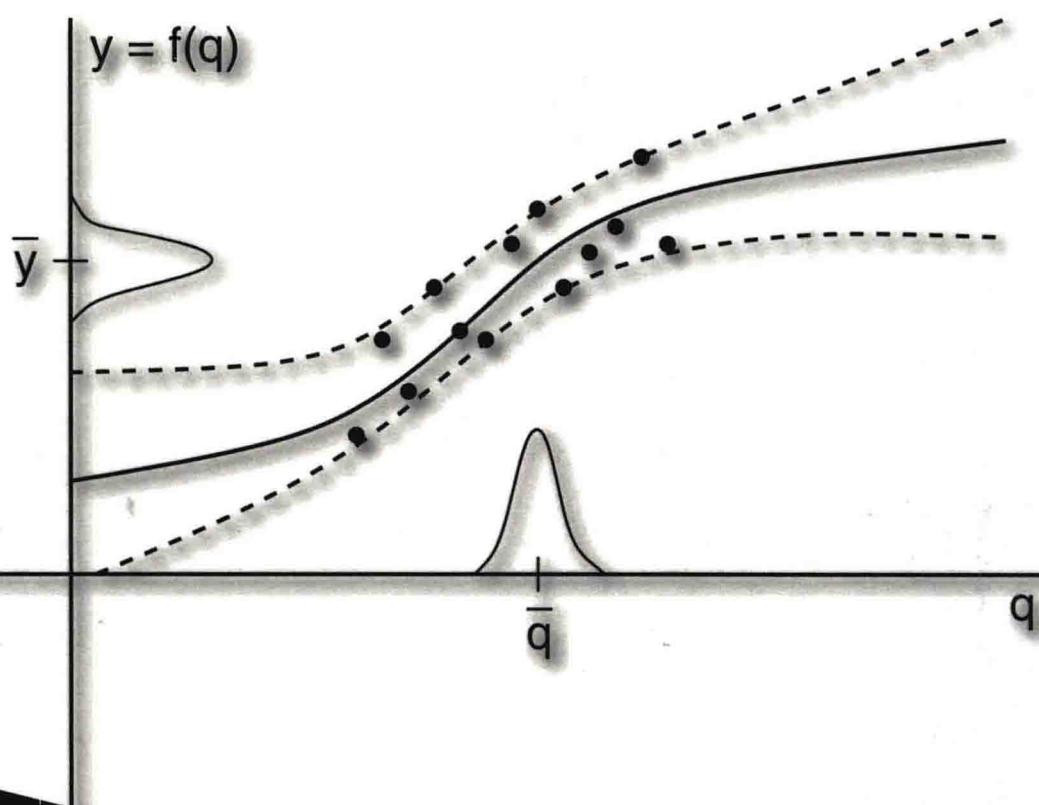


RALPH C. SMITH

Uncertainty Quantification

Theory, Implementation, and Applications



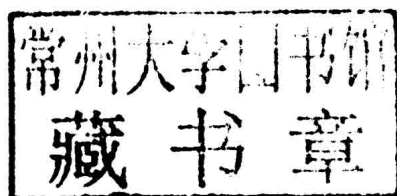
siam
Computational Science & Engineering

RALPH C. SMITH

North Carolina State University
Raleigh, North Carolina

Uncertainty Quantification

Theory, Implementation, and Applications



siam.

Society for Industrial and Applied Mathematics
Philadelphia

Copyright © 2014 by the Society for Industrial and Applied Mathematics

10 9 8 7 6 5 4 3 2 1

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688 USA.

Trademarked names may be used in this book without the inclusion of a trademark symbol. These names are used in an editorial context only; no infringement of trademark is intended.

MATLAB is a registered trademark of The MathWorks, Inc. For MATLAB product information, please contact The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098 USA, 508-647-7000, Fax: 508-647-7001, info@mathworks.com, www.mathworks.com.

Figures 1.2, 2.1, and 2.4 reprinted courtesy of NOAA.

Figure 2.3 reprinted with permission from ECMWF.

