1, \dots, N)$,

$$\sum_{i,j,m=0}^{N} \dot{v_{i4}} v_{j2} v_{m3} \langle \phi_i \phi_j \phi_m, \phi_k \rangle = \gamma q_s(\sum_{i=0}^{N} v_{i3}\langle \phi_i, \phi_k \rangle) - \gamma \sum_{i,j=0}^{N} v_{i3} v_{j4} \langle \phi_i \phi_j, \phi_k \rangle$$

$$+\gamma\kappa\theta_s(\sum_{i=0}^{N} v_{i5}\langle \phi_i, \phi_k \rangle) - \gamma\kappa \sum_{i,j=0}^{N} v_{i1} v_{j5} \langle \phi_i \phi_j, \phi_k \rangle. \quad \text{(F.20)}$$

According to the orthogonality of the Hermite polynomials, the following equation can be obtained

$$\sum_{i,j,m=0}^{N} \dot{v_{i4}} v_{j2} v_{m3} \langle \phi_i \phi_j \phi_m, \phi_k \rangle = \gamma q_s v_{k3} - \gamma \sum_{i,j=0}^{N} v_{i3} v_{j4} \langle \phi_i \phi_j, \phi_k \rangle + \gamma\kappa\theta_s v_{k5} -$$

$$\gamma\kappa \sum_{i,j=0}^{N} v_{i1} v_{j5} \langle \phi_i \phi_j, \phi_k \rangle, (k = 0,1, \dots, N) \quad \text{(F.21)}$$

Equation 5:

Equation (5.5) can be written as

$$\dot{\mu} = \gamma_q \dot{h} - \dot{q} - w\gamma_q. \quad \text{(F.22)}$$

Using PC expansion, the following will be obtained

$$\sum_{i=0}^{N} \dot{v_{i5}}\phi_i = \gamma_q \sum_{i=0}^{N} \dot{v_{i2}}\phi_i - \sum_{i=0}^{N} \dot{v_{i4}}\phi_i - w\gamma_q. \tag{F.23}$$

Taking inner product with both sides by $\phi_k (k = 0,1, \dots, N)$ and using the orthogonality of the Hermite polynomials, the following equation can be obtained

$$\begin{aligned} &\dot{v_{k5}} \\ &= \gamma_\theta \dot{v_{k2}} - \dot{v_{k4}} - w\gamma_q, \text{ (if } k = 0) \\ &\gamma_\theta \dot{v_{k2}} - \dot{v_{k4}}, \text{ (if } k = 1,2, \dots, N) \end{aligned} \tag{F.24}$$

For example, if $N = 2$, then

$$\mathbf{x}(t, \xi) = \sum_{i=0}^{2} \mathbf{v}_i(t)\phi_i(\xi). \tag{F.25}$$

The unknowns are

$$\begin{aligned} \mathbf{v}_0 &= (v_{01}, v_{02}, v_{03}, v_{04}, v_{05})^T, \\ \mathbf{v}_1 &= (v_{11}, v_{12}, v_{13}, v_{14}, v_{15})^T, \\ \mathbf{v}_2 &= (v_{21}, v_{22}, v_{23}, v_{24}, v_{25})^T. \end{aligned} \tag{F.26}$$

Denote

$$a_{i,j} = \langle \phi_i, \phi_j \rangle,\ a_{i,j,k} = \langle \phi_i \phi_j, \phi_k \rangle, a_{i,j,m,k} = \langle \phi_i \phi_j \phi_m, \phi_k \rangle. \tag{F.27}$$

The resulting equations are

$$\begin{aligned} &\dot{v_{01}} v_{02} a_{0,0,0} + \dot{v_{01}} v_{12} a_{0,1,0} + \dot{v_{01}} v_{22} a_{0,2,0} + \dot{v_{11}} v_{02} a_{1,0,0} + \dot{v_{11}} v_{12} a_{1,1,0} + \dot{v_{11}} v_{22} a_{1,2,0} \\ &+ \dot{v_{21}} v_{02} a_{2,0,0} + \dot{v_{21}} v_{12} a_{2,1,0} + \dot{v_{21}} v_{22} a_{2,2,0} \\ &= \alpha\theta_s - \alpha v_{01}, \\ &\dot{v_{01}} v_{02} a_{0,0,1} + \dot{v_{01}} v_{12} a_{0,1,1} + \dot{v_{01}} v_{22} a_{0,2,1} + \dot{v_{11}} v_{02} a_{1,0,1} + \dot{v_{11}} v_{12} a_{1,1,1} + \dot{v_{11}} v_{22} a_{1,2,1} \\ &+ \dot{v_{21}} v_{02} a_{2,0,1} + \dot{v_{21}} v_{12} a_{2,1,1} + \dot{v_{21}} v_{22} a_{2,2,1} \\ &= -\alpha v_{11}, \end{aligned}$$

