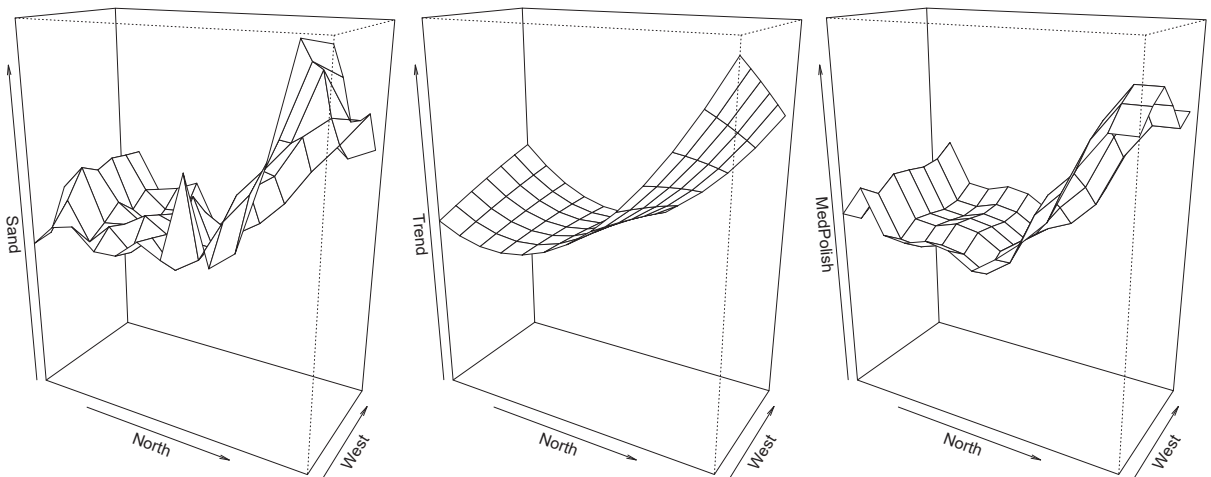


# Spatial Data Analysis in Ecology and Agriculture Using R

SECOND EDITION



Richard E. Plant



CRC Press  
Taylor & Francis Group

# Spatial Data Analysis in Ecology and Agriculture Using R



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

# Spatial Data Analysis in Ecology and Agriculture Using R

## Second Edition

Richard E. Plant

Departments of Plant Sciences and Biological and Agricultural Engineering  
University of California, Davis



**CRC Press**

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2019 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper

International Standard Book Number-978-0-8153-9275-0 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

---

#### Library of Congress Cataloging-in-Publication Data

---

Names: Plant, Richard E., author.  
Title: Spatial data analysis in ecology and agriculture using R / by Richard E. Plant.  
Description: Second edition. | Boca Raton, Florida : CRC Press, [2019] | Includes bibliographical references and index.  
Identifiers: LCCN 2018040387 | ISBN 9780815392750 (hardback : alk. paper) | ISBN 9781351189910 (e-book)  
Subjects: LCSH: Agriculture--Statistical methods. | Spatial analysis (Statistics) | R (Computer program language)  
Classification: LCC S566.55 .P53 2019 | DDC 338.1072/7--dc23  
LC record available at <https://lcn.loc.gov/2018040387>

