Case Studies in Bayesian Statistical Modelling and Analysis

Editors:
Clair L. Alston, Kerrie L. Mengersen and Anthony N. Pettitt, Queensland University of Technology, Brisbane, Australia

Provides an accessible foundation to Bayesian analysis using real world models

This book aims to present an introduction to Bayesian modelling and computation, by considering real case studies drawn from diverse fields spanning ecology, health, genetics and finance. Each chapter comprises a description of the problem, the corresponding model, the computational method, results and inferences as well as the issues that arise in the implementation of these approaches.

Case Studies in Bayesian Statistical Modelling and Analysis:
- Illustrates how to do Bayesian analysis in a clear and concise manner using real-world problems.
- Each chapter focuses on a real-world problem and describes the way in which the problem may be analysed using Bayesian methods.
- Features approaches that can be used in a wide area of application, such as, health, the environment, genetics, information science, medicine, biology, industry and remote sensing.

Case Studies in Bayesian Statistical Modelling and Analysis is aimed at statisticians, researchers and practitioners who have some expertise in statistical modelling and analysis, and some understanding of the basics of Bayesian statistics, but little experience in its application. Graduate students of statistics and biostatistics will also find this book beneficial.

www.wiley.com/go/statistical_modelling
Case Studies in Bayesian Statistical Modelling and Analysis
WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors
David J. Balding, Noel A.C. Cressie, Garrett M. Fitzmaurice, Harvey Goldstein, Iain M. Johnstone, Geert Molenberghs, David W. Scott, Adrian F.M. Smith, Ruey S. Tsay, Sanford Weisberg

Editors Emeriti
Vic Barnett, Ralph A. Bradley, J. Stuart Hunter, J.B. Kadane, David G. Kendall, Jozef L. Teugels

A complete list of the titles in this series appears at the end of this volume.
1 Introduction
Clair L. Alston, Margaret Donald, Kerrie L. Mengersen
and Anthony N. Pettitt
1.1 Introduction 1
1.2 Overview 1
1.3 Further reading 8
1.3.1 Bayesian theory and methodology 8
1.3.2 Bayesian methodology 10
1.3.3 Bayesian computation 10
1.3.4 Bayesian software 11
1.3.5 Applications 13
References 13

2 Introduction to MCMC
Anthony N. Pettitt and Candice M. Hincksman
2.1 Introduction 17
2.2 Gibbs sampling 18
2.2.1 Example: Bivariate normal 18
2.2.2 Example: Change-point model 19
2.3 Metropolis-Hastings algorithms 19
2.3.1 Example: Component-wise MH or MH within Gibbs 20
2.3.2 Extensions to basic MCMC 21
2.3.3 Adaptive MCMC 22
2.3.4 Doubly intractable problems 22
2.4 Approximate Bayesian computation 24
2.5 Reversible jump MCMC 25
2.6 MCMC for some further applications 26
References 27

3 Priors: Silent or active partners of Bayesian inference?
Samantha Low Choy
3.1 Priors in the very beginning 30
3.1.1 Priors as a basis for learning 32
7.3.2 Meta-analysis models
7.3.3 Computation
7.3.4 Results
7.3.5 Discussion
Acknowledgements
References

8 Bayesian mixed effects models
Clair L. Alston, Christopher M. Strickland, Kerrie L. Mengersen and Graham E. Gardner
8.1 Introduction
8.2 Case studies
  8.2.1 Case study 1: Hot carcase weight of sheep carcases
  8.2.2 Case study 2: Growth of primary school girls
8.3 Models and methods
  8.3.1 Model for Case study 1
  8.3.2 Model for Case study 2
  8.3.3 MCMC estimation
8.4 Data analysis and results
8.5 Discussion
References

9 Ordering of hierarchies in hierarchical models: Bone mineral density estimation
Cathal D. Walsh and Kerrie L. Mengersen
9.1 Introduction
9.2 Case study
  9.2.1 Measurement of bone mineral density
9.3 Models
  9.3.1 Hierarchical model
  9.3.2 Model H1
  9.3.3 Model H2
9.4 Data analysis and results
  9.4.1 Model H1
  9.4.2 Model H2
  9.4.3 Implication of ordering
  9.4.4 Simulation study
  9.4.5 Study design
  9.4.6 Simulation study results
9.5 Discussion
References
9.A Appendix: Likelihoods
10 Bayesian Weibull survival model for gene expression data
Sri Astuti Thamrin, James M. McGree
and Kerrie L. Mengersen