$$v_{01}^{\cdot}v_{02}a_{0,0,2} + v_{01}^{\cdot}v_{12}a_{0,1,2} + v_{01}^{\cdot}v_{22}a_{0,2,2} + v_{11}^{\cdot}v_{02}a_{1,0,2} + v_{11}^{\cdot}v_{12}a_{1,1,2} + v_{11}^{\cdot}v_{22}a_{1,2,2}$$

$$+ v_{21}^{\cdot}v_{02}a_{2,0,2} + v_{21}^{\cdot}v_{12}a_{2,1,2} + v_{21}^{\cdot}v_{22}a_{2,2,2}$$

$$= -\alpha v_{21},$$

$$v_{02}^{\cdot}v_{03}a_{0,0,0} + v_{02}^{\cdot}v_{13}a_{0,1,0} + v_{02}^{\cdot}v_{23}a_{0,2,0} + v_{12}^{\cdot}v_{03}a_{1,0,0} + v_{12}^{\cdot}v_{13}a_{1,1,0} + v_{12}^{\cdot}v_{23}a_{1,2,0}$$

$$+ v_{22}^{\cdot}v_{03}a_{2,0,0} + v_{22}^{\cdot}v_{13}a_{2,1,0} + v_{22}^{\cdot}v_{23}a_{2,2,0}$$

$$= \beta\theta_s - \beta v_{01} + wv_{03},$$

$$v_{02}^{\cdot}v_{03}a_{0,0,1} + v_{02}^{\cdot}v_{13}a_{0,1,1} + v_{02}^{\cdot}v_{23}a_{0,2,1} + v_{12}^{\cdot}v_{03}a_{1,0,1} + v_{12}^{\cdot}v_{13}a_{1,1,1} +$$

$$v_{12}^{\cdot}v_{23}a_{1,2,1} + v_{22}^{\cdot}v_{03}a_{2,0,1} + v_{22}^{\cdot}v_{13}a_{2,1,1} + v_{22}^{\cdot}v_{23}a_{2,2,1}.$$

$$= -\beta v_{11} + wv_{13},$$

$$v_{02}^{\cdot}v_{03}a_{0,0,2} + v_{02}^{\cdot}v_{13}a_{0,1,2} + v_{02}^{\cdot}v_{23}a_{0,2,2} + v_{12}^{\cdot}v_{03}a_{1,0,2} + v_{12}^{\cdot}v_{13}a_{1,1,2} + v_{12}^{\cdot}v_{23}a_{1,2,2}$$

$$+ v_{22}^{\cdot}v_{03}a_{2,0,2} + v_{22}^{\cdot}v_{13}a_{2,1,2} + v_{22}^{\cdot}v_{23}a_{2,2,2}$$

$$= -\beta v_{21} + wv_{23},$$

$$v_{03}^{\cdot} = \gamma_\theta v_{02}^{\cdot} - v_{01}^{\cdot} - w\gamma_\theta,$$

$$v_{13}^{\cdot} = \gamma_\theta v_{12}^{\cdot} - v_{11}^{\cdot},$$

$$v_{23}^{\cdot} = \gamma_\theta v_{22}^{\cdot} - v_{21}^{\cdot},$$

$$v_{04}^{\cdot}v_{02}v_{03}a_{0,0,0,0} + v_{04}^{\cdot}v_{02}v_{13}a_{0,0,1,0} + v_{04}^{\cdot}v_{02}v_{23}a_{0,0,2,0} + v_{04}^{\cdot}v_{12}v_{03}a_{0,1,0,0} +$$

$$v_{04}^{\cdot}v_{12}v_{13}a_{0,1,1,0} + v_{04}^{\cdot}v_{12}v_{23}a_{0,1,2,0} + v_{04}^{\cdot}v_{22}v_{03}a_{0,2,0,0} + v_{04}^{\cdot}v_{22}v_{13}a_{0,2,1,0} +$$

$$v_{04}^{\cdot}v_{22}v_{23}a_{0,2,2,0} + v_{14}^{\cdot}v_{02}v_{03}a_{1,0,0,0} + v_{14}^{\cdot}v_{02}v_{13}a_{1,0,1,0} + v_{14}^{\cdot}v_{02}v_{23}a_{1,0,2,0} +$$