Figures 2.6, 2.9, and 2.11 reprinted with permission from IPCC.

Figures 2.12 and 2.15 reprinted courtesy of Kansas Geological Survey. All rights reserved.

Figure 2.13 reprinted courtesy of USGS.

Figure 2.14 reprinted courtesy of Mike Norton.

Figure 2.16 reprinted courtesy of USNRC.

Figure 2.17 reprinted courtesy of the U.S. Department of Energy.

Figures 2.18 and 2.19 reprinted with kind permission from Springer Science and Business Media.

Figures 6.7 (a), 8.8, and 8.9 (b) reprinted with permission from SAGE.

This research was supported by the Air Force Office of Scientific Research (Dynamics and Control Program), the Department of Energy Consortium for Advanced Simulation of Light Water Reactors (CASL), and the National Science Foundation (Research Training Groups in the Mathematical Sciences).

Library of Congress Cataloging-in-Publication Data

Smith, Ralph C., 1960- author.

Uncertainty quantification : theory, implementation, and applications / Ralph Smith, North Carolina State University, Raleigh, North Carolina.

pages cm. – (Computational science and engineering series)

Includes bibliographical references and index.

ISBN 978-1-611973-21-1

1. Measurement uncertainty (Statistics) 2. Estimation theory. I. Title.

QA276.8.S64 2013

519.5'44-dc23

2013034432

siam is a registered trademark.

Uncertainty Quantification

Computational Science & Engineering

The SIAM series on Computational Science and Engineering publishes research monographs, advanced undergraduate- or graduate-level textbooks, and other volumes of interest to an interdisciplinary CS&E community of computational mathematicians, computer scientists, scientists, and engineers. The series includes both introductory volumes aimed at a broad audience of mathematically motivated readers interested in understanding methods and applications within computational science and engineering and monographs reporting on the most recent developments in the field. The series also includes volumes addressed to specific groups of professionals whose work relies extensively on computational science and engineering.

SIAM created the CS&E series to support access to the rapid and far-ranging advances in computer modeling and simulation of complex problems in science and engineering, to promote the interdisciplinary culture required to meet these large-scale challenges, and to provide the means to the next generation of computational scientists and engineers.

Editor-in-Chief

Donald Estep
Colorado State University

Editorial Board

Omar Ghattas
University of Texas at Austin

Max Gunzburger
Florida State University

Des Higham
University of Strathclyde

Michael Holst
University of California, San Diego

David Keyes
Columbia University and KAUST

Max D. Morris
Iowa State University

Alex Pothen
Purdue University

Padma Raghavan
Pennsylvania State University

Karen Willcox
Massachusetts Institute of Technology

Series Volumes

Smith, Ralph C., *Uncertainty Quantification: Theory, Implementation, and Applications*

Dankowicz, Harry and Schilder, Frank, *Recipes for Continuation*

Mueller, Jennifer L. and Siltanen, Samuli, *Linear and Nonlinear Inverse Problems with Practical Applications*

Shapira, Yair, *Solving PDEs in C++: Numerical Methods in a Unified Object-Oriented Approach, Second Edition*

Borzi, Alfio and Schulz, Volker, *Computational Optimization of Systems Governed by Partial Differential Equations*

Ascher, Uri M. and Greif, Chen, *A First Course in Numerical Methods*

Layton, William, *Introduction to the Numerical Analysis of Incompressible Viscous Flows*

Ascher, Uri M., *Numerical Methods for Evolutionary Differential Equations*

Zohdi, T. I., *An Introduction to Modeling and Simulation of Particulate Flows*

Biegler, Lorenz T., Ghattas, Omar, Heinkenschloss, Matthias, Keyes, David, and van Bloemen Waanders, Bart, Editors, *Real-Time PDE-Constrained Optimization*

Chen, Zhangxin, Huan, Guanren, and Ma, Yuanle, *Computational Methods for Multiphase Flows in Porous Media*

Shapira, Yair, *Solving PDEs in C++: Numerical Methods in a Unified Object-Oriented Approach*

Preface

Uncertainty quantification is both a new field and one that is as old as the disciplines of probability and statistics. The present novelty lies in the synthesis of probability, statistics, model development, mathematical and numerical analysis, large-scale simulations, experiments, and disciplinary sciences to provide a computational framework for quantifying input and response uncertainties in a manner that facilitates predictions with quantified and reduced uncertainty. This is the topic of this book.