10.1 Introduction 171
10.2 Survival analysis 172
10.3 Bayesian inference for the Weibull survival model 174
  10.3.1 Weibull model without covariates 174
  10.3.2 Weibull model with covariates 175
  10.3.3 Model evaluation and comparison 176
10.4 Case study 178
  10.4.1 Weibull model without covariates 178
  10.4.2 Weibull survival model with covariates 180
  10.4.3 Model evaluation and comparison 182
10.5 Discussion 182
References 183

11 Bayesian change point detection in monitoring clinical outcomes
Hassan Assareh, Ian Smith and Kerrie L. Mengersen

11.1 Introduction 186
11.2 Case study: Monitoring intensive care unit outcomes 187
11.3 Risk-adjusted control charts 187
11.4 Change point model 188
11.5 Evaluation 189
11.6 Performance analysis 190
11.7 Comparison of Bayesian estimator with other methods 194
11.8 Conclusion 194
References 195

12 Bayesian splines
Samuel Clifford and Samantha Low Choy

12.1 Introduction 197
12.2 Models and methods 197
  12.2.1 Splines and linear models 197
  12.2.2 Link functions 198
  12.2.3 Bayesian splines 198
  12.2.4 Markov chain Monte Carlo 204
  12.2.5 Model choice 206
  12.2.6 Posterior diagnostics 207
12.3 Case studies 207
  12.3.1 Data 207
  12.3.2 Analysis 208
12.4 Conclusion 216
   12.4.1 Discussion 216
   12.4.2 Extensions 217
   12.4.3 Summary 218
References 218

13 Disease mapping using Bayesian hierarchical models 221
   Arul Earnest, Susanna M. Cramb and Nicole M. White

13.1 Introduction 221
13.2 Case studies 224
   13.2.1 Case study 1: Spatio-temporal model examining the incidence of birth defects 224
   13.2.2 Case study 2: Relative survival model examining survival from breast cancer 225
13.3 Models and methods 225
   13.3.1 Case study 1 225
   13.3.2 Case study 2 229
13.4 Data analysis and results 230
   13.4.1 Case study 1 230
   13.4.2 Case study 2 231
13.5 Discussion 234
References 237

14 Moisture, crops and salination: An analysis of a three-dimensional agricultural data set 240
   Margaret Donald, Clair L. Alston, Rick Young and Kerrie L. Mengersen

14.1 Introduction 240
14.2 Case study 241
   14.2.1 Data 242
   14.2.2 Aim of the analysis 242
14.3 Review 243
   14.3.1 General methodology 243
   14.3.2 Computations 243
14.4 Case study modelling 243
   14.4.1 Modelling framework 243
14.5 Model implementation: Coding considerations 246
   14.5.1 Neighbourhood matrices and CAR models 246
   14.5.2 Design matrices vs indexing 246
14.6 Case study results 247
14.7 Conclusions 249
References 250
15 A Bayesian approach to multivariate state space modelling: A study of a Fama–French asset-pricing model with time-varying regressors
Christopher M. Strickland and Philip Gharghori

15.1 Introduction 252
15.2 Case study: Asset pricing in financial markets 253
  15.2.1 Data 254
15.3 Time-varying Fama–French model 254
  15.3.1 Specific models under consideration 255
15.4 Bayesian estimation 256
  15.4.1 Gibbs sampler 256
  15.4.2 Sampling $\Sigma_{\epsilon}$ 257
  15.4.3 Sampling $\beta_{1:n}$ 257
  15.4.4 Sampling $\alpha$ 259
  15.4.5 Likelihood calculation 260
15.5 Analysis 261
  15.5.1 Prior elicitation 261
  15.5.2 Estimation output 261
15.6 Conclusion 264
References 265

16 Bayesian mixture models: When the thing you need to know is the thing you cannot measure
Clair L. Alston, Kerrie L. Mengersen and Graham E. Gardner