---

Visit the Taylor & Francis Web site at  
<http://www.taylorandfrancis.com>

and the CRC Press Web site at  
<http://www.crcpress.com>

*To Kathie, Carolyn, Suzanne, and Eleanor, with all my love*



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

---

# Contents

---

Preface to the First Edition.....	xiii
Preface to the Second Edition.....	xv
Author.....	xvii
<b>1. Working with Spatial Data.....</b>	<b>1</b>
1.1 Introduction.....	1
1.2 Analysis of Spatial Data.....	5
1.2.1 Types of Spatial Data.....	5
1.2.2 The Components of Spatial Data.....	7
1.2.3 Spatial Data Models.....	7
1.2.4 Topics Covered in the Text.....	8
1.3 The Data Sets Analyzed in This Book.....	9
1.3.1 Data Set 1: Yellow-Billed Cuckoo Habitat.....	10
1.3.2 Data Set 2: Environmental Characteristics of Oak Woodlands.....	12
1.3.3 Data Set 3: Uruguayan Rice Farmers.....	13
1.3.4 Data Set 4: Factors Underlying Yield in Two Fields.....	14
1.3.5 Comparing the Data Sets.....	15
1.4 Further Reading.....	16
<b>2. The R Programming Environment.....</b>	<b>19</b>
2.1 Introduction.....	19
2.1.1 Introduction to R.....	19
2.1.2 Setting Yourself Up to Use This Book.....	20
2.2 R Basics.....	21
2.3 Programming Concepts.....	26
2.3.1 Looping and Branching.....	26
2.3.2 Functional Programming.....	28
2.4 Handling Data in R.....	30
2.4.1 Data Structures in R.....	30
2.4.2 Basic Data Input and Output.....	33
2.4.3 Spatial Data Structures.....	34
2.5 Writing Functions in R.....	43
2.6 Graphics in R.....	46
2.6.1 Traditional Graphics in R: Attribute Data.....	47
2.6.2 Traditional Graphics in R: Spatial Data.....	52
2.6.3 Trellis Graphics in R, Attribute Data.....	55
2.6.4 Trellis Graphics in R, Spatial Data.....	57
2.6.5 Using Color in R.....	59
2.7 Continuing on from Here with R.....	61
2.8 Further Reading.....	62
Exercises.....	62



<b>3. Statistical Properties of Spatially Autocorrelated Data .....</b>	<b>67</b>
3.1 Introduction .....	67
3.2 Components of a Spatial Random Process .....	68
3.2.1 Spatial Trends in Data .....	68
3.2.2 Stationarity .....	74
3.3 Monte Carlo Simulation .....	77
3.4 A Review of Hypothesis and Significance Testing .....	79
3.5 Modeling Spatial Autocorrelation .....	84
3.5.1 Monte Carlo Simulation of Time Series .....	84
3.5.2 Modeling Spatial Contiguity .....	88
3.5.3 Modeling Spatial Association in R .....	93
3.6 Application to Field Data .....	98
3.6.1 Setting Up the Data .....	98
3.6.2 Checking Sequence Validity .....	101
3.6.3 Determining Spatial Autocorrelation .....	102
3.7 Further Reading .....	103
Exercises .....	103
<b>4. Measures of Spatial Autocorrelation .....</b>	<b>105</b>
4.1 Introduction .....	105
4.2 Preliminary Considerations .....	105
4.2.1 Measurement Scale .....	105
4.2.2 Resampling and Randomization Assumptions .....	108
4.2.3 Testing the Null Hypothesis .....	109
4.3 Join-Count Statistics .....	110
4.4 Moran's $I$ and Geary's $c$ .....	114
4.5 Measures of Autocorrelation Structure .....	117
4.5.1 The Moran Correlogram .....	117
4.5.2 The Moran Scatterplot .....	119
4.5.3 Local Measures of Autocorrelation .....	121
4.5.4 Geographically Weighted Regression .....	124
4.6 Measuring Autocorrelation of Spatially Continuous Data .....	127
4.6.1 The Variogram .....	127
4.6.2 The Covariogram and the Correlogram .....	132
4.7 Further Reading .....	133
Exercises .....	133
<b>5. Sampling and Data Collection .....</b>	<b>135</b>
5.1 Introduction .....	135
5.2 Preliminary Considerations .....	138
5.2.1 The Artificial Population .....	138
5.2.2 Accuracy, Bias, Precision, and Variance .....	141
5.2.3 Comparison Procedures .....	142
5.3 Developing the Sampling Patterns .....	142
5.3.1 Random Sampling .....	142
5.3.2 Geographically Stratified Sampling .....	144
5.3.3 Sampling on a Regular Grid .....	146
5.3.4 Stratification Based on a Covariate .....	148
5.3.5 Cluster Sampling .....	153