$$v_{14}^{\cdot}v_{12}v_{03}a_{1,1,0,0} + v_{14}^{\cdot}v_{12}v_{13}a_{1,1,1,0} + v_{14}^{\cdot}v_{12}v_{23}a_{1,1,2,0} + v_{14}^{\cdot}v_{22}v_{03}a_{1,2,0,0} +$$

$$v_{14}^{\cdot}v_{22}v_{13}a_{1,2,1,0} + v_{14}^{\cdot}v_{22}v_{23}a_{1,2,2,0} + v_{24}^{\cdot}v_{02}v_{03}a_{2,0,0,0} + v_{24}^{\cdot}v_{02}v_{13}a_{2,0,1,0} +$$

$$v_{24}^{\cdot}v_{02}v_{23}a_{2,0,2,0} + v_{24}^{\cdot}v_{12}v_{03}a_{2,1,0,0} + v_{24}^{\cdot}v_{12}v_{13}a_{2,1,1,0} + v_{24}^{\cdot}v_{12}v_{23}a_{2,1,2,0} +$$

$$v_{24}^{\cdot}v_{22}v_{03}a_{2,2,0,0} + v_{24}^{\cdot}v_{22}v_{13}a_{2,2,1,0} + v_{24}^{\cdot}v_{22}v_{23}a_{2,2,2,0} =$$

$$\gamma q_s v_{03} -$$

$$\gamma\left(v_{03}v_{04}a_{0,0,0} + v_{03}v_{14}a_{0,1,0} + v_{03}v_{24}a_{0,2,0} + v_{13}v_{04}a_{1,0,0} + v_{13}v_{14}a_{1,1,0} + v_{13}v_{24}a_{1,2,0} + v_{2,3}v_{04}a_{2,0,0} + v_{23}v_{14}a_{2,1,0} + v_{23}v_{24}a_{2,2,0}\right) + \gamma\kappa\theta_s v_{05} - \gamma\kappa\left(v_{01}v_{05}a_{0,0,0} + v_{01}v_{15}a_{0,1,0} + v_{01}v_{25}a_{0,2,0} + v_{11}v_{05}a_{1,0,0} + v_{11}v_{15}a_{1,1,0} + v_{11}v_{25}a_{1,2,0} + v_{21}v_{05}a_{2,0,0} + v_{21}v_{15}a_{2,1,0} + v_{21}v_{25}a_{2,2,0}\right),$$

$$\dot{v_{04}}v_{02}v_{03}a_{0,0,0,1} + \dot{v_{04}}v_{02}v_{13}a_{0,0,1,1} + \dot{v_{04}}v_{02}v_{23}a_{0,0,2,1} + \dot{v_{04}}v_{12}v_{03}a_{0,1,0,1} + \dot{v_{04}}v_{12}v_{13}a_{0,1,1,1} + \dot{v_{04}}v_{12}v_{23}a_{0,1,2,1} + \dot{v_{04}}v_{22}v_{03}a_{0,2,0,1} + \dot{v_{04}}v_{22}v_{13}a_{0,2,1,1} + \dot{v_{04}}v_{22}v_{23}a_{0,2,2,1} + \dot{v_{14}}v_{02}v_{03}a_{1,0,0,1} + \dot{v_{14}}v_{02}v_{13}a_{1,0,1,1} + \dot{v_{14}}v_{02}v_{23}a_{1,0,2,1} + \dot{v_{14}}v_{12}v_{03}a_{1,1,0,1} + \dot{v_{14}}v_{12}v_{13}a_{1,1,1,1} + \dot{v_{14}}v_{12}v_{23}a_{1,1,2,1} + \dot{v_{14}}v_{22}v_{03}a_{1,2,0,1} + \dot{v_{14}}v_{22}v_{13}a_{1,2,1,1} + \dot{v_{14}}v_{22}v_{23}a_{1,2,2,1} + \dot{v_{24}}v_{02}v_{03}a_{2,0,0,1} + \dot{v_{24}}v_{02}v_{13}a_{2,0,1,1} + \dot{v_{24}}v_{02}v_{23}a_{2,0,2,1} + \dot{v_{24}}v_{12}v_{03}a_{2,1,0,1} + \dot{v_{24}}v_{12}v_{13}a_{2,1,1,1} + \dot{v_{24}}v_{12}v_{23}a_{2,1,2,1} + \dot{v_{24}}v_{22}v_{03}a_{2,2,0,1} + \dot{v_{24}}v_{22}v_{13}a_{2,2,1,1} + \dot{v_{24}}v_{22}v_{23}a_{2,2,2,1} = \gamma q_s v_{13} - \gamma\left(v_{03}v_{04}a_{0,0,1} + v_{03}v_{14}a_{0,1,0} + v_{03}v_{24}a_{0,2,0} + v_{13}v_{04}a_{1,0,0} + v_{13}v_{14}a_{1,1,0} + v_{13}v_{24}a_{1,2,0} + v_{2,3}v_{04}a_{2,0,0} + v_{23}v_{14}a_{2,1,1} + v_{23}v_{24}a_{2,2,1}\right) + \gamma\kappa\theta_s v_{15} - \gamma\kappa\left(v_{01}v_{05}a_{0,0,1} + v_{01}v_{15}a_{0,1,1} + v_{01}v_{25}a_{0,2,1} + v_{11}v_{05}a_{1,0,1} + v_{11}v_{15}a_{1,1,1} + v_{11}v_{25}a_{1,2,1} + v_{21}v_{05}a_{2,0,1} + v_{21}v_{15}a_{2,1,1} + v_{21}v_{25}a_{2,2,1}\right),$$