16.1 Introduction 267
16.2 Case study: CT scan images of sheep 268
16.3 Models and methods 270
  16.3.1 Bayesian mixture models 270
  16.3.2 Parameter estimation using the Gibbs sampler 273
  16.3.3 Extending the model to incorporate spatial information 274
16.4 Data analysis and results 276
  16.4.1 Normal Bayesian mixture model 276
  16.4.2 Spatial mixture model 278
  16.4.3 Carcase volume calculation 281
16.5 Discussion 284
References 284

17 Latent class models in medicine
Margaret Rolfe, Nicole M. White and Carla Chen

17.1 Introduction 287
17.2 Case studies 288
  17.2.1 Case study 1: Parkinson’s disease 288
  17.2.2 Case study 2: Cognition in breast cancer 288
17.3 Models and methods 289
  17.3.1 Finite mixture models 290
  17.3.2 Trajectory mixture models 292
  17.3.3 Goodness of fit 296
  17.3.4 Label switching 297
  17.3.5 Model computation 298
17.4 Data analysis and results 300
  17.4.1 Case study 1: Phenotype identification in PD 300
  17.4.2 Case study 2: Trajectory groups for verbal memory 302
17.5 Discussion 306
References 307

18 Hidden Markov models for complex stochastic processes: A case study in electrophysiology 310
Nicole M. White, Helen Johnson, Peter Silburn, Judith Rousseau and Kerrie L. Mengersen

18.1 Introduction 310
18.2 Case study: Spike identification and sorting of extracellular recordings 311
18.3 Models and methods 312
  18.3.1 What is an HMM? 312
  18.3.2 Modelling a single AP: Application of a simple HMM 313
  18.3.3 Multiple neurons: An application of a factorial HMM 315
  18.3.4 Model estimation and inference 317
18.4 Data analysis and results 320
  18.4.1 Simulation study 320
  18.4.2 Case study: Extracellular recordings collected during deep brain stimulation 323
18.5 Discussion 326
References 327

19 Bayesian classification and regression trees 330
Rebecca A. O’Leary, Samantha Low Choy, Wenbiao Hu and Kerrie L. Mengersen

19.1 Introduction 330
19.2 Case studies 332
  19.2.1 Case study 1: Kyphosis 332
  19.2.2 Case study 2: Cryptosporidium 332
19.3 Models and methods 334
  19.3.1 CARTs 334
  19.3.2 Bayesian CARTs 335
19.4 Computation
   19.4.1 Building the BCART model – stochastic search 337
   19.4.2 Model diagnostics and identifying good trees 339
19.5 Case studies – results
   19.5.1 Case study 1: Kyphosis 341
   19.5.2 Case study 2: Cryptosporidium 343
19.6 Discussion
References 345

20 Tangled webs: Using Bayesian networks in the fight against infection 348
Mary Waterhouse and Sandra Johnson
20.1 Introduction to Bayesian network modelling 348
   20.1.1 Building a BN 349
20.2 Introduction to case study 351
20.3 Model 352
20.4 Methods 354
20.5 Results 355
20.6 Discussion 357
References 359

21 Implementing adaptive dose finding studies using sequential Monte Carlo 361
James M. McGree, Christopher C. Drovandi and Anthony N. Pettitt
21.1 Introduction 361
21.2 Model and priors 363
21.3 SMC for dose finding studies 364
   21.3.1 Importance sampling 364
   21.3.2 SMC 365
   21.3.3 Dose selection procedure 367
21.4 Example 369
21.5 Discussion 371
References 372
21.A Appendix: Extra example 373

22 Likelihood-free inference for transmission rates of nosocomial pathogens 374
Christopher C. Drovandi and Anthony N. Pettitt
22.1 Introduction 374
22.2 Case study: Estimating transmission rates of nosocomial pathogens 375
   22.2.1 Background 375
   22.2.2 Data 376
   22.2.3 Objective 376
22.3 Models and methods 376
  22.3.1 Models 376
  22.3.2 Computing the likelihood 379
  22.3.3 Model simulation 380
  22.3.4 ABC 381
  22.3.5 ABC algorithms 382
22.4 Data analysis and results 384
22.5 Discussion 385
References 386