5.4	Methods for Variogram Estimation.....	154
5.5	Estimating the Sample Size .....	157
5.6	Sampling for Thematic Mapping.....	158
5.7	Design-Based and Model-Based Sampling .....	159
5.8	Further Reading .....	164
	Exercises .....	164
<b>6.</b>	<b>Preparing Spatial Data for Analysis.....</b>	<b>167</b>
6.1	Introduction .....	167
6.2	Quality of Attribute Data.....	168
6.2.1	Dealing with Outliers and Contaminants .....	168
6.2.2	Quality of Ecological Survey Data .....	170
6.2.3	Quality of Automatically Recorded Data .....	170
6.3	Spatial Interpolation Procedures .....	175
6.3.1	Inverse Weighted Distance Interpolation.....	175
6.3.2	Kriging Interpolation .....	180
6.3.3	Cokriging Interpolation .....	183
6.4	Spatial Rectification and Alignment of Data .....	188
6.4.1	Definitions of Scale Related Processes.....	188
6.4.2	Change of Coverage.....	190
6.4.3	Change of Support.....	193
6.5	Further Reading .....	196
	Exercises .....	197
<b>7.</b>	<b>Preliminary Exploration of Spatial Data.....</b>	<b>199</b>
7.1	Introduction .....	199
7.2	Data Set 1.....	201
7.3	Data Set 2.....	214
7.4	Data Set 3.....	229
7.5	Data Set 4.....	240
7.6	Further Reading .....	251
	Exercises .....	251
<b>8.</b>	<b>Data Exploration Using Non-Spatial Methods: The Linear Model.....</b>	<b>255</b>
8.1	Introduction .....	255
8.2	Multiple Linear Regression .....	255
8.2.1	The Many Perils of Model Selection .....	255
8.2.2	Multicollinearity, Added Variable Plots, and Partial Residual Plots .....	261
8.2.3	A Cautious Approach Model Selection as an Exploratory Tool.....	269
8.3	Building a Multiple Regression Model for Field 4.1.....	270
8.4	Generalized Linear Models .....	281
8.4.1	Introduction to Generalized Linear Models .....	281
8.4.2	Multiple Logistic Regression Model for Data Set 2.....	288
8.4.3	Logistic Regression Model of Count Data for Data Set 1.....	295
8.4.4	Analysis of the Counts of Data Set 1: Zero-Inflated Poisson Data .....	299
8.5	Further Reading .....	303
	Exercises .....	304

<b>9. Data Exploration Using Non-Spatial Methods: Nonparametric Methods .....</b>	<b>307</b>
9.1 Introduction .....	307
9.2 The Generalized Additive Model .....	307
9.3 Classification and Regression Trees (a.k.a. Recursive Partitioning) .....	317
9.3.1 Introduction to the Method .....	317
9.3.2 The Mathematics of Recursive Partitioning .....	320
9.3.3 Exploratory Analysis of Data Set 2 with Regression Trees .....	321
9.3.4 Exploratory Analysis of Data Set 3 with Recursive Partitioning .....	328
9.3.5 Exploratory Analysis of Field 4.1 with Recursive Partitioning .....	334
9.4 Random Forest.....	339
9.4.1 Introduction to Random Forest.....	339
9.4.2 Application to Data Set 2.....	342
9.5 Further Reading .....	345
Exercises .....	345
<b>10. Variance Estimation, the Effective Sample Size, and the Bootstrap.....</b>	<b>347</b>
10.1 Introduction .....	347
10.2 Bootstrap Estimation of the Standard Error .....	351
10.3 Bootstrapping Time Series Data .....	355
10.3.1 The Problem with Correlated Data.....	355
10.3.2 The Block Bootstrap .....	357
10.3.3 The Parametric Bootstrap .....	360
10.4 Bootstrapping Spatial Data.....	362
10.4.1 The Spatial Block Bootstrap.....	362
10.4.2 The Parametric Spatial Bootstrap .....	366
10.4.3 Power of the Tests.....	368
10.5 Application to the EM38 Data .....	368
10.6 Further Reading .....	371
Exercises .....	372
<b>11. Measures of Bivariate Association between Two Spatial Variables.....</b>	<b>373</b>
11.1 Introduction .....	373
11.2 Estimating and Testing the Correlation Coefficient .....	376
11.2.1 The Correlation Coefficient.....	376
11.2.2 The Clifford et al. (1989) Correction .....	378
11.2.3 The Bootstrap Variance Estimate.....	381
11.2.4 Application to the Example Problem .....	383
11.3 Contingency Tables.....	386
11.3.1 Large Sample Size Contingency Tables .....	386
11.3.2 Small Sample Size Contingency Tables .....	393
11.4 The Mantel and Partial Mantel Statistics.....	398
11.4.1 The Mantel Statistic .....	398
11.4.2 The Partial Mantel Test .....	401
11.5 The Modifiable Areal Unit Problem and the Ecological Fallacy .....	404
11.5.1 The Modifiable Areal Unit Problem .....	404
11.5.2 The Ecological Fallacy .....	408
11.6 Further Reading .....	410
Exercises .....	410

