Statistics for Chemical Engineers

Build a firm foundation for studying statistical modeling, data science, and machine learning with this practical introduction to statistics, written with chemical engineers in mind.

Key Features

- Introduces a data–model–decision approach to applying statistical methods to real-world chemical engineering challenges.
- Establishes links between statistics, probability, linear algebra, calculus, and optimization, covering classical and modern topics of statistics such as uncertainty quantification, risk modeling, and decision-making under uncertainty.
- Over 100 worked examples using MATLAB and Python demonstrate how to apply theory to practice, with over 70 end-of-chapter problems to reinforce student learning.
- Introduces key topics using a modular structure, which supports learning at a range of paces and levels.

Requiring only a basic understanding of calculus and linear algebra, this textbook is the ideal introduction for undergraduate students in chemical engineering and a valuable preparatory text for advanced courses in data science and machine learning with chemical engineering applications.

Victor M. Zavala is the Baldovin-DaPra Professor of Chemical and Biological Engineering at the University of Wisconsin–Madison and a senior computational mathematician at Argonne National Laboratory. He is the recipient of the Harvey Spangler Award for Innovative Teaching and Learning Practices from the College of Engineering at the University of Wisconsin–Madison and of the Presidential Early Career Award for Scientists and Engineers (PECASE).

Cambridge Series in Chemical Engineering

Books in the Series
Baldea and Daoutidis, *Dynamics and Nonlinear Control of Integrated Process Systems*
Chamberlin, *Radioactive Aerosols*
Chau, *Process Control: A First Course with Matlab*
Cussler, *Diffusion: Mass Transfer in Fluid Systems, Third Edition*
Cussler and Moggridge, *Chemical Product Design, Second Edition*
De Pablo and Schieber, *Molecular Engineering Thermodynamics*
Deen, *Introduction to Chemical Engineering Fluid Mechanics*
Denn, *Chemical Engineering: An Introduction*
Denn, *Polymer Melt Processing: Foundations in Fluid Mechanics and Heat Transfer*
Dorfman and Daoutidis, *Numerical Methods with Chemical Engineering Applications*
Duncan and Reimer, *Chemical Engineering Design and Analysis: An Introduction 2E*
Fan, *Chemical Looping Partial Oxidation Gasification, Reforming, and Chemical Syntheses*
Fan and Zhu, *Principles of Gas-Solid Flows*
Fox, *Computational Models for Turbulent Reacting Flows*
Franses, *Thermodynamics with Chemical Engineering Applications*
Grossmann, *Advanced Optimization for Process Systems Engineering*
Leal, *Advanced Transport Phenomena: Fluid Mechanics and Convective Transport Processes*
Lim and Shin, *Fed-Batch Cultures: Principles and Applications of Semi-Batch Bioreactors*
Litster, *Design and Processing of Particulate Products*
Maravelias, *Chemical Production Scheduling*
Marchisio and Fox, *Computational Models for Polydisperse Particulate and Multiphase Systems*
Mewis and Wagner, *Colloidal Suspension Rheology*
Morbidelli, Gavriilidis, and Varma, *Catalyst Design: Optimal Distribution of Catalyst in Pellets, Reactors, and Membranes*
Nicoud, *Chromatographic Processes*
Noble and Terry, *Principles of Chemical Separations with Environmental Applications*
Orbey and Sandler, *Modeling Vapor-Liquid Equilibria: Cubic Equations of State and their Mixing Rules*
Petyluk, *Distillation Theory and its Applications to Optimal Design of Separation Units*
Pfister, Nicoud, and Morbidelli, *Continuous Biopharmaceutical Processes:* Chromatography, Bioconjugation, and Protein Stability
Ramkrishna and Song, *Cybernetic Modeling for Bioreaction Engineering*
Rao and Nott, *An Introduction to Granular Flow*
Russell, Robinson, and Wagner, *Mass and Heat Transfer: Analysis of Mass Contactors and Heat Exchangers*
Schobert, *Chemistry of Fossil Fuels and Biofuels*
Shell, *Thermodynamics and Statistical Mechanics*
Sirkar, *Separation of Molecules, Macromolecules and Particles: Principles, Phenomena and Processes*
Slattery, *Advanced Transport Phenomena*
Subramaniam, *Green Catalysis and Reaction Engineering: An Integrated Approach with Industrial Case Studies*
Varma, Morbidelli, and Wu, *Parametric Sensitivity in Chemical Systems*
Vassiliadis et al., *Optimization for Chemical and Biochemical Engineering*
Weatherley, *Intensification of Liquid–Liquid Processes*
Wolf, Bielser, and Morbidelli, *Perfusion Cell Culture Processes for Biopharmaceuticals*
Zavala, *Statistics for Chemical Engineers*
Zhu, Fan, and Yu, *Dynamics of Multiphase Flows*

