

## PRINCIPLES OF DATA ASSIMILATION

Data assimilation is theoretically founded on probability, statistics, control theory, information theory, linear algebra, and functional analysis. At the same time, data assimilation is a very practical subject, given its goal of estimating the posterior probability density function in realistic high-dimensional applications. This puts data assimilation at the intersection between the contrasting requirements of theory and practice. Based on over 20 years of teaching courses in data assimilation, *Principles of Data Assimilation* introduces a unique perspective that is firmly based on mathematical theories, but also acknowledges practical limitations of the theory. With the inclusion of numerous examples and practical case studies throughout, this new perspective will help students and researchers to competently interpret data assimilation results and to identify critical challenges of developing data assimilation algorithms. The benefit of information theory also introduces new pathways for further development, understanding, and improvement of data assimilation methods.

SEON KI PARK is Professor of Meteorology at Ewha Womans University, Seoul, Republic of Korea. His research focuses on storm-scale to mesoscale analysis, parameter estimation, and data assimilation to improve numerical weather and climate prediction. He coedited a series of four volumes titled *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications* (2009, 2013, 2017, 2022).

MILIJAZ ŽUPANSKI is Senior Research Scientist at Colorado State University, Fort Collins. He is a principal developer of two four-dimensional variational data assimilation systems and the maximum likelihood ensemble filter. His research focuses on data assimilation development and applications, including the atmosphere, land surface, aerosols, and combustion.

Cambridge University Press & Assessment  
978-1-108-83176-5 — Principles of Data Assimilation  
Seon Ki Park , Milija Zupanski  
Frontmatter  
[More Information](#)

---

# PRINCIPLES OF DATA ASSIMILATION

SEON KI PARK

*Ewha Womans University*

MILIJA ŽUPANSKI

*Colorado State University*



CAMBRIDGE  
UNIVERSITY PRESS



Shaftesbury Road, Cambridge CB2 8EA, United Kingdom  
One Liberty Plaza, 20th Floor, New York, NY 10006, USA  
477 Williamstown Road, Port Melbourne, VIC 3207, Australia  
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,  
New Delhi – 110025, India  
103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,  
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of  
education, learning and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781108831765](http://www.cambridge.org/9781108831765)

DOI: 10.1017/9781108924238

Cambridge University Press & Assessments © 2022

This publication is in copyright. Subject to statutory exception and to the provisions  
of relevant collective licensing agreements, no reproduction of any part may take  
place without the written permission of Cambridge University Press & Assessment.

First published 2022

*A catalogue record for this publication is available from the British Library.*

*Library of Congress Cataloging-in-Publication data*

Names: Park, Seon Ki, 1961– author. | Zupanski, Milija, 1955– author.

Title: Principles of data assimilation / Seon Ki Park, Ewha Womans  
University, Milija Zupanski, Colorado State University.

Description: New York : Cambridge University Press, 2022. |  
Includes bibliographical references and index.

Identifiers: LCCN 2022009537 (print) | LCCN 2022009538 (ebook) |  
ISBN 9781108831765 (hardback) | ISBN 9781108924238 (epub)

Subjects: LCSH: Meteorology–Data processing. | Meteorology–Observations.

Classification: LCC QC874.3 .P37 2022 (print) | LCC QC874.3 (ebook) |  
DDC 551.630285/63–dc23/eng20220517

LC record available at <https://lcn.loc.gov/2022009537>

LC ebook record available at <https://lcn.loc.gov/2022009538>

ISBN 978-1-108-83176-5 Hardback

Cambridge University Press & Assessment has no responsibility for the persistence  
or accuracy of URLs for external or third-party internet websites referred to in this  
publication and does not guarantee that any content on such websites is, or will  
remain, accurate or appropriate.

*To the memory of our mentor,*

Yoshi K. Sasaki,

and

*to our families,*

Hyunmi, Jun Bum, Shinwon and Kyung Keun;

Dusanka and Sandra

Cambridge University Press & Assessment  
978-1-108-83176-5 — Principles of Data Assimilation  
Seon Ki Park , Milija Zupanski  
Frontmatter

[More Information](#)

---

## Contents

<i>Preface</i>	<i>page xi</i>
<b>Part I General Background</b>	<b>1</b>
1 Data Assimilation: General Background	3
1.1 Introduction	3
1.2 Historical Background	5
1.3 Terminologies and Notation	9
1.4 Basic Estimation Problem	16
1.5 Optimal Interpolation	27
2 Probability and the Bayesian Approach	35
2.1 Introduction	35
2.2 Probability	35
2.3 Probability Density Function Moments	37
2.4 Common Probability Density Functions	38
2.5 Bayes' Formula	40
2.6 Recursive Bayes' Formula for Data Assimilation	41
2.7 Uncertainty and Probability: Shannon Entropy	42
2.8 Data Assimilation with Gaussian Probability Density Functions	46
2.9 Overview	48
3 Filters and Smoothers	54
3.1 Introduction	54
3.2 Discrete Model and Observations	55
3.3 Filters	56
3.4 Smoothers	67
3.5 Bayesian Perspectives	74