Uncertainty quantification for physical models can be motivated in the context of weather modeling. Models for complex phenomena, such as dust-induced cloud formation, are approximate and uncertain, as are the parameters in these models. Additional errors and uncertainties are introduced by the numerical algorithms and experimental data used to approximate and calibrate the models. In the first step of the prediction process, data assimilation or model calibration techniques are used to determine input parameters and initial conditions so that quantities of interest, such as temperature or relative humidity, match current conditions. The second step entails the prediction of future weather conditions with uncertainties quantified by probabilistic statements—e.g., 95% change of rain—or uncertainty cones of the type reported for hurricanes or tropical storms.

Whereas model calibration and uncertainty propagation comprise the primary aspects of the prediction process, their implementation for large-scale applications requires a wide range of supporting topics. These include aspects of probability, statistics, analysis, and numerical analysis as well as the following topics: parameter selection, surrogate model construction, local and global sensitivity analysis, and quantification of model discrepancies. The interdisciplinary nature of the field is augmented by the fact that all of these components must be investigated and implemented in the context of the underlying applications.

The explosive growth of uncertainty quantification as an interdisciplinary field is due to a number of factors: increasing emphasis on models having quantified uncertainties for large-scale applications, novel algorithm development, and new computational architectures that facilitate implementation of these algorithms.

In Chapter 2, we detail five applications where model predictions with quantified uncertainties are critical for understanding and predicting scientific phenomena and making informed decisions and designs based on these predictions. These applications are weather models, climate models, subsurface hydrology and geology

models, nuclear reactor models, and models for biological phenomena. Whereas the presence and role of uncertainties in these applications has long been recognized, the development of computational models that quantify and incorporate uncertainties is receiving increased attention. The reliance of scientists and policy makers on such models is expected to grow rapidly as the field of uncertainty quantification for predictive sciences matures and computational resources evolve.

The relatively recent development of supporting mathematical and statistical theory and algorithms is a second factor supporting the growth of the field. For example, the adaptive DRAM and DREAM algorithms discussed in Chapter 8 for Bayesian model calibration were developed within the last ten years. These algorithms are presently being investigated in the context of climate and groundwater models. Similarly, much of the sparse grid theory discussed in Chapter 11 was developed in the last twenty years, although the original concept is much older.

The availability of massively parallel computer architectures and hardware has further bolstered uncertainty quantification for complex and large-scale applications. The DREAM algorithms are inherently parallel, and recent versions of DRAM are being implemented on parallel architectures. It is anticipated that field programmable gate arrays (FPGAs) will be increasingly utilized for uncertainty quantification as high-level tools are developed to reduce programming overhead. The fact that we operate in increasingly data-rich environments will also benefit uncertainty quantification, and we anticipate increased interaction between data mining, high-dimensional visualization, and uncertainty quantification.

The growth in the field has spawned the introduction of interdisciplinary courses on uncertainty quantification, and this text owes its genesis to the author's development of such a course at North Carolina State University in 2008. This text was written with the goal of introducing advanced undergraduates, graduate students, postdocs, and researchers in mathematics, statistics, engineering, and natural and biological sciences to the various topics comprising uncertainty quantification for predictive models. To achieve this, we motivate a number of the topics using very basic examples that should be familiar to most readers. We have included numerous definitions and significant detail to provide a common footing for a wide range of readers. Because this is a new and evolving field, we indicate open research questions at various points in the text and provide research references in the Notes and References at the end of each chapter.

Various resources will be maintained at the website <http://www.siam.org/books/cs12> to augment the text and provide a mechanism to update the material. This includes data employed in exercises as well as a future erratum.

