

# Contents

<b>List of contributors</b>	v
<b>Introducing Highly Structured Stochastic Systems</b>	1
<i>Peter J. Green, Nils Lid Hjort, and Sylvia Richardson</i>	
1 How to read this book	1
2 What are highly structured stochastic systems?	1
2.1 Background	2
2.2 Hierarchical models	2
2.3 Conditional independence	3
2.4 Graphical models	4
2.5 The Bayesian paradigm	4
2.6 Computational methods	5
2.7 Cross-fertilization	6
3 The ESF-funded HSSS Programme	6
4 Structure of the book	8
5 Expectations of the future	11
Acknowledgements	12
References	12
<b>1 Some modern applications of graphical models</b>	13
<i>Steffen L. Lauritzen</i>	
1 Introduction	13
2 Bayesian networks and probability propagation	13
3 Decision support and expert systems	15
3.1 Medical diagnosis	15
3.2 Troubleshooting with Bayesian networks	17
3.3 Decision networks	18
4 Digital communication	21
4.1 Coding and decoding noisy communication channels	21
4.2 Convolutional codes	22
4.3 Iterative decoding and turbocodes	22
5 Machine learning and variational inference	25
6 Other applications	27
References	28

<b>1A</b>	<b>Analysing social science data with graphical Markov models</b>	33
	<i>Nanny Wermuth</i>	
1	Introduction	33
2	A motivating research example	33
3	Some consequences of directed acyclic graphs	35
	3.1 Consequences in simple cases	35
	3.2 Some consequences of large graphs	36
4	Relations to traditional methods and case studies	38
	Acknowledgements	39
<b>1B</b>	<b>Analysis of DNA mixtures using Bayesian networks</b>	39
	<i>Julia Mortera</i>	
	Acknowledgement	43
	Additional references in discussion	43
<b>2</b>	<b>Causal inference using influence diagrams: the problem of partial compliance</b>	45
	<i>A. Philip Dawid</i>	
1	Introduction	45
2	Causal models	45
3	Extended influence diagrams	47
	3.1 Conditional independence	48
4	Partial compliance	49
	4.1 Encouragement trial	49
	4.2 Minimal representation	52
	4.3 Constraints on $\phi$	52
5	Causal inference	54
	5.1 Intervention experiment	54
	5.2 Combination	55
	5.3 Inference	57
	5.4 Bounding the intervention response probabilities	58
	5.5 Bounds for $\alpha$	60
6	Functional representation	60
7	Conclusions	63
	Acknowledgements	63
	References	63
<b>2A</b>	<b>Commentary: causality and statistics</b>	66
	<i>Elja Arjas</i>	
<b>2B</b>	<b>Semantics of causal DAG models and the identification of direct and indirect effects</b>	70
	<i>James M. Robins</i>	
1	Causal models and their DAG representation	70

2	Direct and indirect effects	73
3	Substantive considerations	78
	Additional references in discussion	80
<b>3</b>	<b>Causal inference via ancestral graph models</b>	<b>83</b>
	<i>Thomas S. Richardson and Peter Spirtes</i>	
1	Introduction	83
1.1	Graphical models	83
1.2	Directed acyclic graphs and data-generating processes	83
1.3	Causal models	84
1.4	Inferring causal structure	84
1.5	The problem of unobserved variables	84
1.6	Ancestral graphs	85
1.7	Overview	85
2	Basic graphical terms and concepts	86
3	Graphical Markov models	86
3.1	Example: undirected graphical Markov models	87
3.2	Example: directed acyclic graphical Markov models	87
4	Ancestral graph Markov models	88
4.1	Definition of an ancestral graph	88
4.2	The global Markov property for ancestral graphs	88
4.3	Pairwise Markov property: maximal ancestral graphs	89
4.4	Gaussian parameterization via recursive equations with correlated errors	90
4.5	Gaussian independence models	91
4.6	Future work: the discrete case	92
5	Causal interpretation of DAG models	92
5.1	Causal DAGs	92
5.2	Examples	93
6	Identification of causal effects when the causal DAG is unknown	94
6.1	Marginalizing and conditioning independence models ( $\mathcal{J}_L^s$ )	94
6.2	Marginalizing and conditioning for ancestral graphs	95
6.3	Relating $\mathcal{G}$ and $\mathcal{G}_L^s$	97
6.4	Markov equivalence classes of maximal ancestral graphs	99
6.5	PAGs over four variables	100
7	Relation to other work	102
8	Discussion	102
	Acknowledgements	103
	References	103