<b>Part II Practical Tools</b>	81
4 Tangent Linear and Adjoint Models	83
4.1 Introduction	83
4.2 Taylor Series Approximation	84
4.3 Tangent Linear Model	87
4.4 Adjoint Model	90
4.5 Uncertainty and Validity of the TLM/ADJM	95
4.6 Example: Continuous ADJM from Burgers' Equation	99
4.7 Continuous versus Discrete ADJM	103
4.8 General Rules of Constructing the TLM/ADJM	105
5 Automatic Differentiation	111
5.1 Introduction	111
5.2 Background on Automatic Differentiation	112
5.3 Forward versus Reverse Mode of Differentiation	116
5.4 Automatic Differentiation Approaches and Tools	120
6 Numerical Minimization Process	128
6.1 Introduction	128
6.2 Mathematical Background on Minimization	129
6.3 Algorithms for Minimization	138
<b>Part III Methods and Issues</b>	161
7 Variational Data Assimilation	163
7.1 Introduction	163
7.2 Variational Data Assimilation as an Optimal Control Problem	170
7.3 Elements of Variational Data Assimilation	172
7.4 From 3DVAR to 4DVAR	191
8 Ensemble and Hybrid Data Assimilation	210
8.1 Introduction	210
8.2 Role of Forecast Error Covariance	211
8.3 Ensemble Data Assimilation	216
8.4 Hybrid Data Assimilation	223
9 Coupled Data Assimilation	232
9.1 Introduction	232
9.2 Coupled Modeling System	233
9.3 Classification Based on Coupling Strength	233
9.4 Challenges	237



*Contents*

ix

9.5	One-Point, Two-Component Coupled Data Assimilation System	241
9.6	Structure of Coupled Forecast Error Covariance	243
9.7	Summary and Future	246
10	Dynamics and Data Assimilation	251
10.1	Introduction	251
10.2	Probabilistic Prediction	251
10.3	Error Growth and the Lyapunov Exponent	252
10.4	Recursive Data Assimilation	255
10.5	Addressing Dynamical Imbalances in the Analysis	259
<b>Part IV Applications</b>		263
11	Sensitivity Analysis and Adaptive Observation	265
11.1	Introduction	265
11.2	Deterministic Sensitivity Analysis	267
11.3	Adaptive Observations	280
12	Satellite Data Assimilation	305
12.1	Current Status	305
12.2	All-sky Radiance Assimilation Challenges	308
12.3	Summary and Future	323
<b>Part V Appendices</b>		331
<i>Appendix A</i> Linear Algebra and Functional Analysis		333
A.1	Introduction	333
A.2	Vector Space and Functional Analysis	333
A.3	Covariance and Correlation Matrices	340
A.4	State Vector, Control Vector, and Uncertainty	341
A.5	Overview and Examples	343
<i>Appendix B</i> Discretization of Partial Differential Equations		347
B.1	Grid-point System	347
B.2	Finite Difference Methods	347
B.3	Example – 1D Advection Equation	349
<i>Appendix C</i> Lab Practice I		353
C.1	Logistic Map	353
C.2	Development of the Tangent Linear Model and the Adjoint Model	353
C.3	Tangent Linear Model/Adjoint Model Codes by Hand	354

C.4	Tangent Linear Model/Adjoint Model Codes Using an Automatic Differentiation Tool	355
C.5	Suggested Experiments	358
<i>Appendix D</i>	Lab Practice II	361
D.1	Data Assimilation Terminology	361
D.2	Data Assimilation Steps	362
D.3	Other Data Assimilation Considerations	368
D.4	Data Assimilation Code with a 1D Burgers' Equation	370
<i>Index</i>		384

## Preface

Data assimilation has become a major component of today's numerical weather prediction (NWP). It followed the development of numerical modeling in 1950s–1970s, mostly in terms of estimating the initial conditions. This is not surprising given that NWP models are a system of partial differential equations (PDEs). However, in later applications, in particular with climate models, data assimilation has included parameter estimation, either standalone or in addition to initial conditions. From the 1990s the role of data assimilation in NWP has been steadily increasing, closely related to the advancement of numerical modeling with ever increasing complexity. In addition to meteorology, data assimilation has also been getting more attention within other disciplines. One may follow the number of published books addressing data assimilation and notice an almost exponential increase starting with only a couple of books in the 1990s, to probably 10–15 books in the 2010s. This indirectly suggests that data assimilation has transformed from obscure to mainstream over the last 30 years.