This text has benefited significantly from graduate students, postdocs, and colleagues whose comments have improved the exposition and reduced the number of typos by orders of magnitude. Specifically, sincere thanks are extended to Nate Burch, Amanda Coons, John Crews, John Harlim, Zhengzheng Hu, Zack Kenz, Christine Latten, Jerry McMahan Jr., Keri Rehm, Mami Wentworth, and Lucas Van Blaircum for their attention to detail and candid feedback regarding parts of the manuscript. The author is also extremely grateful to Brian Adams and Karen Willcox for their feedback during the review process; the book is significantly improved due to their detailed comments.

The support provided by several funding agencies has been instrumental both for related research and the writing of this text. These agencies include the Air Force Office of Scientific Research (Dynamics and Control Program), the Department of Energy Consortium for Advanced Simulation of Light Water Reactors (CASL), and the National Science Foundation (Research Training Groups in the Mathematical Sciences). Part of this text was written while the author was a Faculty Fellow in the 2011–12 Statistical and Applied Mathematical Sciences Institute (SAMSI) *Program on Uncertainty Quantification*. Collaboration and interactions during this year significantly influenced aspects of the book, and the author very gratefully acknowledges the scientific and financial contributions from this program. Finally, I would like to thank Elizabeth Greenspan from SIAM for her assistance and encouragement throughout the process of writing this book.

Ralph C. Smith
North Carolina State University
Raleigh, NC
July 1, 2013

Notation

This compilation does not include all of the symbols used throughout the text, and we neglect those that appear one time in a specific context such as those in the models of Chapter 2. Instead, it is meant to clarify the role of symbols that appear multiple times throughout the discussion.

| Symbol | Meaning | Page |
|---------------------------------------|--|----------|
| $\partial\mathcal{D}, \partial\Omega$ | Boundaries of regions \mathcal{D} and Ω | 62, 63 |
| $\alpha(q^* q^{k-1})$ | Probability of accepting candidate q^* | 159 |
| γ_i | Normalization factor for $\langle \cdot, \cdot \rangle_\rho$ | 209 |
| Γ_i, Γ | Range of i^{th} random parameter, random vector | 108 |
| $\delta, \delta(x_i), \delta(t_i)$ | Model discrepancy or error | 133, 257 |
| ε, ϵ | Random and realized measurement errors | 82, 132 |
| Λ_{M_1} | Lebesgue constant | 252 |
| μ | Mean | 70 |
| μ_i, μ_i^* | Morris sensitivity measures | 332 |
| ν | Dimension of model response $y(t, q)$ | 61 |
| $\pi_0(q), \pi(q y)$ | Bayesian prior and posterior density | 100 |
| $\pi(y q)$ | Bayesian likelihood function | 100 |
| $\rho_{Q_i}(q_i), \rho_Q(q)$ | Density for i^{th} random parameter, random vector | 108 |
| σ_0^2 | Unknown measurement error variance | 135 |
| $\hat{\sigma}^2, \sigma^2$ | Estimator and estimate for σ_0^2 | 135 |
| σ_j | Singular values of the matrix A | 117 |
| Σ | Matrix of singular values of matrix A | 117 |
| v | Realized measurements of Υ | 132, 156 |
| Υ | Random variable for measurements | 82 |
| $\phi_i(x)$ | Spatial basis functions | 219 |
| χ | Independent variables $\chi = [x, t] \in \mathcal{D} \times \mathcal{T} \equiv \Omega$ | 63 |
| $\psi_k(Q), \Psi_k(Q)$ | Univariate, multivariate orthogonal polynomials | 209, 213 |
| $\mathcal{A}(q, p)$ | Sparse grid quadrature operator | 247 |
| $B(u, q), B(q)u$ | Boundary operators | 62, 63 |