Statistics for Chemical Engineers

From Data to Models to Decisions

Victor M. Zavala
University of Wisconsin–Madison

CAMBRIDGE
UNIVERSITY PRESS

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
Information on this title: www.cambridge.org/highereducation/isbn/9781009541893

DOI: 10.1017/9781009541916

When citing this work, please include a reference to the DOI 10.1017/9781009541916

First published 2026

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

A Cataloging-in-Publication data record for this book is available from the Library of Congress

ISBN 978-1-009-54189-3 Hardback

Additional resources for this publication at www.cambridge.org/zavala

Cover image: iStock / Getty Images aleksey_rezin

To Katie, Owen, and Frankie
To Irma, Karla, and Victor
To all my students, friends, and collaborators

Contents

Preface

Background and Motivation

This book aims to *change the way we think about statistics* and *teach statistics* in chemical engineering. Statistics is one of the pillars of modern science and engineering and of emerging topics such as data science and machine learning; despite this, its scope and relevance has remained misunderstood and underappreciated in chemical engineering education (and in engineering education at large). Specifically, statistics is often taught by placing emphasis on *data analysis*. I would like to convince the reader that statistics is much more than that; particularly, statistics is a *mathematical modeling* paradigm that complements physical modeling paradigms used in chemical engineering (e.g., thermodynamics, transport phenomena, conservation, and reaction kinetics). Statistics, in particular, can help model random phenomena that might not be predictable from physics alone (or from deterministic physical laws), quantify the uncertainty of predictions obtained with physical models, discover physical models from data, and create models directly from data (in the absence of physical knowledge).

The *fusion* of statistical and physical modeling paradigms provides a powerful framework for analyzing data, extracting knowledge from data, and making informed predictions and decisions. In addition, I would like to convince the reader that statistics is a *foundational* topic in that it *fundamentally transforms the way* we perceive the world and it touches every aspect of chemical engineering. More broadly, I would like to convince the reader that statistics provides a *conceptual framework* that can help chemical engineers do what they do best: abstract, understand, design, and control *complex systems*.

The desire to write this book came from my personal experience in learning statistics in college and in identifying the significant gaps in my understanding of statistics throughout my professional career. Similar feelings are often shared with me by professionals working in industry and academia. Like many chemical engineers, I took a course in statistics in college that covered classical topics such as random variables, descriptive statistics, regression, design of experiments, and basic probability. This course, while I found it to be interesting, felt disconnected from the rest of the chemical engineering curriculum. Specifically, with the exception of linear regression and design of experiments, I did not encounter other major uses of statistics in the curriculum. This left me with a perception that statistics was an intellectual “curiosity.” Throughout my professional career, I have been exposed to a broad range of applications in which knowledge of statistics has proven to be essential: uncertainty quantification, quality control, risk assessment, modeling of random phenomena, process monitoring, forecasting, machine learning, computer vision, and decision-making under uncertainty. These are applications that are pervasive in industry and academia. It is also important to recognize that the field of statistics continues to evolve and many new concepts/tools have become available; for example, the fields of uncertainty quantification, Bayesian analysis, statistical learning, and decision-making under uncertainty have experienced significant growth in recent years.

I believe that *there is a need to modernize the scope and approach to teach statistics.* This should be done carefully by finding a suitable blend of classical and new topics, finding points of connection with the rest of the chemical engineering curriculum, and thinking about what are the unique skills and interests of chemical engineers. As such, a pedagogical question that motivates this book is:

> *How can we better teach statistics to chemical engineers?*

Statistics is typically taught by mathematics/statistics departments to a broad range of engineers; as such, it might be difficult for instructors to standardize the teaching material and make explicit connections to applications that are of interest to chemical engineers. Specifically, it is important to understand that chemical engineers are trained to use *physical knowledge* (e.g., thermodynamics, reaction kinetics, and transport phenomena) to analyze systems and make decisions (e.g., design an experiment or a chemical process). As such, when teaching statistics, it is important to emphasize when and how these tools can complement (or substitute) physics. Moreover, it is important to remember that a *key skill of chemical engineers* is the ability to develop mathematical abstractions (models) to analyze complex systems, and by *complex*, I mean systems that involve heterogeneous phenomena (e.g., reactions, flows, heating/cooling, and separation) and that might involve different scales (e.g., molecular level, unit level, process level, and enterprise level). As such, when teaching statistics, it is important to emphasize how this can provide tools to facilitate the modeling/understanding of complex systems. It is also important to remember that, when chemical engineers analyze data, they are ultimately interested in understanding phenomena and extracting underlying principles; in other words, engineers aim to attribute physical meaning/origin of behavior that helps them make decisions (e.g., design a new material or microbe to conduct a function). As such, when teaching statistics, it is important not to discard the physical and decision-making context. Finally, with the advent of machine learning and data science, it is important to remember that statistics (together with calculus and linear algebra) provides key mathematical fundamentals that are the *building blocks* of these tools. As such, when teaching statistics, it is important to explain how foundations can be used to develop or select tools and to analyze the outcomes of such tools.

Statistics is typically taught early in the chemical engineering curriculum, together with other mathematics courses such as calculus, linear algebra, and numerical methods. Teaching statistics early in the curriculum is important, as it provides the foundations and tools that are needed in other courses. However, if taught too early (e.g., sophomore year), instructors will be limited in the content that they can cover, as students do not yet have the proper foundations to cover advanced topics. For example, if students have not taken linear algebra or multivariate calculus, it will be difficult for them to understand principal component analysis and parameter estimation. Moreover, if statistics is taught too late (e.g., senior year), it will be difficult for students to appreciate the connections between statistics and the rest of the curriculum. I thus think that a suitable timing to teach statistics is in the junior year, or alternatively, one could spread the course into smaller modules that are taught in different years. Teaching statistics in the junior year has several *strategic benefits*; for instance, given that statistics makes extensive use of linear algebra, numerical methods, and calculus, it can provide a venue to reinforce these concepts. For instance, data science and machine learning can help students better understand and appreciate the relevance of all the mathematical concepts learned so far in the curriculum. Moreover, by teaching statistics in the junior year, students will be in a better position to use statistical tools in core courses (e.g., parameter estimation in reactor engineering, data analysis in laboratories, uncertainty quantification in process design, and data-driven modeling in process control). Moreover, by teaching statistics at the junior level, students can begin to question the origin and quality of data and models (e.g., empirical vs. first-principles); this can help reinforce the understanding of physical principles and can help them appreciate aspects of complex systems that are less/more difficult to predict.

Design and Organization

The *design of this book* follows a *data–models–decisions* pipeline. The intent of this design is to emphasize that statistics is a modeling paradigm that maps data to models and models to decisions; this design also aims to "connect the dots" between different branches of statistics and engineering. The focus on the pipeline is also important in reminding the reader that understanding the application context matters. For instance, the data and type of model used for process design can be quite different from the data and type of model used for experimental design. Similarly, the nature of the decision and the data available influence the type of model used. The book design is also intended for the reader to understand the close interplay between statistical and physical modeling; specifically, we emphasize how statistics provides tools to model aspects of a system that cannot be fully predicted from physics. Moreover, we emphasize how physical systems might naturally exhibit random behavior due to uncontrollable aspects (e.g., defects of materials, noise, fluctuations, and vibrations).

The book design is also intended to help the reader appreciate how statistics provides an *important foundation* for a broad range of modern tools of data science and machine learning (e.g., neural networks and logistic regression). More broadly, the book emphasizes how statistics provides a way to think about the world. For instance, we discuss how statistical thinking is fundamentally different from deterministic thinking (which ignores uncertainty). Making this distinction is extremely important, as chemical engineering courses tend to follow a *deterministic mindset* (e.g., there is no uncertainty/variability in the data and the model used for analysis is perfect). In this context, the book discusses how ambiguity can arise when one faces uncertainty, as key variables of interest (e.g., cost and carbon emissions) are no longer numbers (they are distributions) and thus cannot be compared so easily. Moreover, we discuss how statistics provides tools that can help make decisions that *mitigate/control uncertainty*.