<b>3A Other approaches to the description of conditional independence structures</b>	106
<i>Milan Studený</i>	
1 Introduction	106
2 Chain graphs	107
3 Annotated graphs	108
4 Structural insets	108
Acknowledgements	108
<b>3B On ancestral graph Markov models</b>	109
<i>Jan T. A. Koster</i>	
Additional references in discussion	112
<b>4 Causality and graphical models in time series analysis</b>	115
<i>Rainer Dahlhaus and Michael Eichler</i>	
1 Introduction	115
2 Graphical models for multivariate time series	116
2.1 Time series chain graphs	117
2.2 Granger causality graphs	118
2.3 Partial correlation graphs	120
3 Markov properties	121
3.1 Global Markov properties for PC graphs	122
3.2 Global Markov properties for TSC graphs	122
3.3 Global Markov properties for causality graphs	124
3.4 The p-separation criterion	126
4 Inference for graphical time series models	126
4.1 Conditional likelihood for graphical autoregressive models	126
4.2 Partial directed and partial contemporaneous correlation	127
4.3 Partial correlation graphs	128
5 Applications	129
5.1 Application to air pollution data	129
5.2 Application to human tremor data	131
6 Discussion	133
References	134
<b>4A Graphical models for stochastic processes</b>	138
<i>Vanessa Didelez</i>	
1 Introduction	138
2 Granger causality versus intervention causality	138
3 The continuous-time situation	139
<b>4B Conditional association in time series</b>	141
<i>Hans R. Künsch</i>	
1 Concepts of conditional association in time series	141

2	Discovery and interpretation of structure	142
3	Reducing the number of parameters	143
4	Model fitting	143
	Additional references in discussion	144
<b>5</b>	<b>Linking theory and practice of MCMC</b>	<b>145</b>
	<i>Gareth O. Roberts</i>	
1	Introduction	145
2	Motivating example	147
3	Mathematical framework	150
3.1	Rates of convergence of Markov chains	152
4	Drift and minorization conditions	152
4.1	Central limit theorems	154
5	MCMC and drift conditions	155
5.1	Practical implications	156
5.2	Other algorithms	158
6	Examples	159
7	Computable bounds	159
8	Conclusions	162
	Acknowledgements	164
	References	164
<b>5A</b>	<b>Advances in MCMC: a discussion</b>	<b>167</b>
	<i>Christian P. Robert</i>	
1	Introduction	167
2	Parameterization and reparameterization	167
3	Perfect sampling	170
4	Conclusion	171
	Acknowledgements	171
<b>5B</b>	<b>On some current research in MCMC</b>	<b>172</b>
	<i>Arnoldo Frigessi</i>	
1	Introduction	172
2	Adaptive MCMC	172
3	Non-reversible Markov chains	174
4	Post-processing	175
5	More ideas	175
	Acknowledgements	175
	Additional references in discussion	175
<b>6</b>	<b>Trans-dimensional Markov chain Monte Carlo</b>	<b>179</b>
	<i>Peter J. Green</i>	
1	Introduction	179
2	Reversible jump MCMC	180
2.1	Metropolis–Hastings on a general state space	181

2.2	A constructive representation in terms of random numbers	182
2.3	The trans-dimensional case	182
2.4	Details of application to the model-choice problem	183
2.5	Some remarks and ramifications	183
3	Relations to other across-model approaches	184
3.1	Jump diffusion	184
3.2	Point processes, with and without marks	185
3.3	Product-space formulations	185
4	Alternatives to joint model-parameter sampling	186
4.1	Integrating out the parameters	186
4.2	Within-model simulation	187
5	Some issues in choosing a sampling strategy	188
5.1	Is it good to jump?	189
5.2	The two-model case	189
5.3	Finite state space example	190
5.4	Tempering-by-embedding	190
6	An automatic generic trans-dimensional sampler	191
6.1	Examples	193
6.2	Limitations of this approach	194
7	Methodological extensions	194
7.1	Delayed rejection	194
7.2	Efficient proposal choice for reversible jump MCMC	195
7.3	Diagnostics for reversible jump MCMC	195
	Acknowledgements	196
	References	196
<b>6A</b>	<b>Proposal densities, and product-space methods</b>	<b>199</b>
	<i>Simon J. Godsill</i>	
1	Introduction	199
2	Construction of proposal densities	199
3	Product-space methods	201
<b>6B</b>	<b>Trans-dimensional Bayesian non-parametrics with spatial point processes</b>	<b>203</b>
	<i>Juha Heikkinen</i>	
1	Introduction	203
2	Non-parametric Bayesian function estimation	203
3	On constraints and dimension matching	205
	Acknowledgements	206
	Additional references in discussion	206