$$\dot{v_{04}}v_{02}v_{03}a_{0,0,0,2} + \dot{v_{04}}v_{02}v_{13}a_{0,0,1,2} + \dot{v_{04}}v_{02}v_{23}a_{0,0,2,2} + \dot{v_{04}}v_{12}v_{03}a_{0,1,0,2} + \dot{v_{04}}v_{12}v_{13}a_{0,1,1,2} + \dot{v_{04}}v_{12}v_{23}a_{0,1,2,2} + \dot{v_{04}}v_{22}v_{03}a_{0,2,0,2} + \dot{v_{04}}v_{22}v_{13}a_{0,2,1,2} + \dot{v_{04}}v_{22}v_{23}a_{0,2,2,2} + \dot{v_{14}}v_{02}v_{03}a_{1,0,0,2} + \dot{v_{14}}v_{02}v_{13}a_{1,0,1,2} + \dot{v_{14}}v_{02}v_{23}a_{1,0,2,2} + \dot{v_{14}}v_{12}v_{03}a_{1,1,0,2} + \dot{v_{14}}v_{12}v_{13}a_{1,1,1,2} + \dot{v_{14}}v_{12}v_{23}a_{1,1,2,2} + \dot{v_{14}}v_{22}v_{03}a_{1,2,0,2} + \dot{v_{14}}v_{22}v_{13}a_{1,2,1,2} + \dot{v_{14}}v_{22}v_{23}a_{1,2,2,2} + \dot{v_{24}}v_{02}v_{03}a_{2,0,0,2} + \dot{v_{24}}v_{02}v_{13}a_{2,0,1,2} + \dot{v_{24}}v_{02}v_{23}a_{2,0,2,2} + \dot{v_{24}}v_{12}v_{03}a_{2,1,0,2} + \dot{v_{24}}v_{12}v_{13}a_{2,1,1,2} + \dot{v_{24}}v_{12}v_{23}a_{2,1,2,2} +$$

$$v_{24}^{\cdot} v_{22} v_{03} a_{2,2,0,2} + v_{24}^{\cdot} v_{22} v_{13} a_{2,2,1,2} + v_{24}^{\cdot} v_{22} v_{23} a_{2,2,2,2} =$$

$$\gamma q_s v_{23} -$$

$$\gamma\big(v_{03} v_{04} a_{0,0,2} + v_{03} v_{14} a_{0,1,2} + v_{03} v_{24} a_{0,2,2} + v_{13} v_{04} a_{1,0,2} + v_{13} v_{14} a_{1,1,2} +$$

$$v_{13} v_{24} a_{1,2,2} + v_{2,3} v_{04} a_{2,0,2} + v_{23} v_{14} a_{2,1,2} + v_{23} v_{24} a_{2,2,2}\big) + \gamma\kappa\theta_s v_{25} -$$

$$\gamma\kappa\big(v_{01} v_{05} a_{0,0,2} + v_{01} v_{15} a_{0,1,2} + v_{01} v_{25} a_{0,2,2} + v_{11} v_{05} a_{1,0,2} + v_{11} v_{15} a_{1,1,2} +$$

$$v_{11} v_{25} a_{1,2,2} + v_{21} v_{05} a_{2,0,2} + v_{21} v_{15} a_{2,1,2} + v_{21} v_{25} a_{2,2,2}\big),$$

$$v_{05}^{\cdot} = \gamma_\theta v_{02}^{\cdot} - v_{04}^{\cdot} - w\gamma_q,$$

$$v_{15}^{\cdot} = \gamma_\theta v_{12}^{\cdot} - v_{14}^{\cdot},$$

$$v_{25}^{\cdot} = \gamma_\theta v_{22}^{\cdot} - v_{24}^{\cdot}. \qquad \text{(F.28)}$$

