Contents

Preface

As recently as two decades ago, the impact of hierarchical Bayesian methods outside of a small group of theoretical probabilists and statisticians was minimal at best. Realistic models for challenging data sets were easy enough to write down, but the computations associated with these models required integrations over hundreds or even thousands of unknown parameters, far too complex for existing computing technology. Suddenly, around 1990, the "Markov chain Monte Carlo (MCMC) revolution" in Bayesian computing took place. Methods like the Gibbs sampler and the Metropolis algorithm, when coupled with ever-faster workstations and personal computers, enabled evaluation of the integrals that had long thwarted applied Bayesians. Almost overnight, Bayesian methods became not only feasible, but the method of choice for almost any model involving multiple levels incorporating random effects or complicated dependence structures. The growth in applications has also been phenomenal, with a particularly interesting recent example being a Bayesian program to delete spam from your incoming email (see `popfile.sourceforge.net`).

Our purpose in writing this book is to describe hierarchical Bayesian methods for one class of applications in which they can pay substantial dividends: spatial (and spatiotemporal) statistics. While all three of us have been working in this area for some time, our motivation for writing the book really came from our experiences teaching courses on the subject (two of us at the University of Minnesota, and the other at the University of Connecticut). In teaching we naturally began with the textbook by Cressie (1993), long considered the standard as both text and reference in the field. But we found the book somewhat uneven in its presentation, and written at a mathematical level that is perhaps a bit high, especially for the many epidemiologists, environmental health researchers, foresters, computer scientists, GIS experts, and other users of spatial methods who lacked significant background in mathematical statistics. Now a decade old, the book also lacks a current view of hierarchical modeling approaches for spatial data.

But the problem with the traditional teaching approach went beyond the mere need for a less formal presentation. Time and again, as we presented the traditional material, we found it wanting in terms of its flexibility to deal with realistic assumptions. Traditional Gaussian kriging is obviously the most important method of point-to-point spatial interpolation, but extending the paradigm beyond this was awkward. For areal (block-level) data, the problem seemed even more acute: CAR models should most naturally appear as priors for the parameters in a model, not as a model for the observations themselves.

This book, then, attempts to remedy the situation by providing a fully Bayesian treatment of spatial methods. We begin in Chapter 1 by outlining and providing illustrative examples of the three types of spatial data: point-level (geostatistical), areal (lattice), and spatial point process. We also provide a brief introduction to map projection and the proper calculation of distance on the earth's surface (which, since the earth is round, can differ markedly from answers obtained using the familiar notion of Euclidean distance). Our statistical presentation begins in earnest in Chapter 2, where we describe both exploratory data analysis tools and traditional modeling approaches for point-referenced data. Modeling approaches from traditional geostatistics (variogram fitting, kriging, and so forth) are covered here. Chapter 3 offers a similar presentation for areal data models, again starting with choropleth maps and other displays and progressing toward more formal statistical models. This chapter also presents Brook's Lemma and Markov random fields, topics that underlie the conditional, intrinsic, and simultaneous autoregressive (CAR, IAR, and SAR) models so often used in areal data settings.

Chapter 4 provides a review of the hierarchical Bayesian approach in a fairly generic setting, for readers previously unfamiliar with these methods and related computing and software. (The penultimate sections of Chapters 2, 3, and 4 offer tutorials in several popular software packages.) This chapter is not intended as a replacement for a full course in Bayesian methods (as covered, for example, by Carlin and Louis, 2000, or Gelman et al., 2004), but should be sufficient for readers having at least some familiarity with the ideas. In Chapter 5 then we are ready to cover hierarchical modeling for univariate spatial response data, including Bayesian kriging and lattice modeling. The issue of nonstationarity (and how to model it) also arises here.

Chapter 6 considers the problem of spatially misaligned data. Here, Bayesian methods are particularly well suited to sorting out complex interrelationships and constraints and providing a coherent answer that properly accounts for all spatial correlation and uncertainty. Methods for handling multivariate spatial responses (for both point- and block-level data) are discussed in Chapter 7. Spatiotemporal models are considered in Chapter 8, while Chapter 9 presents an extended application of areal unit data modeling in the context of survival analysis methods. Chapter 10 considers novel methodology associated with spatial process modeling, including spatial directional derivatives, spatially varying coefficient models, and spatial cumulative distribution functions (SCDFs). Finally, the book also features two useful appendices. Appendix A reviews elements of matrix theory and important related computational techniques, while Appendix B contains solutions to several of the exercises in each of the book's chapters.

Our book is intended as a research monograph, presenting the "state of the art" in hierarchical modeling for spatial data, and as such we hope readers will find it useful as a desk reference. However, we also hope it will be of benefit to instructors (or self-directed students) wishing to use it as a textbook. Here we see several options. Students wanting an introduction to methods for point-referenced data (traditional geostatistics and its extensions) may begin with Chapter 1, Chapter 2, Chapter 4, and Section 5.1 to Section 5.3. If areal data models are of greater interest, we suggest beginning with Chapter 1, Chapter 3, Chapter 4, Section 5.4, and Section 5.5. In addition, for students wishing to minimize the mathematical presentation, we have also marked sections containing more advanced material with a star ($\star$). These sections may be skipped (at least initially) at little cost to the intelligibility of the subsequent narrative. In our course in the Division of Biostatistics at the University of Minnesota, we are able to cover much of the book in a 3-credit-hour, single-semester (15-week) course. We encourage the reader to check `http://www.biostat.umn.edu/~brad/` on the web for many of our data sets and other teaching-related information.

We owe a debt of gratitude to those who helped us make this book a reality. Kirsty Stroud and Bob Stern took us to lunch and said encouraging things (and more importantly, picked up the check) whenever we needed it. Cathy Brown, Alex Zirpoli, and Desdamona Racheli prepared significant portions of the text and figures. Many of our current and former graduate and postdoctoral students, including Yue Cui, Xu Guo, Murali Haran, Xiaoping Jin, Andy Mugglin, Margaret Short, Amy Xia, and Li Zhu at Minnesota, and Deepak Agarwal, Mark Ecker, Sujit Ghosh, Hyon-Jung Kim, Ananda Majumdar, Alexandra Schmidt, and Shanshan Wu at the University of Connecticut, played a big role. We are also grateful to the Spring 2003 *Spatial Biostatistics* class in the School of Public Health at the University of Minnesota for taking our draft for a serious "test drive." Colleagues Jarrett Barber, Nicky Best, Montserrat Fuentes, David Higdon, Jim Hodges, Oli Schabenberger, John Silander, Jon Wakefield, Melanie Wall, Lance Waller, and many others provided valuable input and assistance. Finally, we thank our families, whose ongoing love and support made all of this possible.

SUDIPTO BANERJEE — Minneapolis, Minnesota
BRADLEY P. CARLIN — Durham, North Carolina
ALAN E. GELFAND — October 2003