The design of the book is also intended to *reinforce mathematical fundamentals (linear algebra, optimization, and calculus)* that are found in a broad range of applications; I believe that statistics and data science provide an excellent framework for doing this, as one can relate mathematical concepts (e.g., eigenvalues) to statistical concepts (e.g., information content). Statistics also provides a suitable framework because it connects different mathematical concepts that are often treated separately (e.g., estimation connects linear algebra and optimization).

The book content is structured in three major modules/parts: **Part I:** Basic Concepts (Chapters 1–3), **Part II:** Intermediate Concepts (Chapters 4–5), and **Part III:** Advanced Concepts (Chapters 6–7). The chapter contents can be summarized as follows:

- **Chapter 1** provides a *"big picture" introduction* to statistics by following the *data–model–decisions pipeline*. This aims to explain how statistics provides tools to use data to model uncertainty and variability, understand how uncertainty propagates through systems, and make decisions that help mitigate and control the effect of uncertainty. This chapter also introduces fundamental concepts of random variables; this discussion emphasizes how a random variable is a model (a theoretical model) that can be used to explain actual data of a system that we observe. Moreover, this discussion introduces key elements of random variable models, such as probability density functions, domains, parameters, and summarizing statistics (e.g., mean, variance, quantiles, and probability of loss). This chapter also discusses *structural models*, which are models that capture systematic relationships between variables.
- **Chapter 2** discusses how to *characterize uncertainty using univariate (single-variable) random variable models*. Here, we highlight the properties that different models have, what type of behavior they aim to model, and when it is appropriate to use them. Moreover, we emphasize connections between models and behavior that is typically observed in physical systems (e.g., due to aspects that cannot be fully controlled or predicted). This chapter also highlights connections between random variable models (e.g., some models are generalizations or special cases of other models). This discussion aims to highlight how the selection of statistical models is directly analogous to the selection of physical models (e.g., equations

of state or reaction kinetic models) in that one needs to understand the assumptions that each model makes and under what circumstances they can be (or not) used.

- **Chapter 3** discusses how to *model uncertainty using multivariate random variable models*. Here, we highlight concepts of correlation and covariance and of joint and conditional probabilities that help explain how random variables are interrelated. This discussion is intended for the reader to understand how connections between random variables can be used to gain understanding of a complex system, develop models, and make predictions. This chapter will also show how to compute probabilities of complex events and how events can be used to model decision-making logic. This chapter also highlights that there are few multivariate models available (e.g., multivariate Gaussian) but that one can still conduct useful analysis based on data alone.
- **Chapter 4** discusses how to *estimate the parameters for random variable models* from data using a variety of techniques. This chapter also shows how to compute estimates for theoretical quantities of random variables (e.g., expected value, variance, and probabilities) from available data and how to quantify the accuracy of such estimates. This discussion will lead us to important computational tools such as Monte Carlo approximations and important asymptotic properties (e.g., the law of large numbers, the central limit theorem, and the extreme value theorem) that explain how estimates behave as we accumulate data.
- **Chapter 5** discusses how to *estimate parameters for structural models* from data. Here, we will highlight how structural models exploit interconnections between variables to make predictions and how estimation techniques aim to separate the structural form of the model (the "trend") from inherent random effects present in the data (e.g., sensor noise). This discussion will also show how one can use random variable models to quantify the uncertainty of structural model predictions. This chapter discusses a powerful, optimization-based estimation technique known as maximum likelihood estimation in depth and applies it to linear and nonlinear models. This chapter also introduces the notion of information content and highlights how information is inherently linked to data availability and quality and how it can be quantified using optimization and linear algebra techniques. We also discuss the importance of incorporating prior knowledge in the estimation procedure (in the form of constraints and regularization terms) in order to obtain better parameter estimates and avoid spurious behavior (e.g., overfitting or predictions that do not obey physics). The chapter closes with a discussion of Bayesian estimation, which provides an alternative paradigm to maximum likelihood estimation that can help capture a broader set of settings (e.g., reconcile data to physics).
- **Chapter 6** discusses *advanced tools of statistical data analysis and learning*. Here, the goal is to understand how to apply the fundamentals of statistics from previous chapters to analyze complex datasets and to build sophisticated predictive models. This chapter explains how to use different types of techniques (eigenvalue decompositions, convolutions, and Fourier transforms) to

extract information from different types of data (e.g., images and time series). In addition, the chapter discusses advanced machine learning models (logistic regression, kernel models, and neural networks) and shows how to use statistics to help explain their design and behavior.