| Symbol | Meaning | Page |
|--|---|------|
| \mathcal{C} | Observation matrix or vector | 61 |
| $d_i(q), d_i^\sigma(q)$ | Morris elementary effects for i^{th} input | 331 |
| D, D_i, D_{ij} | Total and partial variances of response Y | 324 |
| \mathcal{D} | Spatial domain in $\mathbb{R}^1, \mathbb{R}^2$, or \mathbb{R}^3 | 62 |
| $f(q), f(t, x, q)$ | Model response | 132 |
| $\tilde{f}(q), \tilde{f}(t, x, q)$ | Surrogate model | 274 |
| $F(q)$ | Source terms | 62 |
| H_u, H_q, H_F | Hilbert spaces for state, parameters, and source | 63 |
| $H_i(Q)$ | Hermite polynomials | 210 |
| $\mathcal{H}(q, p)$ | Sparse quadrature grid | 248 |
| $\mathbf{i}', \mathbf{j}', \mathbf{k}'$ | Multi-indices | 212 |
| $I^{(p)}f$ | Integral operator in \mathbb{R}^p | 240 |
| $\mathbb{I}(\ell)$ | Multi-index sets | 246 |
| $\mathcal{I}_\ell^{(p)}u$ | Interpolation operator in \mathbb{R}^p | 254 |
| $I(q)$ | Identifiable subspace | 113 |
| $\mathcal{I}(q)$ | Space of influential parameters | 114 |
| $\mathcal{J}(q)$ | Least squares functional | 135 |
| $J(q^* q^{k-1})$ | Proposal or jumping distribution | 159 |
| $\ell(q v)$ | Log-likelihood function | 83 |
| $L(q v)$ | Likelihood function | 83 |
| $L(q)u$ | Linear operator | 63 |
| $L_m(\dot{q})$ | Lagrange interpolating polynomial | 251 |
| $L_{\rho_i}^2(\Gamma_i), L_\rho^2(\Gamma)$ | Square integrable functions on Γ_i, Γ | 215 |
| M, M_ℓ | Number of collocation points or samples | 253 |
| n | Number of measurements or model evaluations | 61 |
| N | Dimension of state u | 61 |
| $NI(q)$ | Unidentifiable subspace | 113 |
| $\mathcal{NI}(q)$ | Space of noninfluential parameters | 114 |
| $\mathcal{N}(A)$ | Null space of the matrix A | 116 |
| $\mathcal{N}(u, q)$ | Linear or nonlinear differential operator | 62 |
| p | Number of parameters | 100 |
| $P_i(Q)$ | Legendre polynomials | 211 |
| \mathbb{P}_k | Space of polynomials with argument less than or equal to k | 208 |
| $\widehat{\mathbb{P}}_k$ | Polynomials in \mathbb{P}_k that are orthogonal to \mathbb{P}_{k-1} | 208 |

| Symbol | Meaning | Page |
|------------------------------------|---|----------|
| q_0 | True but unknown parameter | 82 |
| $q = [q_1, \dots, q_p]$ | Realizations of Q | 100 |
| q^* | Proposed Markov chain parameter | 159 |
| q^{k-1} | Parameter at $k - 1$ step in Markov chain | 159 |
| q^r, q^m | Quadrature, collocation, and sample points | 211, 217 |
| \hat{q}_{OLS}, q_{OLS} | Least squares estimator, estimate for q_0 | 82 |
| q_{MAP} | Maximum a posteriori estimate | 157 |
| q_{MLE} | Maximum likelihood estimate | 84 |
| $Q = [Q_1, \dots, Q_p]$ | Random vector of parameters | 100 |
| Q | Orthogonal matrix in QR factorization | 118 |
| \mathbb{Q} | Admissible parameter space | 82 |
| \mathcal{Q} | Sample space | 82 |
| $Q^{(p)}$ | Quadrature operator in \mathbb{R}^p | 240 |
| r | Rank of matrix A | 117 |
| R, R_ℓ | Number of quadrature points | 243 |
| R | Upper triangular matrix in QR factorization | 118 |
| \hat{R}, R | Residual estimator and estimate | 136 |
| \mathcal{R} | Number of sparse grid quadrature points | 248 |
| $\mathcal{R}(u, q)$ | General observation or response | 63 |
| $\mathcal{R}(A)$ | Range of the matrix A | 116 |
| s_i | Local sensitivity indices | 192, 322 |
| S_i^σ | Sigma-normalized sensitivity indices | 322 |
| S_i, S_{ij}, S_{T_i} | Sobol sensitivity indices | 324 |
| SS_q | Sum of squares error | 156 |
| \mathcal{T} | Temporal domain | 63 |
| $u(q), u(t, x, q)$ | State variable | 61 |
| $\tilde{u}(q), \tilde{u}(t, x, q)$ | Surrogate state representation | 279 |
| V, V^J | Spaces of spatial test functions | 219 |
| V_k | Chain covariance matrix | 172 |
| w^r | Quadrature weights | 211 |
| X | Deterministic $n \times p$ design matrix | 131 |
| $\mathcal{X}(q)$ | Sensitivity matrix | 144 |
| y | Realizations of Y | 132 |
| Y | Random variable for model response | 321 |
| Z, Z^K | Spaces of parameter test functions | 219 |