23 Variational Bayesian inference for mixture models 388
Clare A. McGrory

  23.1 Introduction 388
  23.2 Case study: Computed tomography (CT)
    scanning of a loin portion of a pork carcase 390
  23.3 Models and methods 392
  23.4 Data analysis and results 397
  23.5 Discussion 399
  References 399
  23.A Appendix: Form of the variational posterior
    for a mixture of multivariate normal densities 401

24 Issues in designing hybrid algorithms 403
Jeong E. Lee, Kerrie L. Mengersen and
Christian P. Robert

  24.1 Introduction 403
  24.2 Algorithms and hybrid approaches 406
    24.2.1 Particle system in the MCMC context 407
    24.2.2 MALA 407
    24.2.3 DRA 408
    24.2.4 PS 409
    24.2.5 Population Monte Carlo (PMC) algorithm 410
  24.3 Illustration of hybrid algorithms 412
    24.3.1 Simulated data set 412
    24.3.2 Application: Aerosol particle size 415
  24.4 Discussion 417
  References 418

25 A Python package for Bayesian estimation
  using Markov chain Monte Carlo 421
Christopher M. Strickland, Robert J. Denham,
Clair L. Alston and Kerrie L. Mengersen

  25.1 Introduction 421
  25.2 Bayesian analysis 423
Bayesian statistics is now an established statistical methodology in almost all research disciplines and is being applied to a very wide range of problems. These approaches are endemic in areas of health, the environment, genetics, information science, medicine, biology, industry, remote sensing, and so on. Despite this, most statisticians, researchers and practitioners will not have encountered Bayesian statistics as part of their formal training and often find it difficult to start understanding and employing these methods. As a result of the growing popularity of Bayesian statistics and the concomitant demand for learning about these methods, there is an emerging body of literature on Bayesian theory, methodology, computation and application. Some of this is generic and some is specific to particular fields. While some of this material is introductory, much is at a level that is too complex to be replicated or extrapolated to other problems by an informed Bayesian beginner.

As a result, there is still a need for books that show how to do Bayesian analysis, using real-world problems, at an accessible level.

This book aims to meet this need. Each chapter of this text focuses on a real-world problem that has been addressed by members of our research group, and describes the way in which the problem may be analysed using Bayesian methods. The chapters generally comprise a description of the problem, the corresponding model, the computational method, results and inferences, as well as the issues arising in the implementation of these approaches. In order to meet the objective of making the approaches accessible to the informed Bayesian beginner, the material presented in these chapters is sometimes a simplification of that used in the full projects. However, references are typically given to published literature that provides further details about the projects and/or methods.

This book is targeted at those statisticians, researchers and practitioners who have some expertise in statistical modelling and analysis, and some understanding of the basics of Bayesian statistics, but little experience in its application. As a result, we provide only a brief introduction to the basics of Bayesian statistics and an overview of existing texts and major published reviews of the subject in Chapter 2, along with references for further reading. Moreover, this basic background in statistics and Bayesian concepts is assumed in the chapters themselves.

Of course, there are many ways to analyse a problem. In these chapters, we describe how we approached these problems, and acknowledge that there may be alternatives or improvements. Moreover, there are very many models and a vast number of applications that are not addressed in this book. However, we hope that the material presented here provides a foundation for the informed Bayesian beginner to
engage with Bayesian modelling and analysis. At the least, we hope that beginners will become better acquainted with Bayesian concepts, models and computation, Bayesian ways of thinking about a problem, and Bayesian inferences. We hope that this will provide them with confidence in reading Bayesian material in their own discipline or for their own project. At the most, we hope that they will be better equipped to extend this learning to do Bayesian statistics. As we all learn about, implement and extend Bayesian statistics, we all contribute to ongoing improvement in the philosophy, methodology and inferential capability of this powerful approach.

