



Statistical Methods in the Atmospheric Sciences by Daniel S. Wilks

Review by: Richard A. Levine

Journal of the American Statistical Association, Vol. 95, No. 449 (Mar., 2000), pp. 344-345

Published by: [American Statistical Association](#)

Stable URL: <http://www.jstor.org/stable/2669579>

Accessed: 15/06/2014 10:51

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*.

<http://www.jstor.org>

principle and the Bellman equation are introduced in their most natural setting. The developed techniques are illustrated in the chapter on inventory models. The book's second half deals with ergodic cost-optimization problems. Those problems in general require a rather delicate treatment even in the simplest situation, due to a number of pitfalls that exist within the "ergodic part" of MDP area. First, the necessary tools are developed for analysis of the ergodic problems for finite state-space processes, and then these tools are extended to infinite state-spaces. Two chapters cover application of the developed techniques to various queueing models. The last chapter can be viewed as an introduction to continuous-time, discrete, state-space dynamic programming.

In short, *Stochastic Dynamic Programming and the Control of Queueing Systems* is a well-written, easy-to-read book that is neither overloaded with heavy analytics nor with excessive number of distractive computations. Written as a textbook for master's-level students, it will be also a welcome addition to the library of any specialist on queueing theory or stochastic control.

Michael TAKSAR
State University of New York at Stony Brook

Statistical Methods in the Atmospheric Sciences.

Daniel S. WILKS. New York: Academic Press, 1995. ISBN 0-12-751965-3. xi + 467 pp. \$75.

Daniel Wilks writes in the second sentence of this book that "Students (and others) often resist statistics, and the subject is perceived by many to be the epitome of dullness." He then concludes the first paragraph by hoping his text will overcome this (unfortunately) widely shared point of view. I applaud Professor Wilks for making this effort in such diversified fields as the "deterministic" physical sciences and "random" statistical sciences. In my opinion he has achieved this goal admirably, plus much more, with a clearly written, well-organized, and lively text introducing statistics for the atmospheric sciences in its full glory.

Statistics as applied to the atmospheric sciences is an immense field to tackle in one book. A look at the chapter titles—Introduction, Review of Probability, Empirical Distributions and Exploratory Data Analysis, Theoretical Probability Distributions, Hypothesis Testing, Statistical Weather Forecasting, Forecast Verification, Time Series, and Methods for Multivariate Data—suggests the vast spectrum of statistical methods covered. As a statistics textbook for upper-level undergraduates and beginning graduate students in the atmospheric sciences, the coverage of meteorology subject matter is equally extensive. Furthermore, the audience is expected to have had a course in statistics and perhaps experience in meteorology, although this is not necessary, so the book is appealing to the novice and experienced statistician and geophysical scientist alike.

The fields of statistics and the atmospheric sciences encompass different training tactics, approaches, and philosophies, although the methodological overlaps are substantive. Climatology is a geophysical science where many ideas and techniques are developed independently of other fields through subject specific problems. Statistics is an art influenced by many scientific fields, a large subset of which is the mathematical sciences. Consequently, the terminology and practices, though solving similar problems, appear on the surface much different. Professor Wilks successfully presents statistics to atmospheric science students and researchers in his book by understanding and embracing these differences, tying together and clarifying common themes underlying the two disciplines. The pathway to enlightenment may be described through three themes played out in the book: scientific problem solving, the many roles of statistics in the atmospheric sciences, and the concern of statistics with uncertainty.

The book is written for researchers, practitioners, and students in the atmospheric sciences as both a teaching and a reference tool. The audience consists of scientists needing statistics to understand "what their datasets mean" and solve geophysical problems. Professor Wilks thus introduces most of the statistical subject material as a means to solve climatological problems. One motivating example used throughout the book is weather forecasting, an inherently statistical problem. Regression analysis is motivated beautifully in Chapter 6 as a method to enhance forecast model output and ultimately predict the weather. Following a practical illustra-

tion of statistical weather forecasting, the basics of regression analysis are presented in Section 6.2 with examples relating back to the forecasting problem. The rest of the chapter details weather forecasting from a regression standpoint, in the process explaining the shortfalls and advantages of the method toward analyzing weather data. In particular, Wilks presents regression diagnostics, model selection, cross-validation, statistical prediction, and logistic regression within this framework. The reader can then more easily follow the mathematics through a substantive application underpinning the basic theory.

This method of presentation is used repeatedly in the text with great success. For instance, probability concepts and distributions in Chapter 2 are motivated through an illustration of the probability of precipitation on a given day. This familiar example allows for an easy description of probability as a long-run frequency, Bayes theorem, and Bayesian modeling perspectives. Chapter 5 differentiates between parametric and non-parametric tests and then introduces resampling schemes through a study of the effect of cloud seeding on lightening strikes. Chapter 7 introduces the "curse of dimensionality" common in multivariate and image analysis through an application in weather forecast verification. Datasets used in multiple chapters are presented in an appendix at the end of the book.