<b>12. The Mixed Model .....</b>	<b>413</b>
12.1 Introduction .....	413
12.2 Basic Properties of the Mixed Model .....	417
12.3 Application to Data Set 3.....	419
12.4 Incorporating Spatial Autocorrelation.....	422
12.5 Generalized Least Squares .....	429
12.6 Spatial Logistic Regression.....	431
12.6.1 Upscaling Data Set 2 in the Coast Range .....	431
12.6.2 The Incorporation of Spatial Autocorrelation.....	436
12.7 Further Reading .....	443
Exercises .....	444
<b>13. Regression Models for Spatially Autocorrelated Data .....</b>	<b>445</b>
13.1 Introduction .....	445
13.2 Detecting Spatial Autocorrelation in a Regression Model.....	450
13.3 Models for Spatial Processes .....	452
13.3.1 The Spatial Lag Model.....	452
13.3.2 The Spatial Error Model.....	454
13.4 Determining the Appropriate Regression Model .....	455
13.4.1 Formulation of the Problem.....	455
13.4.2 The Lagrange Multiplier Test .....	456
13.5 Fitting the Spatial Lag and Spatial Error Models.....	458
13.6 The Conditional Autoregressive Model.....	460
13.7 Application of Simultaneous Autoregressive and Conditional Autoregressive Models to Field Data .....	462
13.7.1 Fitting the Data.....	462
13.7.2 Comparison of the Mixed Model and Spatial Autoregression.....	465
13.8 Further Reading .....	466
Exercises .....	466
<b>14. Bayesian Analysis of Spatially Autocorrelated Data .....</b>	<b>467</b>
14.1 Introduction .....	467
14.2 Markov Chain Monte Carlo Methods.....	471
14.3 Introduction to WinBUGS.....	478
14.3.1 WinBUGS Basics.....	478
14.3.2 WinBUGS Diagnostics.....	481
14.3.3 Introduction to R2WinBUGS.....	483
14.3.4 Generalized Linear Models in WinBUGS.....	490
14.4 Hierarchical Models .....	492
14.5 Incorporation of Spatial Effects.....	498
14.5.1 Spatial Effects in the Linear Model .....	498
14.5.2 Application to Data Set 3.....	501
14.5.3 The spBayes Package .....	505
14.6 Comparison of the Methods.....	509
14.7 Further Reading .....	510
Exercises .....	511