Acronyms and Initialisms

| Term | Meaning | Page |
|--------|--|------|
| ANOVA | Analysis of variance | 291 |
| AR | Autoregressive (model) | 89 |
| ASAP | Adjoint sensitivity analysis procedure | 306 |
| BWR | Boiling water reactor | 36 |
| càdlàg | Continue à droite, limite à gauche | 69 |
| CASL | Consortium for Advanced Simulation of Light Water Reactors | 37 |
| cdf | Cumulative distribution function | 68 |
| CESM | Community Earth System Model | 30 |
| CFCs | chlorofluorocarbons | 25 |
| CRUD | Chalk River unidentified deposit | 43 |
| CVTs | Centroidal Voronoi tessellations | 285 |
| DAKOTA | Design Analysis Kit for Optimization and Terascale Applications | 236 |
| DOE | Department of Energy | 37 |
| DRAM | Delayed rejection adaptive Metropolis | 172 |
| DREAM | DiffeRential Evolution Adaptive Metropolis | 181 |
| ECMWF | European Centre for Medium-Range Weather Forecasts | 16 |
| FPGAs | Field programmable gate arrays | x |
| FSAP | Forward sensitivity analysis procedure | 306 |
| gcd | Greatest common divisor | 94 |
| GCR | Gas-cooled reactor | 36 |
| GP | Gaussian process | 89 |
| gPC | Generalized polynomial chaos | 207 |
| HDMR | High-dimensional model representation | 289 |
| HIV | Human immunodeficiency virus | 45 |
| iid | Independent and identically distributed | 79 |

| Term | Meaning | Page |
|-------|---|------|
| IPCC | Intergovernmental Panel on Climate Change | 32 |
| kde | Kernel density estimation | 75 |
| LANL | Los Alamos National Laboratory | 41 |
| LQR | Linear quadratic regulator | 50 |
| MAP | Maximum a posteriori (estimate) | 157 |
| MCMC | Markov chain Monte Carlo | 159 |
| MLE | Maximum likelihood estimate | 84 |
| NISP | Nonintrusive spectral projection | 225 |
| NWP | Numerical weather prediction | 16 |
| ODE | Ordinary differential equation | 51 |
| OLS | Ordinary least squares | 82 |
| ORNL | Oak Ridge National Laboratory | 41 |
| PC | Polynomial chaos | 207 |
| PCA | Principal component analysis | 109 |
| PDE | Partial differential equation | 51 |
| pdf | Probability density function | 69 |
| POD | Proper orthogonal decomposition | 285 |
| PRA | Probabilistic risk assessment | 44 |
| PWR | Pressurized water reactor | 36 |
| Q-Q | Quantile-quantile | 74 |
| QoI | Quantity of interest | 4 |
| SAMSI | Statistical and Applied Mathematical Sciences Institute | xi |
| SDE | Stochastic differential equation | 97 |
| SIR | Susceptible, infected, recovered (model) | 55 |
| SVD | Singular value decomposition | 117 |
| WMO | World Meteorological Organization | 17 |