- **Chapter 7** discusses how to *make decisions in the face of uncertainty*. Here, we discuss how to model different attitudes toward risk, how to compare decisions, and how to obtain optimal decisions. Specifically, we will see that making decisions under uncertainty is challenging because variables of interest (e.g., cost) are no longer numbers but are random variables; as such, selecting a suitable decision requires the comparison of the probability density functions or of summarizing statistics. We will also introduce a technique known as *stochastic optimization* that aims to make decisions that shape the distribution of variables of interest in desirable ways. An important goal of this discussion will be to convince the reader that *deterministic* decision-making paradigms can lead to decisions that are vulnerable/non-robust when facing different types of circumstances. Finally, the book closes with a discussion on *Bayesian optimization*, which is a powerful technique that combines diverse topics covered in the book. Specifically, we will see that this technique is a *closed-loop learning* paradigm that aims to strategically collect data to learn (to develop a predictive model) and to make decisions. This learning paradigm mimics how humans and living systems naturally leverage data (e.g., collected from our sensory systems) to improve predictions and decisions.

The book is intended to provide a framework to develop a *statistical modeling* course that complements physical modeling courses covered in chemical engineering education. Along these lines, the book aims to place *statistical modeling at the core of chemical engineering*. Specifically, it is essential for engineers to model random phenomena and understand the origins of such phenomena, properly quantify uncertainty and risk of models/predictions when making critical decisions, and understand what data is most useful (or not useful) when making a decision. This is becoming increasingly relevant as societal problems become increasingly complex and higher volumes of data become available (e.g., better sensors and instrumentation are becoming available). It is also important to highlight that this book differs from other engineering statistics books in that it places emphasis on modeling and chemical engineering applications; moreover, the book covers classical and modern topics of statistics and places emphasis on mathematical fundamentals.

Audience

The book is primarily intended for undergraduate chemical engineers (ideally at the junior level) who have a basic background in calculus, linear algebra, and programming. The content of the book can be covered in a single semester (by adjusting the level of depth/difficulty in each topic), but the book has been designed in a modular form so that the content can be spread throughout the curriculum: sophomore year

(Part I), junior year (Part II), and senior year (Part III). The content should also be useful for undergraduate statistics courses in other engineering fields (e.g., mechanical, electrical, industrial, and biomedical engineering). Some advanced content can also be used in graduate courses (primarily for students who seek to reinforce their knowledge of statistics and mathematical fundamentals) and for developing short courses for industrial practitioners.

Approach

In discussing the different topics, we will make special emphasis on the *key role that data plays* in helping us select appropriate models and make decisions. In this context, we emphasize how to determine if sufficient and good-quality data is available to develop models and make decisions; moreover, we will discuss how to best select data to conduct different types of functions (e.g., estimation vs. optimization). We also place emphasis on Monte Carlo simulations, which provide a general framework to use data to quantify uncertainty and make decisions. In discussing the different topics, we will also make emphasis on the deep connections that exist between statistics and other mathematical fields such as calculus (e.g., optimization, integration, and differential equations) and linear algebra (e.g., vectors, matrices, and eigenvalues). This will highlight how statistics, data science, and machine learning provide an excellent venue to reinforce and understand mathematical concepts. For instance, we discuss how second derivatives allow us to measure information content and how eigenvalues can help reveal redundancies in data. Reinforcing concepts of optimization and linear algebra is particularly important, as such concepts are broadly applicable in science and engineering.

We provide examples to connect all the topics discussed with a broad range of subjects of interest in chemical engineering. These examples are accompanied by their *software implementation*, which can help the reader understand how to implement the concepts in a practical setting and can help reinforce knowledge of computer programming and numerical computations. It is important to recognize that statistics, when seen from a perspective of modeling, can become rather mathematical in nature; as such, having a blend of theory and computational views is important. The book has a mathematical feel; this was both intentional and necessary; one of the reasons for this is to ensure that the reader understands the mathematics behind the software tools available. This mathematical understanding is necessary to interpret results, understand the applicability of specific tools, and facilitate troubleshooting.

A broad range of exercise problems are provided to reinforce the concepts and help show the reader how to apply statistical tools to solve realistic problems; some of these problems have been derived from real data and settings (in some instances based on research by our research group). The selection of examples and exercises is biased by my own experience, but I am hoping that the principles taught are broadly applicable.