This approach to statistics as a scientific problem-solving tool lends itself to a casual expository. The reader receives lucid and concise descriptions of difficult statistical concepts through the author's eloquent writing style. The two-paragraph description of sampling distributions in Chapter 5 as motivated by analyzing January temperatures illustrates this point. The conversational style is most conducive to enjoyable reading, with smatterings of the author's wit, such as "it is worth taking a moment to understand why one would voluntarily commit the violence of shoe-horning real data into an abstract mold" and "some people feel more secure with the results of objective forecasting procedures, apparently taking comfort from their lack of contamination by the vagaries of human judgment." I applaud Professor Wilks for encouraging students to enjoy the *art* of statistics and approach the subject as an enthused scientist rather than a formula cruncher. In fact, just by reading the first sections of each chapter, the reader cannot help but get excited about statistical methods for atmospheric research in the presence of informative real life problems to be solved.

Statistics plays an important role in the atmospheric sciences. The interplay of the physical deterministic systems and mathematical models often applied in climatological research with statistical models is crucial for solving problems. The numerous entries concerning statistical methods in the atmospheric sciences in the *Meteorological and Geophysical Abstracts*, the *Current Index of Statistics*, and other relevant databases is a testament to this fact, as is the number of books published in the past decade under the heading of statistical meteorology presenting either basic statistical methods at a slightly lower level than Wilks (e.g., Thiébaux 1994) or more advanced, subject-specific methods (e.g., Daley 1991; Von Storch and Navarra 1995; Polyak 1996; Rao, Priestley, and Lessi 1997; Berliner, Nychka, and Hoar 2000). (Note, however, that no text at the level of Wilks's book has appeared since Panofsky and Brier 1958.)

Professor Wilks emphasizes the importance of statistical methods in all aspects of climatological research. The motivating application of statistics in atmospheric sciences in Chapter 1 is that "if it were possible to make perfect [weather] forecasts even one day into the future (i.e., if there were no uncertainty involved), the practice of meteorology would be very dull." Chapter 3 emphasizes the value of statistics to summarize the "torrents of numerical data" present in the atmospheric sciences. An issue often confronted by atmospheric scientists is temporal dependence in high-dimensional data. Chapters 8 and 9 present statistical tools beyond the introductory level to handle such data both graphically and analytically. In particular, frequency and time domain methods, including stationarity, harmonic and spectral analysis, Markov chain modeling, and ARMA processes, as well as data reduction and classification methods, are introduced. The presentation is oriented toward scientific problem solving, with goals such as teasing apart annual cycles in a temperature series and studying geopotential heights over spatial fields/arrays. The conversational style again allows for a fuller understanding of the underlying statistical concepts without losing site of the primary issues among mathematical formulations.

Of course, part of the confusion in learning statistics for application in a scientific field is the abundance of terminology, much of which uses different phrases to identify the same method. For example, transformations via normalization and centering in the statistics literature are forms of (standardized) anomalies in climatology. Principal component analysis is in essence empirical orthogonal function analysis as used by atmospheric scientists, not to be confused with factor analysis or canonical correlation analysis. The text points to these difficulties as they arise, clearly and concisely explaining the differences. Furthermore, for the reader unfamiliar with meteorological concepts and techniques, the text explains terms such as teleconnection, persistence, harmonic dials, and one-point correlation maps for understanding by scientifically oriented readers without any loss of flow in the discussion.

Wilks underlines the importance of statistics in the atmospheric sciences through persuasive arguments toward the combination of physical and statistical thinking and modeling. The theme is why and how do we incorporate uncertainty into atmospheric processes. Uncertainty plagues the atmospheric sciences in three areas: atmospheric variability, chaotic systems, and imprecise predictions. Wilks elucidates each in Chapter 1, to motivate the modeling of uncertainty in climatology for the remainder of the book.

The first aspect of uncertainty about the atmosphere is obvious in that "the atmosphere exhibits variations and fluctuations that are irregular. This uncertainty is the driving force behind the collection and analysis of large [atmospheric] data sets." However, incorporation of statistical models in the deterministic physical systems used by climate modelers is not so clear. Wilks presents an excellent discussion on *dynamical chaos*, "the death knell for the dream of perfect (uncertainty-free) weather forecasts," illustrating the need for statistical modeling even in these deterministic systems. The idea is that given the sensitivity of a nonlinear, dynamical system to initial conditions, it is impossible for even the perfect mathematical model to calculate what the atmosphere will do. "Even if the atmosphere is not fundamentally a random system, for many practical purposes it might as well be." And of course, random systems are "not precisely predictable." Thus the need for statistical and probabilistic models to analyze climate model data and infer the precision of predictions from the model and observed data. (As a shameless self-reference, see the discussion in Levine and Berliner 1999 for more details on this point in the area of climate change.)