After calculating the coefficients in (F.27) and substituting them into the above equations, the resulted equations become

$$v_{01}^{\cdot} v_{02} + v_{11}^{\cdot} v_{12} + v_{21}^{\cdot} v_{22} = \alpha\theta_s - \alpha v_{01},$$

$$v_{01}^{\cdot} v_{12} + v_{11}^{\cdot} v_{02} + v_{11}^{\cdot} v_{22}\sqrt{2} + v_{21}^{\cdot} v_{12}\sqrt{2} = -\alpha v_{11},$$

$$v_{01}^{\cdot} v_{22} + v_{11}^{\cdot} v_{12}\sqrt{2} + v_{21}^{\cdot} v_{02} + v_{21}^{\cdot} v_{22} 2\sqrt{2} = -\alpha v_{21},$$

$$v_{02}^{\cdot} v_{03} + v_{12}^{\cdot} v_{13} + v_{22}^{\cdot} v_{23} = \beta\theta_s - \beta v_{01} + w v_{03},$$

$$v_{02}^{\cdot} v_{13} + v_{12}^{\cdot} v_{03} + v_{12}^{\cdot} v_{23}\sqrt{2} + v_{22}^{\cdot} v_{13}\sqrt{2} = -\beta v_{11} + w v_{13},$$

$$v_{02}^{\cdot} v_{23} + v_{12}^{\cdot} v_{13}\sqrt{2} + v_{22}^{\cdot} v_{03} + v_{22}^{\cdot} v_{23} 2\sqrt{2} = -\beta v_{21} + w v_{23},$$

$$v_{03}^{\cdot} = \gamma_\theta v_{02}^{\cdot} - v_{01}^{\cdot} - w\gamma_\theta,$$

$$v_{13}^{\cdot} = \gamma_\theta v_{12}^{\cdot} - v_{11}^{\cdot},$$

$$v_{23}^{\cdot} = \gamma_\theta v_{22}^{\cdot} - v_{21}^{\cdot},$$

$$\dot{v_{04}}v_{02}v_{03} + \dot{v_{04}}v_{12}v_{13} + \dot{v_{04}}v_{22}v_{23} + \dot{v_{14}}v_{02}v_{13} + \dot{v_{14}}v_{12}v_{03} + \dot{v_{14}}v_{12}v_{23}\sqrt{2} +$$

$$\dot{v_{14}}v_{22}v_{13}\sqrt{2} + \dot{v_{24}}v_{02}v_{23} + \dot{v_{24}}v_{12}v_{13}\sqrt{2} + \dot{v_{24}}v_{22}v_{03} + \dot{v_{24}}v_{22}v_{23}2\sqrt{2} = \gamma q_s v_{03} -$$

$$\gamma(v_{03}v_{04} + v_{13}v_{14} + v_{23}v_{24}) + \gamma\kappa\theta_s v_{05} - \gamma\kappa(v_{01}v_{05} + v_{11}v_{15} + v_{21}v_{25}),$$

$$\dot{v_{04}}v_{02}v_{13} + \dot{v_{04}}v_{12}v_{03} + \dot{v_{04}}v_{12}v_{23}\sqrt{2} + \dot{v_{04}}v_{22}v_{13}\sqrt{2} + \dot{v_{14}}v_{02}v_{03} +$$

$$\dot{v_{14}}v_{02}v_{23}\sqrt{2} + \dot{v_{14}}v_{12}v_{13}3 + \dot{v_{14}}v_{22}v_{03}\sqrt{2} + \dot{v_{14}}v_{22}v_{23}5 + \dot{v_{24}}v_{02}v_{13}\sqrt{2} +$$

$$\dot{v_{24}}v_{12}v_{03}\sqrt{2} + \dot{v_{24}}v_{12}v_{23}5 + \dot{v_{24}}v_{22}v_{13}5 = \gamma q_s v_{13} - \gamma\left(v_{13}v_{14} + v_{23}v_{14}\sqrt{2}\right) +$$