<b>7</b>	<b>Particle filtering methods for dynamic and static Bayesian problems</b>	<b>207</b>
	<i>Carlo Berzuini and Walter R. Gilks</i>	
1	Introduction	207
2	Dynamic Bayesian problems	208
3	Motivating example	209
4	General particle filtering strategy	211
5	The Sampling Importance Resampling filter	213
6	Why resample?	214
7	Efficient resampling	215
8	Impoverishment	215
9	Introducing Markov iterations	216
10	Enhancing particle diversity (resampling and moving)	217
11	An illustration	221
12	Particle filtering for static problems	222
	Acknowledgements	224
	References	224
<b>7A</b>	<b>Some further topics on Monte Carlo methods for dynamic Bayesian problems</b>	<b>228</b>
	<i>Geir Storvik</i>	
1	Introduction	228
2	When to resample	228
3	Theoretical properties of Monte Carlo filters	230
4	Static parameters	231
<b>7B</b>	<b>General principles in sequential Monte Carlo methods</b>	<b>232</b>
	<i>Peter Clifford</i>	
1	Particle filters	232
2	Generalizations and enhancements	234
2.1	Monte Carlo integration	234
2.2	Regularized sampling and effective dimension	234
2.3	Parametric particle filters	234
2.4	Self-avoiding particles	235
3	Bearings-only tracking	236
	Additional references in discussion	236
<b>8</b>	<b>Spatial models in epidemiological applications</b>	<b>237</b>
	<i>Sylvia Richardson</i>	
1	Introduction	237
2	Level of spatial modelling for disease outcomes	239
2.1	Types of data	239
2.2	Hierarchical models	240
3	Models for point data	240
3.1	Framework	240

3.2	Inhomogeneous Poisson processes	241
3.3	Cox processes	242
3.4	Spatial variability of the disease rate	243
4	Aggregation of point level models	244
4.1	Stratification	244
4.2	Models linked to Gaussian random fields (GRFs)	245
4.3	Models linked to gamma random fields	245
5	Models directly specified at the aggregated level	246
5.1	Markov random fields models	246
5.2	Direct modelling of covariance structure	247
5.3	Partition, allocation, and mixture models	248
6	Covariate effects in spatial models	249
6.1	Point process models	250
6.2	Aggregation and ecological bias	251
6.3	Covariate effects in models specified at the aggregated level	252
6.4	Confounding between covariates and random effects	252
7	Concluding remarks	254
	Acknowledgements	255
	References	255
<b>8A</b>	<b>Some remarks on Gaussian Markov random field models for disease mapping</b>	<b>260</b>
	<i>Leonhard Knorr-Held</i>	
	Acknowledgements	264
<b>8B</b>	<b>A comparison of spatial point process models in epidemiological applications</b>	<b>264</b>
	<i>Jesper Møller</i>	
1	Brief description of models	264
2	General properties of the models	265
3	Aggregation coherence and natural scaling	267
4	Computational aspects	267
5	Concluding remarks	268
	Acknowledgements	268
	Additional references in discussion	268
<b>9</b>	<b>Spatial hierarchical Bayesian models in ecological applications</b>	<b>271</b>
	<i>Antti Penttinen, Fabio Divino, and Anne Riihali</i>	
1	Introduction	271
2	Modelling the suitability of sites	273
3	Extension of the suitability model to two species	275
4	A conditional model for association of species	276
5	Empirical results	280