This book includes an accompanying website. Please visit www.wiley.com/go/statistical_modelling

Clair L. Alston
Kerrie L. Mengersen
Anthony N. Pettitt
List of contributors

Clair L. Alston
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Hassan Assareh
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Carla Chen
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Samuel Clifford
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

David A. Cook
Princess Alexandra Hospital
Brisbane, Australia

Susanna M. Cramb
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Margaret Donald
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Christopher C. Drovandi
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Arul Earnest
Tan Tock Seng Hospital, Singapore &
Duke-NUS Graduate Medical School
Singapore

Graham E. Gardner
School of Veterinary and Biomedical Sciences
Murdoch University
Perth, Australia

Philip Gharghori
Department of Accounting and Finance
Monash University
Melbourne, Australia

Petra L. Graham
Department of Statistics
Macquarie University
North Ryde, Australia

Robert J. Denham
Department of Environment and Resource Management
Brisbane, Australia

Candice M. Hincksman
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia
Wenbiao Hu
School of Population Health and Institute of Health and Biomedical Innovation
University of Queensland
Brisbane, Australia

Katja Ickstadt
Faculty of Statistics
TU Dortmund University
Germany

Helen J Johnson
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Sandra Johnson
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Jonathan M. Keith
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia
and
Monash University
Melbourne, Australia

Jeong E. Lee
School of Computing and Mathematical Sciences
Auckland University of Technology
New Zealand

Samantha Low Choy
Cooperative Research Centre for National Plant Biosecurity, Australia
and
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

James M. McGree
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Clare A. McGrory
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia
and
School of Mathematics
University of Queensland
St. Lucia, Australia

Kerrie L. Mengersen
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Rebecca A. O’Leary
Department of Agriculture and Food
Western Australia, Australia

Anthony N. Pettitt
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Jegar O. Pitchforth
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Christian P. Robert
Université Paris-Dauphine
Paris, France
and
Centre de Recherche en Économie et Statistique (CREST), Paris, France

Margaret Rolfe
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Judith Rousseau
Université Paris-Dauphine
Paris, France
and
Centre de Recherche en Économie et Statistique (CREST), Paris, France
Peter Silburn
St. Andrew’s War Memorial Hospital and Medical Institute
Brisbane, Australia

Ian Smith
St. Andrew’s War Memorial Hospital and Medical Institute
Brisbane, Australia

Christopher M. Strickland
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia

Sri Astuti Thamrin
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia
and
Hasanuddin University, Indonesia

Cathal D. Walsh
Department of Statistics
Trinity College Dublin
Ireland

Mary Waterhouse
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia
and
Wesley Research Institute
Brisbane, Australia

Nicole M. White
School of Mathematical Sciences
Queensland University of Technology
Brisbane, Australia
and
CRC for Spatial Information, Australia

Rick Young
Tamworth Agricultural Institute
Department of Primary Industries
Tamworth, Australia
1

Introduction

Clair L. Alston, Margaret Donald, Kerrie L. Mengersen and Anthony N. Pettitt

Queensland University of Technology, Brisbane, Australia

1.1 Introduction

This book aims to present an introduction to Bayesian modelling and computation, by considering real case studies drawn from diverse fields spanning ecology, health, genetics and finance. As discussed in the Preface, the chapters are intended to be introductory and it is openly acknowledged that there may be many other ways to address the case studies presented here. However, the intention is to provide the Bayesian beginner with a practical and accessible foundation on which to build their own Bayesian solutions to problems encountered in research and practice.

In the following, we first provide an overview of the chapters in the book and then present a list of texts for further reading. This book does not seek to teach the novice about Bayesian statistics per se, nor does it seek to cover the whole field. However, there is now a substantial literature on Bayesian theory, methodology, computation and application that can be used as support and extension. While we cannot hope to cover all of the relevant publications, we provide a selected review of texts now available on Bayesian statistics, in the hope that this will guide the reader to other reference material of interest.

1.2 Overview

In this section we give an overview of the chapters in this book. Given that the models are developed and described in the context of the particular case studies, the first
two chapters focus on the other two primary cornerstones of Bayesian modelling: computational methods and prior distributions. Building on this foundation, Chapters 4–9 describe canonical examples of Bayesian normal linear and hierarchical models. The following five chapters then focus on extensions to the regression models for the analysis of survival, change points, nonlinearity (via splines) and spatial data. The wide class of latent variables models is then illustrated in Chapters 15–19 by considering multivariate linear state space models, mixtures, latent class analysis, hidden Markov models and structural equation models. Chapters 20 and 21 then describe other model structures, namely Bayesian classification and regression trees, and Bayesian networks. The next four chapters of the book focus on different computational methods for solving diverse problems, including approximate Bayesian computation for modelling the transmission of infection, variational Bayes methods for the analysis of remotely sensed data and sequential Monte Carlo to facilitate experimental design. Finally, the last chapter describes a software package, PyMCMC, that has been developed by researchers in our group to provide accessible, efficient Markov chain Monte Carlo algorithms for solving some of the problems addressed in the book.