$$\gamma\kappa\theta_s v_{15} - \gamma\kappa\left(v_{01}v_{15} + v_{11}v_{05} + v_{11}v_{25}\sqrt{2} + v_{21}v_{15}\sqrt{2}\right),$$

$$\dot{v_{04}}v_{02}v_{23} + \dot{v_{04}}v_{12}v_{13}\sqrt{2} + \dot{v_{04}}v_{22}v_{03} + \dot{v_{04}}v_{22}v_{23}2\sqrt{2} + \dot{v_{14}}v_{02}v_{13}\sqrt{2} +$$

$$\dot{v_{14}}v_{12}v_{03}\sqrt{2} + \dot{v_{14}}v_{12}v_{23}5 + \dot{v_{14}}v_{22}v_{13}5 + \dot{v_{24}}v_{02}v_{03} + \dot{v_{24}}v_{02}v_{23}2\sqrt{2} +$$

$$\dot{v_{24}}v_{12}v_{13}5 + \dot{v_{24}}v_{22}v_{03}2\sqrt{2} + \dot{v_{24}}v_{22}v_{23}15 = \gamma q_s v_{23} - \gamma\left(v_{03}v_{24} + v_{13}v_{14}\sqrt{2} +\right.$$

$$\left.v_{23}v_{04} + v_{23}v_{24}2\sqrt{2}\right) + \gamma\kappa\theta_s v_{25} - \gamma\kappa\left(v_{01}v_{25} + v_{11}v_{15}\sqrt{2} + v_{21}v_{05} + v_{21}v_{25}2\sqrt{2}\right),$$

$$\dot{v_{05}} = \gamma_\theta \dot{v_{02}} - \dot{v_{04}} - w\gamma_q,$$

$$\dot{v_{15}} = \gamma_\theta \dot{v_{12}} - \dot{v_{14}},$$

$$\dot{v_{25}} = \gamma_\theta \dot{v_{22}} - \dot{v_{24}}. \qquad \text{(F.29)}$$

Appendix G

Legendre Polynomials

This appendix provides a detailed introduction of Legendre polynomials and their properties in single and multiple variables.

1. Legendre Polynomial – Scalar Case

In mathematics, the Legendre functions of the first kind, sometimes called Legendre coefficients or zonal harmonics (Whittaker and Watson 1990) are solutions to the Legendre differential equation,

$$(1-x^2)y'' - 2xy' + uy = [(1-x^2)y']' + uy = 0, \tag{G.1}$$

which are possible only if

$$u = m(m+1), m \text{ is a real number.} \tag{G.2}$$

The solutions of this equation are called Legendre Functions of degree m. If m is a non-negative integer, i.e., $m = 0, 1, 2, 3, ...$, the Legendre functions are often referred as Legendre polynomials $P_m(x)$.

The Legendre polynomials $P_m(x)$ of degree m in a scalar variable x can be expressed by Rodrigues' formula as

$$P_m(x) = \frac{1}{2^m m!}\frac{d^m}{dx^m}(x^2-1)^m, m = 0, 1, 2, ... \tag{G.3}$$

One generating function of a Legendre Polynomial is

$$\frac{1}{\sqrt{1-2tx+t^2}} = \sum_{m=0}^{\infty} P_m(x)t^m, \tag{G.4}$$

where P_m can be defined as the coefficients of the Taylor series expansion in (G.4).

In generating a Legendre polynomial for a specific degree m, the following formula is useful,

$$P_m(x) = \frac{1}{2^m}\sum_{k=0}^{m}\binom{m}{k}^2 (x-1)^{m-k}(x+1)^k$$

$$= \sum_{k=0}^{m}\binom{m}{k}\binom{-m-1}{k}\left(\frac{1-x}{2}\right)^k$$

$$= 2^m \sum_{k=0}^{m} x^k \binom{m}{k}\binom{\frac{m+k-1}{2}}{m}, \tag{G.5}$$

where

$$\binom{m}{k} = \frac{m!}{k!(m-k)!}. \tag{G.6}$$

2. Orthogonality Property

Let's first define an inner product which is similar as that defined in Appendix A for Hermite Polynomials. For integers m and k, the inner product of $P_m(x)$ and $P_k(x)$ is defined as

$$\langle P_m P_k \rangle_w = \int_{-1}^{1} P_m(x) P_k(x)\, w(x) dx, \tag{G.7}$$

where $w(x)$ is the weighting function. Here, in consistent with the definition for Hermite polynomials, $w(x)$ is defined as the weighting function for uniform distribution in range [-1, 1], i.e.,