Wilks illustrates the link between physical and statistical modeling most perceptibly in his discussion of forecast methods in Chapter 6. The reader is stepped through the history of statistical weather forecasting, starting with a purely statistical regression model to the incorporation of output from numerical weather prediction models to ensemble and probability forecasts. The exposition demonstrates the marriage between statistics and meteorology necessary to successfully solve these difficult geophysical problems.

In my view, Wilks not only achieves overcoming the perception of statistics as the "epitome of dullness," but also in fact portrays the subject as an exciting field to be appreciated for its relevance to geophysical problems. I recommend this book without hesitation to statisticians and atmospheric scientists alike as a learning tool or reference text or for the casual reader interested in statistical meteorology/meteorological statistics. With more books like these in the applied sciences, perhaps we can someday overcome the stigma that "statistics is dull."

Comment: As a statistician writing for a statistics journal, I approached this review from the view of a statistician dabbling or with an interest in atmospheric sciences. For the climatologist point of view, see the review by Livezey (1995) in the *Bulletin of the American Meteorological Society*.

Richard A. LEVINE
University of California-Davis

REFERENCES

Berliner, L. M., Nychka, D. W., and Hoar, T. J. (Eds.) (2000), *Statistics in the Climate Sciences*, New York: Springer.

Daley, R. (1991), *Atmospheric Data Analysis*, New York: Cambridge University Press.

Levine, R. A., and Berliner, L. M. (1999), "Statistical Principles for Climate Change Studies," *Journal of Climate*, 12, 565–574.

Livezey, R. E. (1995), Review of *Statistical Methods in the Atmospheric Sciences* by D. S. Wilks, *Bulletin of the American Meteorological Society*, 76, 1820–1822.

Panofsky, H., and Brier, G. W. (1958), *Some Applications of Statistics to Meteorology*, University Park, PA: Pennsylvania State University.

Polyak, I. (1996), *Computational Statistics in Climatology*, New York: Oxford University Press.

Thiébaux, H. J. (1994), *Statistical Data Analysis for Ocean and Atmospheric Sciences*, New York: Academic Press.

Von Storch, H., and Navarra, A. (1995), *Analysis of Climate Variability*, New York: Springer.

Case Studies in Environmental Statistics.

Douglas NYCHKA, Walter W. PIEGORSCH, and Lawrence H. COX (Eds.). New York: Springer-Verlag, 1998. ISBN 0-387-98478-X. xi + 196 pp. \$49 (P).

This book grew out of a program of the National Institute of Statistical Sciences, in conjunction with the Environmental Protection Agency, on problems in environmental science and regulation. The main content is a collection of six papers on statistical design, modeling, and analysis in specific environmental situations. Also included are an introductory overview, notes from a workshop, and appendices. The six main chapters deal with various aspects of statistics applied to problems of air pollution.

Chapter 2, the first of the six main chapters, covers modeling ozone levels in the Chicago area as a function of variables such as temperature, relative humidity, and so on. The topics include nonlinear regression models, semiparametric models, and models of extreme values. Chapter 3 discusses modeling ozone levels in the Gulf Coast region, specifically in the Houston area. The focus is on modeling diurnal trends, as opposed to the modeling of maximum daily ozone levels in the previous chapter. Singular value decomposition methods are used to reduce the dimension of hourly ozone measures to components of diurnal trends. Chapter 4 returns to Chicago and describes methods for selecting air quality monitoring sites to augment existing sites in the greater Chicago area. Spatial statistics methods are used to choose new sites. Chapter 5 is on estimating trends in atmospheric deposition of pollutants, with a focus on acid rain. Various regression models are used to relate pollution to meteorological variables. Chapter 6 relates mortality to airborne particles, illustrated with data from Birmingham, Alabama, and other cities. Chapter 7, the last of the six main chapters, covers quantitative risk assessment using regression analysis with categorical response variables and refers to the chemical accident in Bhopal, India.

The workshop chapter deals with methods of combining environmental information. It discusses some well-known methods of combining estimates and *p* values, as well as some ideas specific to environmental applications. Appendix A describes FUNFITS, which augments S-PLUS curve-fitting capabilities. Appendix B describes DI, a package for constructing and analyzing spatial designs.

Chapter authors include recognized authorities on the subjects. Among other attributes, a wide variety of contemporary regression and curve-fitting techniques are covered. The book is well organized, and the case studies illustrate a good selection of topics in air pollution. Perhaps the major shortcoming is that the title implies coverage of environmental topics outside the realm of air pollution. Anyone interested in air pollution data would benefit from reading this book.

Ramon C. LITTELL
University of Florida

REFERENCES

Billingsley, P. (1995), *Probability and Measure* (3rd ed.), New York: Wiley.
Chung, K. L. (1974), *A Course in Probability Theory* (2nd ed.), New York: Academic Press.