Contents

| | |
|--|-------------|
| Preface | ix |
| Notation | xiii |
| Acronyms and Initialisms | xvii |
| 1 Introduction | 1 |
| 1.1 Nature of Uncertainties and Errors | 4 |
| 1.2 Predictive Estimation | 8 |
| 2 Large-Scale Applications | 11 |
| 2.1 Weather Models | 11 |
| 2.2 Climate Models | 21 |
| 2.3 Subsurface Hydrology and Geology | 33 |
| 2.4 Nuclear Reactor Design | 36 |
| 2.5 Biological Models | 44 |
| 3 Prototypical Models | 51 |
| 3.1 Models | 51 |
| 3.2 Evolution, Stationary, and Algebraic Models | 61 |
| 3.3 Abstract Modeling Framework | 63 |
| 3.4 Notation for Parameters and Inputs | 65 |
| 3.5 Exercises | 66 |
| 4 Fundamentals of Probability, Random Processes, and Statistics | 67 |
| 4.1 Random Variables, Distributions, and Densities | 67 |
| 4.2 Estimators, Estimates, and Sampling Distributions | 79 |
| 4.3 Ordinary Least Squares and Maximum Likelihood Estimators | 82 |
| 4.4 Modes of Convergence and Limit Theorems | 85 |
| 4.5 Random Processes | 87 |
| 4.6 Markov Chains | 90 |
| 4.7 Random versus Stochastic Differential Equations | 96 |
| 4.8 Statistical Inference | 98 |
| 4.9 Notes and References | 104 |
| 4.10 Exercises | 105 |

| | | |
|-----------|---|------------|
| 5 | Representation of Random Inputs | 107 |
| 5.1 | Mutually Independent Random Parameters | 107 |
| 5.2 | Correlated Random Parameters | 108 |
| 5.3 | Finite-Dimensional Representation of Random Coefficients . . . | 109 |
| 5.4 | Exercises | 112 |
| 6 | Parameter Selection Techniques | 113 |
| 6.1 | Linearly Parameterized Problems | 115 |
| 6.2 | Nonlinearly Parameterized Problems | 122 |
| 6.3 | Parameter Correlation versus Identifiability | 125 |
| 6.4 | Notes and References | 127 |
| 6.5 | Exercises | 128 |
| 7 | Frequentist Techniques for Parameter Estimation | 131 |
| 7.1 | Parameter Estimation from a Frequentist Perspective | 133 |
| 7.2 | Linear Regression | 134 |
| 7.3 | Nonlinear Parameter Estimation Problem | 141 |
| 7.4 | Notes and References | 152 |
| 7.5 | Exercises | 153 |
| 8 | Bayesian Techniques for Parameter Estimation | 155 |
| 8.1 | Parameter Estimation from a Bayesian Perspective | 155 |
| 8.2 | Markov Chain Monte Carlo (MCMC) Techniques | 159 |
| 8.3 | Metropolis and Metropolis–Hastings Algorithms | 159 |
| 8.4 | Stationary Distribution and Convergence Criteria | 168 |
| 8.5 | Parameter Identifiability | 171 |
| 8.6 | Delayed Rejection Adaptive Metropolis (DRAM) | 172 |
| 8.7 | DiffeRential Evolution Adaptive Metropolis (DREAM) | 181 |
| 8.8 | Notes and References | 184 |
| 8.9 | Exercises | 184 |
| 9 | Uncertainty Propagation in Models | 187 |
| 9.1 | Direct Evaluation for Linear Models | 188 |
| 9.2 | Sampling Methods | 191 |
| 9.3 | Perturbation Methods | 192 |
| 9.4 | Prediction Intervals | 197 |
| 9.5 | Notes and References | 203 |
| 9.6 | Exercises | 204 |
| 10 | Stochastic Spectral Methods | 207 |
| 10.1 | Spectral Representation of Random Processes | 207 |
| 10.2 | Galerkin, Collocation, and Discrete Projection Frameworks . . . | 214 |
| 10.3 | Stochastic Galerkin Method—Examples | 226 |
| 10.4 | Discrete Projection Method—Example | 234 |
| 10.5 | Stochastic Polynomial Packages | 235 |
| 10.6 | Exercises | 236 |