$$w(x) = \frac{1}{2}. \tag{G.8}$$

This inner product induces a norm $\|P_m(x)\|_w$ of $P_m(x)$ defined by

$$\|P_m(x)\|_w^2 = \int_{-1}^{1} P_m^2(x)\, w(x) dx. \tag{G.9}$$

It can be verified that

$$\langle P_m P_k \rangle_w = \|P_m\|_w^2 \delta_{mk}, \tag{G.10}$$

where $\delta_{mk} = 0$ if $m \neq k$, and $= 1$ if $m = k$. That is, P_m and P_k are orthogonal for $m \neq k$, and the norm of $P_m(x)$ is

$$\|P_m\|_w = \frac{1}{\sqrt{2m+1}}. \tag{G.11}$$

Consequently, $\{P_m(x)\}_{m\geq 0}$ constitute an orthogonal system of polynomials and $\left\{\sqrt{2m+1}P_m(x)\right\}_{m\geq 0}$ constitute an orthonormal system. $\left\{\sqrt{2m+1}P_m(x)\right\}_{m\geq 0}$ are called normalized Legendre polynomials.

As an example, the following Table G.1 gives a list of the first few Legendre polynomials ($0 \leq m \leq 6$) with their norms as defined in formula (G.9),

Table G.1 A list of $P_m(x)$, $0 \leq m \leq 6$

Degree m	$P_m(x)$	$\|P_m\|_w^2$
0	1	1
1	x	$\frac{1}{3}$
2	$\frac{1}{2}(3x^2-1)$	$\frac{1}{5}$
3	$\frac{1}{2}(5x^3-3x)$	$\frac{1}{7}$
4	$\frac{1}{8}(35x^4-30x^2+3)$	$\frac{1}{9}$
5	$\frac{1}{8}(63x^5-70x^3+15x)$	$\frac{1}{11}$
6	$\frac{1}{16}(231x^6-315x^4+105x^2-5)$	$\frac{1}{13}$

3. Legendre Polynomials – Multivariate Case

Similar as the multivariate case for Hermite polynomials, let $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in R^n$ be a *n*-variate random variable with weight function

$$W(\mathbf{x}) = \frac{1}{2^n} = \prod_{i=1}^n \frac{1}{2} = \prod_{i=1}^n w(x_i), \qquad \text{(G.12)}$$

where $w(x_i)$ is defined in (G.8).

Given m and one of its partitions $(p_1, p_2, \dots, p_n)$, define an n-variate homogeneous Legendre polynomial of degree m:

$$P_{p_1 p_2 \dots p_n}(\mathbf{x}) = \prod_{i=1}^n P_{p_i}(x_i) = \prod_{i=1}^n \frac{1}{2^{p_i} p_i!} \frac{d^{p_i}}{dx_i{}^{p_i}} (x_i^2-1)^{p_i}. \qquad \text{(G.13)}$$

By combining the multiplicative decomposition of the multi-variate Legendre polynomials in terms of the univariate Legendre polynomials and using the orthogonality of the latter, the orthogonality of the multivariate Legendre polynomials can be readily inferred. Therefore, for $m = p_1 + p_2 + \cdots + p_n$ and $k = q_1 + q_2 + \cdots + q_n$, then

$$\langle P_m(\mathbf{x}) P_k(\mathbf{x}) \rangle_{\mathrm{W}} = 0, \text{ (if } m \neq k)$$

$$= \prod_{i=1}^{n} \left\| P_{p_i}(x_i) \right\|_w^2 \delta_{p_i q_i}, \text{ (if } m = k) \qquad \text{(G.14)}$$

Clearly,

$$\left\| \mathrm{P}_{p_1 p_2 \ldots p_n}(\mathbf{x}) \right\|_W^2 = \prod_{i=1}^{n} \frac{1}{2p_i + 1}. \qquad \text{(G.15)}$$

It can be verified that there are exactly $\binom{m+n-1}{n}$ linearly independent n -variate Legendre polynomials of degree m. Further, it can be verified that the total number of linearly independent n -variate Legendre polynomials of degree less than or equal to m is $\binom{m+n}{n}$.

Table G.2 provides a list of the set of all 15 two-variate ($n = 2$) Legendre polynomials of degree less than or equal to 4. The last column in Table G.2 gives the norm $\left\| P_{p_1 p_2}(x_1, x_2) \right\|_W^2$.