The relevance of data assimilation to the big data paradigm also increases, since realistic data assimilation is a method of processing large data sets and modeling of physical phenomena requires powerful high-performance computing. The interest in estimating the uncertainty of these data is also common to big data science and data assimilation. Our approach in this book is to relate data assimilation to big data and high-dimensional problems. While there is a substantial and critical data assimilation development suitable for low-dimensional applications, we are mostly interested in practical data assimilation methods that have a proven utility and/or potential for realistic high-dimensional applications. While in some interpretations "practical" may imply favoring the solutions that work in practice but only loosely follow mathematical reasoning and proofs, we are trying to rather strictly follow the mathematical methods that can work in practice. Although this may appear as a play on words, we believe that closely following mathematical formalism is fundamental for further progress of data assimilation. Adopting such an approach does not necessarily imply that this is always possible, but that all attempts should be made to satisfy mathematical requirements and to never be satisfied with an ad hoc solution that may work in practice.

In addition to big data, one should not dismiss the relevance of artificial intelligence (AI) for data assimilation. This is quite an important new area of research and development that is also based on big data and high-performance computing, and is currently seen as a

major new step in data assimilation development. However, we decided not to include the applications of AI in data assimilation because it is quite an intense area of development with fast-changing applications that may be practically impossible to present fairly without missing something important. Current AI applications are very broad and include the following: observation operators, derived observation products, data thinning, bias and model error corrections, as well as model dynamics. New areas may be added as we write this. In an eventual further edition of this book, we will likely consider adding a chapter on AI in data assimilation since by that time it will hopefully become clear what main areas of data assimilation can benefit from AI.

As will be seen in the book, data assimilation is multidisciplinary and it is closely related to applications that may be known under different names in different disciplines. For example, data assimilation can be interpreted as an inverse problem with applicability in optics, acoustics, communication and signal processing, medical imaging, geophysics, astronomy, and other disciplines.

This book originated from the tutorials, presentations, and exercises that the authors and their colleagues conducted during the five years of the *Ewha International School of Data Assimilation (EISDA 2012–2016)*, which gathered students and young researchers mostly from Asia and Africa, as well as the graduate-level courses taught by the authors during last 20 years or so at Ewha Womans University and Colorado State University, respectively. Therefore, this book is also written for use by students and scientists interested in applying data assimilation in realistic high-dimensional problems. Some mathematical aspects may be too advanced for undergraduate students and may require some teacher assistance. Given the focus on methods for high-dimensional applications, this book can be of help to users and developers of data assimilation. Although meteorological applications dominate, we believe that the topics presented can also benefit researchers from various other disciplines.

We begin by presenting a general background of data assimilation with a brief historical overview and basic estimation problem, and by introducing common terminologies and notation in Chapter 1. Chapter 2 introduces the basic formalism of data assimilation through probability and the Bayesian paradigm. Fundamental data assimilation algorithms, filters, and smoothers, are presented in Chapter 3. Mathematical and practical details of tangent linear and adjoint models, which are fundamental for understanding variational and hybrid ensemble-variational data assimilation methods, are given in Chapter 4. Their practical implementation and development is closely related to automatic differentiation described in Chapter 5. Numerical minimization algorithms that are fundamental for applying data assimilation in realistic nonlinear problems are described in Chapter 6. In Chapter 7, variational methods, still the workhorse in NWP operational data assimilation, are introduced in various forms along with the optimal control theory. Recent methods of ensemble and hybrid data assimilation are presented in Chapter 8. Coupled data assimilation that is applied to coupled modeling systems is described in Chapter 9. The relation between dynamics and data assimilation, quite important but sometimes neglected given the dominant role of observations, is presented in Chapter 10. Some special topics, such as sensitivity analysis, and adaptive observations in Chapter 11 and satellite data assimilation in Chapter 12, are also discussed given their relevance in numerous applications. The Appendices are

composed in order to introduce relevant tools and methods that are inherent in data assimilation, such as linear algebra and functional analysis in Appendix A, discretization of PDEs in Appendix B, and the lab practices aimed at understanding the intricate details of a practical data assimilation algorithm in Appendices C and D.

Our gratitude is extended to all our students and colleagues that contributed, directly or indirectly, in developing the material for this book. In particular, we would like to thank our colleagues, Liang Xu and Adrian Sandu, who delivered excellent lectures in EISDA. We also appreciate the Ewha graduate students who had attended EISDA and have been the main resources of our long-time and ongoing collaborative research: Sujeong Lim, Ebony Lee, Ji Won Yoon, Seung Yeon Lee, Yoonjin Lee, Sojung Park, Seungwon Chung, and Mariam Hussain.

Cambridge University Press & Assessment  
978-1-108-83176-5 — Principles of Data Assimilation  
Seon Ki Park , Milija Zupanski  
Frontmatter  
[More Information](#)

---