People often ask me if textbooks (such as this one) should continue to be used in education, now that we have access to myriad online resources. In my opinion, a textbook conveys not only knowledge but also a thought process that some students (and instructors) can find beneficial. Moreover, a textbook aims to present a coherent vision of a subject (something difficult to do by simple compilation of resources); in other words, a textbook aims to "connect the dots" between topics/subjects. The book is designed with this in mind; it aims to present an end-to-end and coherent vision on how statistical principles can be used for a wide variety of applications.

Statistics is a broad area of mathematics, and it is inevitable that some topics needed to be left out. For instance, this book does not cover topics of stochastic processes (e.g., Markov chains and time series), statistical mechanics, and information theory. Moreover, the book only covers basic topics of machine learning and data science. However, I am hoping that the book provides the necessary foundations to explore these more advanced topics; again, the intention of this book is to connect the dots between different topics (as opposed to going deep into each topic).

Online Resources

The book incorporates documented code for all examples. A solution manual and code for exercises are also available for instructors. The availability of this code is intended for students to quickly see how to apply abstract concepts and develop "computational thinking" needed for actual implementation.

Acknowledgments

There are many people that helped me put together this book. First and foremost, I want to thank all my undergraduate and graduate students at the University of Wisconsin–Madison; the design of this book was largely inspired by the many things that I have learned while teaching different courses in chemical engineering.

I want to thank all my current and former PhD trainees and postdocs (particularly David Cole, Leo Gonzalez, Ugo Ikegwu, Lisa Je, Bruce Jiang, Kibaek Kim, Jiaze Ma, Ashley McCullough, Amy Qin, Alex Smith, Jaron Thompson, Elvis Umana, Bo-Xun Wang, and Weiqi Zhang) for all their help in proofreading earlier versions of the book and for their help in developing exercise problems and example code. Many of the concepts and problems discussed in this book are inspired by the research conducted at the Scalable Systems Laboratory (a.k.a. ZavaLab) at the University of Wisconsin–Madison.

The design of this book has been inspired by the work of Mihai Anitescu, Larry Biegler, Richard Braatz, Ignacio Grossmann, Joe Qin, Tunde Ogunnaike, and Jim Rawlings (and so many others). I have learned so much from you about statistics, optimization, data science, and linear algebra, and I hope that this is well captured in the contents of this book.

I want to thank my colleagues in the Department of Chemical and Biological Engineering at the University of Wisconsin–Madison and the Mathematics and Computer Science Division at Argonne National Laboratory for creating collaborative environments that have helped me appreciate the many applications of statistics and data science and for creating an environment that welcomes and fosters textbook writing. I dedicate this book to the legendary textbook writers of Wisconsin and Argonne, who have been a great source of inspiration: Bird, Box, Hill, Hougen, Lightfoot, Mangasarian, Moré, Rawlings, Stewart, and Wright. Special thanks go to Mike Graham, Dan Klingenberg, and Thatcher Root, who encouraged and supported me in developing a statistics course and book targeted at chemical engineers in Wisconsin. I also would like to acknowledge financial support from the Hougen Fellowship, which was instrumental in allocating time for writing a significant portion of this book.

I would like to thank my family (particularly Katie Zavala) for supporting me throughout the development of this book; this has been one of the most challenging projects that I have pursued in my career. I have spent many hours (often during weekends) writing and polishing this book.

Funding Support

I acknowledge financial support from the Department of Chemical and Biological Engineering at the University of Wisconsin–Madison under the Hougen program; this program provided funding for a research leave under which significant portions of this book were developed and written. I acknowledge partial financial support from the US National Science Foundation (NSF) under grants CBET-1748516 and IIS-1837812. The design of this book was motivated by research conducted under an NSF CAREER grant (CBET-1748516), and a significant portion of the technical content is inspired and based on research conducted under an NSF BIGDATA grant (IIS-1837812). I acknowledge partial support from the US Department of Energy, Office of Science, Advanced Scientific Computing Research, under contract number DE-AC02-06CH11357. This funding contributed toward my sabbatical year at Argonne National Laboratory in which parts of this book were developed and written.

Sources and Permissions

The following are data/content sources and associated permissions that I would like to explicitly acknowledge.