<b>15. Analysis of Spatiotemporal Data .....</b>	<b>513</b>
15.1 Introduction .....	513
15.2 Spatiotemporal Data Interpolation.....	513
15.2.1 Representing Spatiotemporal Data.....	513
15.2.2 The Spatiotemporal Variogram.....	518
15.2.3 Interpolating Spatiotemporal Data.....	523
15.3 Spatiotemporal Process Models.....	525
15.3.1 Models for Dispersing Populations.....	525
15.3.2 A Process Model for the Yield Data.....	526
15.4 Finite State and Time Models.....	529
15.4.1 Determining Finite State and Time Models Using Clustering.....	529
15.4.2 Factors Underlying Finite State and Time Models.....	538
15.5 Bayesian Spatiotemporal Analysis .....	543
15.5.1 Introduction to Bayesian Updating .....	543
15.5.2 Application of Bayesian Updating to Data Set 3.....	546
15.6 Further Reading .....	550
Exercises .....	551
<b>16. Analysis of Data from Controlled Experiments .....</b>	<b>553</b>
16.1 Introduction .....	553
16.2 Classical Analysis of Variance .....	554
16.3 The Comparison of Methods.....	559
16.3.1 The Comparison Statistics .....	559
16.3.2 The Papadakis Nearest-Neighbor Method.....	561
16.3.3 The Trend Method .....	562
16.3.4 The “Correlated Errors” Method.....	563
16.3.5 Published Comparisons of the Methods .....	565
16.4 Pseudoreplicated Data and the Effective Sample Size.....	566
16.4.1 Pseudoreplicated Comparisons .....	566
16.4.2 Calculation of the Effective Sample Size .....	567
16.4.3 Application to Field Data .....	569
16.5 Further Reading .....	571
Exercises .....	572
<b>17. Assembling Conclusions.....</b>	<b>573</b>
17.1 Introduction .....	573
17.2 Data Set 1.....	573
17.3 Data Set 2.....	578
17.4 Data Set 3.....	583
17.5 Data Set 4.....	586
17.6 Conclusions.....	590
<b>Appendix A: Review of Mathematical Concepts .....</b>	<b>593</b>
<b>Appendix B: The Data Sets .....</b>	<b>619</b>
<b>Appendix C: An R Thesaurus .....</b>	<b>627</b>
<b>References .....</b>	<b>635</b>
<b>Index .....</b>	<b>657</b>

---

## *Preface to the First Edition*

---

This book is intended for classroom use or self-study by graduate students and researchers in ecology, geography, and agricultural science who wish to learn about the analysis of spatial data. The book originated in a course entitled “Spatial Data Analysis in Applied Ecology” that I taught for several years at UC Davis. Although most of the students were enrolled in Ecology, Horticulture and Agronomy, or Geography, there was a smattering of students from other programs such as Entomology, Soil Science, and Agricultural and Resource Economics. The book assumes that the reader has a background in statistics at the level of an upper division undergraduate applied linear models course. This is also, in my experience, the level at which ecologists and agronomists teach graduate applied statistics courses to their own students. To be specific, the book assumes a statistical background at the level of Kutner et al. (2005). I do not assume that the reader has had exposure to the general linear model or modern mixed-model analysis.

The book is intended for those who want to make use of these methods in their research, not for statistical or geographical specialists. It is always wise to seek out a specialist’s help when such help is available, and I strongly encourage this practice. Nevertheless, the more one knows about a specialist’s area of knowledge, the better able one is to make use of that knowledge. Because this is a user’s book, I have elected on some occasions to take small liberties with technical points and details of the terminology. To dot every *i* and cross every *t* would drag the presentation down without adding anything useful.

The book does not assume any prior knowledge of the R programming environment. All of the R code for all of the analyses carried out in this book is available on the book’s companion website, <https://psfaculty.plantsciences.ucdavis.edu/plant/sda.htm>. One of the best features of R is also its most challenging: the vast number of functions and contributed packages that provide a multitude of ways to solve any given problem. This provides a special challenge for a textbook, namely, how to find the best compromise between exposition via manual coding and the use of contributed package functions, and which functions to choose. I have tried to use manual coding when it is easy or there is a point to be made, and save contributed functions for more complex operations. As a result, I sometimes have provided “homemade” code for operations that can also be carried out by a function from a contributed package.

The book focuses on data from four case studies, two from uncultivated ecosystems and two from cultivated ecosystems. The data sets are also available on the book’s companion website. Each of the four data sets is drawn from my own research. My reason for this approach is a conviction that if one wants to really get the most out of a data set, one has to live with it for a while and get to know it from many different perspectives, and I want to give the reader a sense of that process as I have experienced it. I make no claim of uniqueness for this idea; Griffith and Layne (1999), for example, have done it before me. I used data from projects in which I participated not because I think they are in any way special, but because they were available and I already knew them well when I started to write the book.