The chapters are now described in more detail.

Modern Bayesian computation has been hailed as a ‘model-liberating’ revolution in Bayesian modelling, since it facilitates the analysis of a very wide range of models, diverse and complex data sets, and practically relevant estimation and inference. One of the fundamental computational algorithms used in Bayesian analysis is the Markov chain Monte Carlo (MCMC) algorithm. In order to set the stage for the computational approaches described in subsequent chapters, Chapter 2 provides an overview of the Gibbs and Metropolis–Hastings algorithms, followed by extensions such as adaptive MCMC, approximate Bayesian computation (ABC) and reversible jump MCMC (RJ MCMC).

One of the distinguishing features of Bayesian methodology is the use of prior distributions. In Chapter 3 the range of methodology for constructing priors for a Bayesian analysis is described. The approach can broadly be categorized as one of the following two: (i) priors are based on mathematical criteria, such as conjugacy; or (ii) priors model the existing information about the unknown quantity. The chapter shows that in practice a balance must be struck between these two categories. This is illustrated by case studies from the author’s experience. The case studies employ methodology for formulating prior models for different types of likelihood models: binomial, logistic regression, normal and a finite mixture of multivariate normal distributions. The case studies involve the following: time to submit research dissertations; surveillance for exotic plant pests; species distribution models; and delineating ecoregions. There is a review of practical issues. One aim of this chapter is to alert the reader to the important and multi-faceted role of priors in Bayesian inference. The author argues that, in practice, the prior often assumes a silent presence in many Bayesian analyses. Many practitioners or researchers often passively select an ‘inoffensive prior’. This chapter provides practical approaches towards more active selection and evaluation of priors.

Chapter 4 presents the ubiquitous and important normal linear regression model, firstly under the usual assumption of independent, homoscedastic, normal residuals,
and secondly for the situation in which the error covariance matrix is not necessarily diagonal and has unknown parameters. For the latter case, a first-order serial correlation model is considered in detail. In line with the introductory nature of this chapter, two well-known case studies are considered, one involving house prices from a cross-sectional study and the other a time series of monthly vehicle production data from Australia. The theory is extended to the situation where the error covariance matrix is not necessarily diagonal and has unknown parameters, and a first-order serial correlation model is considered in detail. The problem of covariate selection is considered from two perspectives: the stochastic search variable selection approach and a Bayesian lasso. MCMC algorithms are given for the various models. Results are obtained for the two case studies for the fixed model and the variable selection methods.

The application of Bayesian linear regression with informed priors is described in Chapter 5 in the context of modelling patient risk. Risk stratification models are typically constructed via ‘gold-standard’ logistic regressions of health outcomes of interest, often based on a population that has different characteristics to the patient group to which the model is applied. A Bayesian model can augment the local data with priors based on the gold-standard models, resulting in a locally calibrated model that better reflects the target patient group.

A further illustration of linear regression and variable selection is presented in Chapter 6. This concerns a case study involving a genome-wide association (GWA) study. This involves regressing the trait or disease status of interest (a continuous or binary variable) against all the single nucleotide polymorphisms (SNPs) available in order to find the significant SNPs or effects and identify important genes. The case studies involve investigations of genes associated with Type 1 diabetes and breast cancer. Typical SNP studies involve a large number of SNPs and the diabetes study has over 26,000 SNPs while the number of cases is relatively small. A main effects model and an interaction model are described. Bayesian stochastic search algorithms can be used to find the significant effects and the search algorithm to find the important SNPs is described, which uses Gibbs sampling and MCMC. There is an extensive discussion of the results from both case studies, relating the findings to those of other studies of the genetics of these diseases.