- The equilibrium Gibbs reactor model that is used to motivate diverse examples and exercises in the book is an adaptation of the equilibrium model presented by Seider, Seader, Lewin, and Widagdo [1, chapter 7].
- The problem and data of Exercise 2.9 (sieve analysis) were based on the content reported in [2].
- The problem of Exercise 2.7 (stress failure) uses data reported in [3]. The problem is also inspired by the tutorial notes reported in [4].
- The problem of Exercise 2.11 (residence time distribution) is an adaptation of the example problem reported by Professor Scott Fogler in his textbook [5].
- Exercise 2.1 was inspired by the experiment reported in [6].
- The flow cytometry examples and exercises in Chapters 3 and 6 are based on research reported in [7]. The data obtained from this research is used with permission from the authors. I thank Bruce Jiang for his help in putting together these examples/exercises; I also thank Nicholas Abbott for the collaboration that led to this paper.
- The thermostat examples provided in Chapter 3 are adaptations and extensions of an example provided in the textbook of Professor Tunde Ogunnaike [8].

- The electricity market Exercises 3.2, 3.5, and 3.9 use public data reported by the Electric Reliability Council of Texas (ERCOT) [9]. I thank Jiaze Ma for his help in putting together these exercises.
- The brain analysis problem of Exercise 3.7 is based on data reported in [10]. I thank Alexander Smith for his help in identifying this dataset.
- The problem of Exercise 4.8 is based on the content and data reported in [11].
- The Hougen–Watson estimation examples of Chapters 5 and 6 use data that is reprinted (adapted) with permission from [12]. Copyright 1960, American Chemical Society.
- Exercise 5.1 (lake quality) uses public data reported by the Wisconsin Department of Natural Resources [13]. I thank David Cole for his help in putting together this example.
- Exercise 5.5 (predicting solubility) uses data that is reprinted (adapted) with permission from [14]. Copyright 1999, American Chemical Society. I thank Montgomery Laky and Daniel Laky for their help in putting together this example.
- Exercises 5.6 and 5.7 (predicting CMC) use data that is reprinted (adapted) with permission from [15]. Copyright 2021, American Chemical Society. I thank Amy Qin for her help in putting together these exercises.
- Exercise 5.8 uses public data reported by the US National Oceanic and Atmospheric Association [16].
- Exercise 5.11 and Example 7.11 (experimental design) use data reported in [17].
- Exercise 5.12 is based on research reported in [18]. The data obtained from this research is used with permission from the authors. I thank Raka Dastidar and George Huber for their help in putting together this exercise.
- Examples 6.6, 6.19, and 6.29 (liquid crystal droplets) use data that is reprinted (adapted) with permission from [19]. Copyright 2023, American Chemical Society. I thank Amy Qin for her help in putting together these examples; I also thank Claribel Acevedo-Velez, David Lynn, Reid Van Lehn, and Fengrui Wang for the collaboration that led to this paper.
- Exercises 6.1, 6.2, and 6.3 (plastic characterization) use data that is reprinted (adapted) with permission from [20]. Copyright 2022, American Chemical Society. I thank Bruce Jiang for his help in putting together these examples; I also thank Fei Long and Ezra Bar-Ziv for the collaboration that led to this paper.
- Exercise 6.5 (asset placement) uses public data reported by the California Independent System Operator (CAISO) [21].
- Exercise 6.8 (sulfide removal) uses data that is reprinted with permission from [22]. Copyright 2013, Springer Nature. I thank Ryan Ettie for identifying this dataset as part of his final project for my statistics course at the University of Wisconsin–Madison.
- Exercise 6.10 (optimizing footwear materials) uses data that is published in [23]. I thank Amy Qin for her help in putting together this exercise; I also thank Allen Roman and Tim Osswald for the collaboration that led to this paper.

- Exercise 6.11 (applying convolutional filters) uses data that is reprinted (adapted) with permission from [24]. Copyright 2023, American Chemical Society. I thank Bruce Jiang for his help in putting together this exercise; I also thank Nicholas Abbott, Manos Mavrikakis, Alexander Smith, and Reid Van Lehn for the collaboration that led to this paper.
- Exercise 7.10 is inspired by the model reported in [25]. I thank Jiaze Ma for his help in putting together this exercise.
- Exercises 7.4 and 7.5 are based on the model reported in [26]. I thank Javier Tovar-Facio for his help in developing this model.