For most of my career, I have had a joint appointment in the Department of Biological and Agricultural Engineering and a department that, until it was gobbled up in a fit of academic consolidation, bore the name Agronomy and Range Science. Faculty in this second department were fortunate in that, as the department name implies, we were able to

work in both cultivated and uncultivated ecosystems, and this enabled me to include two of each in the book. I was originally motivated to write the book based on my experiences working in precision agriculture. We in California entered this arena considerably later than our colleagues in the Midwest, but I was in at the beginning in California. As was typical of academic researchers, I developed methods for site-specific crop management, presented them to farmers at innumerable field days, and watched with bemusement, as they were not adopted very often. After a while I came to the realization that farmers can figure out how best to use this new technology in the ways that suit their needs, and indeed they are beginning to do so. This led to the further realization that we researchers should be using this technology for what we do best: research. This requires learning how to analyze the data that the technology provides, and that is what this book is about.

I have been very fortunate to have had some truly outstanding students work with me, and their work has contributed powerfully to my own knowledge of the subject. I particularly want to acknowledge those students who contributed to the research that resulted in this book, including (in alphabetical order) Steven Greco, Peggy Hauselt, Randy Horney, Claudia Marchesi, Ali Mermer, Jorge Perez-Quezada, Alvaro Roel, and Marc Vayssières. In particular, Steven Greco provided Data Set 1, Marc Vayssières provided Data Set 2, and Alvaro Roel provided Data Set 3. I also want to thank the students in my course who read through several versions of this book and made many valuable suggestions. In particular, thanks go to Kimberley Miller, who read every chapter of the final draft and made many valuable comments. I have benefited from the interaction with a number of colleagues, too many to name, but I particularly want to thank Hugo Firpo, who collected the data on Uruguayan rice farmers for Data Set 3, Tony Turkovich, who let us collect the data for Data Set 4 in his fields, Stuart Pettygrove, who managed the large-scale effort that led to that data set and made the notes and data freely available, and Robert Hijmans, who introduced me to his raster package and provided me with many valuable comments about the book in general. Finally, I want to thank Roger Bivand, who, without my asking him to do so, took the trouble to read one of the chapters and made several valuable suggestions. Naturally, these acknowledgments in no way imply that the persons acknowledged, or anyone else, endorses what I have written. Indeed, some of the methods presented in this book are very *ad hoc* and may turn out to be inappropriate. For that, as well as for the mistakes in the book and bugs in the code that I am sure are lurking there, I take full responsibility.

**Davis**  
*California*

---

## *Preface to the Second Edition*

---

I am grateful for the opportunity to write a second edition of this book. R is very dynamic, and there have been sufficient changes to degrade the book's usefulness for learning R. There have also been major advances on several fronts in spatial data analysis. The most dramatic changes have been in the analysis of spatiotemporal data. These have been sufficient for me to completely revise that chapter. Major advances have also been made in the application of Bayesian methods to spatial data. The use of the generalized additive model has been rapidly gaining ground in ecology. Finally, but not least importantly, the use of packages associated with the tidyverse of Hadley Wickham and his colleagues have made graphical analysis much simpler. I have incorporated all of these, although, for reasons I have elaborated in [Chapter 2](#), I have decided to continue to use the traditional R graphics to construct the figures in this book. Of course this means that some things have had to be deleted to make room. Most prominent has been the removal of the section on principal components analysis. The substitution of a section on the generalized additive model has prompted me to switch the order of this chapter and the one on linear models. I have placed the section on principal components analysis on the book's companion website, <https://psfaculty.plantsciences.ucdavis.edu/plant/sda2.htm> in a section called "Additional Topics." Additional material to accompany the book can be accessed at <https://psfaculty.plantsciences.ucdavis.edu/plant/additionaltopics.htm>. I hope to add discussions of other relevant topics there as well. I have also had the opportunity to correct numerous errors in the first edition; I hope I have not introduced too many new ones into the second. In addition to the people I acknowledged in the first edition, I want to thank Meili Baragatti, James Graham, and Andrew Latimer for their thoughtful reviews of the first edition. I would also like to thank my editors, John Sulzycki and Alice Oven, for their guidance and assistance. They have been a true pleasure to work with.

**Davis**  
*California*