The ease with which hierarchical models are constructed in a Bayesian framework is illustrated in Chapter 7 by considering the problem of Bayesian meta-analysis. Meta-analysis involves a systematic review of the relevant literature on the topic of interest and quantitative synthesis of available estimates of the associated effect. For one of the case studies in the chapter this is the association between red meat consumption and the incidence of breast cancer. Formal studies of the association have reported conflicting results, from no association between any level of red meat consumption to a significantly raised relative risk of breast cancer. The second case study is illustrative of a range of problems requiring the synthesis of results from time series or repeated measures studies and involves the growth rate and size of fish. A multivariate analysis is used to capture the dependence between parameters of interest. The chapter illustrates the use of the WinBUGS software to carry out the computations.
Mixed models are a popular statistical model and are used in a range of disciplines to model complex data structures. Chapter 8 presents an exposition of the theory and computation of Bayesian mixed models.

Considering the various models presented to date, Chapter 9 reflects on the need to carefully consider the way in which a Bayesian hierarchical model is constructed. Two different hierarchical models are fitted to data concerning the reduction in bone mineral density (BMD) seen in a sample of patients attending a hospital. In the sample, one of three distinct methods of measuring BMD is used with a patient and patients can be in one of two study groups, either outpatient or inpatient. Hence there are six combinations of data, the three BMD measurement methods and in- or outpatient. The data can be represented by covariates in a linear model, as described in Chapter 2, or can be represented by a nested structure. For the latter, there is a choice of two structures, either method measurement within study group or vice versa, both of which provide estimates of the overall population mean BMD level. The resulting posterior distributions, obtained using WinBUGS, are shown to depend substantially on the model construction.

Returning to regression models, Chapter 10 focuses on a Bayesian formulation of a Weibull model for the analysis of survival data. The problem is motivated by the current interest in using genetic data to inform the probability of patient survival. Issues of model fit, variable selection and sensitivity to specification of the priors are considered.

Chapter 11 considers a regression model tailored to detect change points. The standard model in the Bayesian context provides inferences for a change point and is relatively straightforward to implement in MCMC. The motivation of this study arose from a monitoring programme of mortality of patients admitted to an intensive care unit (ICU) in a hospital in Brisbane, Australia. A scoring system is used to quantify patient mortality based on a logistic regression and the score is assumed to be correct before the change point and changed after by a fixed amount on the odds ratio scale. The problem is set within the context of the application of process control to health care. Calculations were again carried out using WinBUGS software.

The parametric regression models considered so far are extended in Chapter 12 to smoothing splines. Thin-plate splines are discussed in a regression context and a Bayesian hierarchical model is described along with an MCMC algorithm to estimate the parameters. B-splines are described along with an MCMC algorithm and extensions to generalized additive models. The ideas are illustrated with an adaptation to data on the circle (averaged 24 hour temperatures) and other data sets. MATLAB code is provided on the book’s website.

Extending the regression model to the analysis of spatial data, Chapter 13 concerns disease mapping which generally involves modelling the observed and expected counts of morbidity or mortality and expressing each as a ratio, a standardized mortality/morbidity rate (SMR), for an area in a given region. Crude SMRs can have large variances for sparsely populated areas or rare diseases. Models that have spatial correlation are used to smooth area estimates of disease risk and the chapter shows how appropriate Bayesian hierarchical models can be formulated. One case study involves the incidence of birth defects in New South Wales, Australia. A conditional
autoregressive (CAR) model is used for modelling the observed number of defects in an area and various neighbour weightings considered and compared. WinBUGS is used for computation. A second case study involves survival from breast cancer in Queensland and excess mortality, a count, is modelled using a CAR model. Various priors are used and sensitivity analyses carried out. Again WinBUGS is used to estimate the relative excess risk. The approach is particularly useful when there are sparsely populated areas, as is the situation in the two case studies.

The focus on spatial data is continued in Chapter 14 with a description of the analysis carried out to investigate the effects of different cropping systems on the moisture of soil at varying depths up to 300 cm below the surface at 108 different sites, set out in a row by column design. The experiment involved collecting daily data on about 60 occasions over 5 years but here only one day’s data are analysed. The approach uses a Gaussian Markov random field model defined using the CAR formulation to model the spatial dependence for each horizontal level and linear splines to model the smooth change in moisture with depth. The analysis was carried out using the WinBUGS software and the code on the book’s website is described.