4. **Legendre Polynomials Chaos**

Similarly, the properties of the Legendre polynomials $P_m(x)$ can directly carry over to $P_m(\xi)$ where ξ is a uniform distributed random variable over [-1, 1].

The following properties of $P_m(\xi)$ can be easily verified.

(1) Examples of $P_m(\xi)$ are obtained by replacing x by ξ in Table G.1.

(2) $E[P_m(\xi)] = 0$, if $m > 0$.

Table G.2 Two-variate (*n*=2) Legendre polynomials, degree less than or equal to 4

Degree m	**Multi index** $(p_1 p_2)$	$P_{p_1 p_2}(x_1, x_2)$	$P_{p_1}(x_1)P_{p_2}(x_2)$	$\|P_{p_1 p_2}(x_1, x_2)\|_W^2$
0	0 0	1	1	1
1	1 0	x_1	$P_1(x_1)P_0(x_2)$	$\frac{1}{3}$
	0 1	x_2	$P_0(x_1)P_1(x_2)$	$\frac{1}{3}$
2	2 0	$\frac{1}{2}(3x_1^2 - 1)$	$P_2(x_1)P_0(x_2)$	$\frac{1}{5}$
	1 1	$x_1 x_2$	$P_1(x_1)P_1(x_2)$	$\frac{1}{9}$
	0 2	$\frac{1}{2}(3x_2^2 - 1)$	$P_0(x_1)P_2(x_2)$	$\frac{1}{5}$
3	3 0	$\frac{1}{2}(5x_1^3 - 3x_1)$	$P_3(x_1)P_0(x_2)$	$\frac{1}{7}$
	2 1	$\frac{1}{2}(3x_1^2 - 1)x_2$	$P_2(x_1)P_1(x_2)$	$\frac{1}{15}$
	1 2	$\frac{1}{2}x_1(3x_2^2 - 1)$	$P_1(x_1)P_2(x_2)$	$\frac{1}{15}$
	0 3	$\frac{1}{2}(5x_2^3 - 3x_2)$	$P_0(x_1)P_3(x_2)$	$\frac{1}{7}$
4	4 0	$\frac{1}{8}(35x_1^4 - 30x_1^2 + 3)$	$P_4(x_1)P_0(x_2)$	$\frac{1}{9}$
	3 1	$\frac{1}{2}(5x_1^3 - 3x_1)x_2$	$P_3(x_1)P_1(x_2)$	$\frac{1}{21}$
	2 2	$\frac{1}{4}(3x_1^2 - 1)(3x_2^2 - 1)$	$P_2(x_1)P_2(x_2)$	$\frac{1}{25}$
	1 3	$\frac{1}{2}x_1(5x_2^3 - 3x_2)$	$P_1(x_1)P_3(x_2)$	$\frac{1}{21}$
	0 4	$\frac{1}{8}(35x_2^4 - 30x_2^2 + 3)$	$P_0(x_1)P_4(x_2)$	$\frac{1}{9}$

(3) $\{P_m(\xi)\}_{m\geq 0}$ are orthogonal, that is,

$$\langle P_m(\xi), P_k(\xi)\rangle = E[P_m(\xi)P_k(\xi)] = 0 \text{ if } m \neq k.$$

(4) The norm $\|P_m(\xi)\|$ of $P_m(\xi)$ is defined by

$$\langle P_m(\xi), P_m(\xi)\rangle = \|P_m(\xi)\|^2 = E[P_m^2(\xi)] = \frac{1}{2m+1},$$

$$\text{i.e., } \|P_m(\xi)\| = \frac{1}{\sqrt{2m+1}}.$$

(5) $\left\{\sqrt{2m+1}P_m(\xi)\right\}_{m\geq 0}$ form an orthonormal system of Legendre polynomials.

The above properties can be readily extended to multivariate Legendre polynomials over a set of n standard uniform distributed random variables $\xi_1, \xi_2, \dots, \xi_n$ (independent with each other):

(6) Let $p_1 + p_2 + \cdots + p_n = m$, where $0 \leq p_i \leq m$ for $1 \leq i \leq n$. Then

$$P_{p_1,p_2,\dots,p_n}(\xi_1, \xi_2, \dots, \xi_n) = P_{p_1}(\xi_1)P_{p_2}(\xi_2) \dots P_{p_n}(\xi_n).$$

(7) $\left\|P_{p_1,p_2,\dots,p_n}(\xi_1, \xi_2, \dots, \xi_n)\right\|^2 = \prod_{i=1}^{n} \frac{1}{2p_i+1}$.