Complex data structures can be readily modelled in a Bayesian framework by extending the models considered to data to include latent structures. This concept is illustrated in Chapter 15 by describing a Bayesian analysis for multivariate linear state space modelling. The theory is developed for the Fama–French model of excess return for asset portfolios. For each portfolio the excess return is explained by a linear model with time-varying regression coefficients described by a linear state space model. Three different models are described which allow for different degrees of dependence between the portfolios and across time. A Gibbs algorithm is described for the unknown parameters while an efficient algorithm for simulating from the smoothing distribution for the system parameters is provided. Discrimination between the three possible models is carried out using a likelihood criterion. Efficient computation of the likelihood is also considered. Some results for the regression models for different contrasting types of portfolios are given which confirm the characteristics of these portfolios.

The interest in latent structure models is continued in Chapter 16 with an exposition of mixture distributions, in particular finite normal mixture models. Mixture models can be used as non-parametric density estimates, for cluster analysis and for identifying specific components in a data set. The latent structure in this model indicates mixture components and component membership. A Gibbs algorithm is described for obtaining samples from the posterior distribution. A case study describes the application of mixtures to image analysis for computer tomography (CT) for scans taken from a sheep’s carcase in order to determine the quantities of bone, muscle and fat. The basic model is extended so that the spatial smoothness of the image can be taken into account and a Potts model is used to spatially cluster the different components. A brief description of how the method can be extended to estimate the volume of bone, muscle and fat in a carcase is given. Some practical hints on how to set up the models are also given.

Chapter 17 again involves latent structures, this time through latent class models for clustering subgroups of patients or subjects, leading to identification of meaningful
clinical phenotypes. Between-subject variability can be large and these differences can be modelled by an unobservable, or latent, process. The first case study involves the identification of subgroups for patients suffering from Parkinson’s disease using symptom information. The second case study involves breast cancer patients and their cognitive impairment possibly as a result of therapy. The latent class models involving finite mixture models and trajectory mixture models are reviewed, and various aspects of MCMC implementation discussed. The finite mixture model is used to analyse the Parkinson’s disease data using binary and multinomial models in the mixture. The trajectory mixture model is used with regression models to analyse the cognitive impairment of breast cancer patients. The methods indicate two or three latent classes in the case studies. Some WinBUGS code is provided for the trajectory mixture model on the book’s website.

A related form of latent structure representation, described in Chapter 18, is hidden Markov models (HMMs) which have been extensively developed and used for the analysis of speech data and DNA sequences. Here a case study involves electrophysiology and the application of HMMs to the identification and sorting of action potentials in extracellular recordings involving firing neurons in the brain. Data have been collected during deep brain stimulation, a popular treatment for advanced Parkinson’s disease. The HMM is described in general and in the context of a single neuron firing. An extension to a factorial HMM is considered to model several neurons firing, essentially each neuron having its own HMM. A Gibbs algorithm for posterior simulation is described and applied to simulated data as well as the deep brain stimulation data.

Bayesian models can extend to other constructs to describe complex data structures. Chapter 19 concerns classification and regression trees (CARTs) and, in particular, the Bayesian version, BCART. The BCART model has been found to be highly rated in terms of interpretability. Classification and regression trees give sets of binary rules, repeatedly splitting the predictor variables, to finally end at the predicted value. The case studies here are from epidemiology, concerning a parasite living in the human gut (cryptosporidium), and from medical science, concerning disease of the spine (kyphosis), and extensive analyses of the data sets are given. The CART approach is described and then the BCART is detailed. The BCART approach employs a stochastic search over possible regression trees with different structures and parameters. The original BART employed reversible jump MCMC and is compared with a recent implementation. MATLAB code is available on the book’s website and a discussion on implementation is provided. The kyphosis data set involves a binary indicator for disease for subjects after surgery and a small number of predictor variables. The cryptosporidiosis case study involves predicting incidence rates of the disease. The results of the BCART analyses are described and details of implementation provided.

As another example of alternative model constructs, the idea of a Bayesian network (BN) for modelling the relationship between variables is introduced in Chapter 20. A BN can also be considered as a directed graphical model. Some details about software for fitting BNs are given. A case study concerns MRSA transmission in hospitals (see also Chapter 19). The mechanisms behind MRSA transmission