6	Discussion	284
	Acknowledgements	286
	References	286
<b>9A</b>	<b>Likelihood analysis of binary data in space and time</b>	<b>289</b>
	<i>Julian Besag and Jeremy Tantrum</i>	
1	Introduction	289
2	Markov random field formulation	290
3	Parameter estimation	291
	3.1 Maximum pseudo-likelihood estimates for the $\theta_t$ s	291
	3.2 Maximum likelihood estimates for the $\theta_t$ s	292
	3.3 Time homogeneity	294
4	Discussion	294
	Acknowledgements	295
<b>9B</b>	<b>Some further aspects of spatio-temporal modelling</b>	<b>295</b>
	<i>Alexandra Mello Schmidt</i>	
1	Introduction	295
2	Modelling of heterogeneous spatial processes	295
3	Non-separable covariance matrices	297
	Acknowledgement	298
	Additional references in discussion	298
<b>10</b>	<b>Advances in Bayesian image analysis</b>	<b>301</b>
	<i>Merrilee Hurn, Oddvar Husbym, and Håvard Rue</i>	
1	Introduction	301
2	Modelling	301
	2.1 Prior modelling	302
	2.2 Data modelling	309
3	Treatment of parameters	309
	3.1 Estimation techniques for low-level models	310
	3.2 Estimation techniques for high-level models	310
	3.3 Fully Bayesian approaches	310
4	Point estimation for images	311
5	Example	312
	5.1 A model for diffuse scattering in ultrasound imaging	313
	5.2 Image restoration and MCMC algorithms	314
	5.3 Contour detection	315
	Acknowledgements	316
	References	316
<b>10A</b>	<b>Probabilistic image modelling</b>	<b>323</b>
	<i>M. N. M. van Lieshout</i>	
1	Introduction	323
2	Pixel-level modelling	323

3	High-level modelling	324
4	Concluding remarks	325
<b>10B</b>	<b>Prospects in Bayesian image analysis</b>	<b>326</b>
	<i>Alain Trubuil</i>	
1	Introduction	326
2	Texture modelling	326
3	Level curves and segmentation	327
4	Applications	328
	Additional references in discussion	330
<b>11</b>	<b>Preventing epidemics in heterogeneous environments</b>	<b>333</b>
	<i>Niels G. Becker and Sergey Utev</i>	
1	Introduction	333
2	The standard epidemic model	335
2.1	Model formulation	335
2.2	A threshold parameter	336
2.3	Critical immunity coverage	336
2.4	Estimation	337
2.5	Dependence, heterogeneity, and structure	338
3	Variation in infectivity and susceptibility	338
3.1	Unidentifiable types	338
3.2	Identifiable types	341
3.3	Variable vaccine response	343
4	Community of households	346
5	Contact networks and other models	348
5.1	Disease transmission via contact sets	348
5.2	Variation over time	349
5.3	Other models	350
	Acknowledgements	350
	References	350
<b>11A</b>	<b>MCMC methods for stochastic epidemic models</b>	<b>354</b>
	<i>Philip D. O'Neill</i>	
1	Introduction	354
1.1	Advantages and challenges of MCMC	354
2	Examples	355
2.1	Temporal data, general stochastic epidemic	355
2.2	Temporal data, network model	356
2.3	Final size data, households model	356
2.4	Future directions	357
<b>11B</b>	<b>Towards Bayesian inference in epidemic models</b>	<b>358</b>
	<i>Kari Auranen</i>	
	Additional references in discussion	361

<b>12 Genetic linkage analysis using Markov chain Monte Carlo techniques</b>	<b>363</b>
<i>Simon C. Heath</i>	
1 Introduction	363
1.1 Markers and trait loci	365
1.2 The computational difficulties of linkage analysis	365
1.3 Relationship of linkage to linkage disequilibrium analysis	366
2 Exact probability calculations	367
2.1 The peeling algorithm	367
2.2 The Lander–Green algorithm	369
3 Markov chain Monte Carlo sampling schemes	373
3.1 Single locus genotype samplers	373
3.2 Multilocus genotype samplers	374
3.3 Sampling segregation indicators	375
3.4 General genetic samplers	375
3.5 Samplers for imperfect data	376
4 Sampling trait locus number	377
5 Discussion	377
References	378
<b>12A Graphical models for mapping continuous traits</b>	<b>382</b>
<i>Nuala A. Sheehan and Daniel A. Sorensen</i>	
1 Introduction	382
2 QTL mapping on a granddaughter design	382
3 Conclusion	386
Acknowledgements	386
<b>12B Statistical approaches to genetic mapping</b>	<b>386</b>
<i>David A. Stephens</i>	
1 Association mapping and genome-wide maps	388
2 Future developments: data mining approaches to genetic mapping	389
Additional references in discussion	391
<b>13 The genealogy of a neutral mutation</b>	<b>393</b>
<i>R. C. Griffiths and Simon Tavaré</i>	
1 Introduction	393
1.1 Coalescent trees	394
1.2 Variable population size	395
2 Relationship between coalescent trees and urn models	396
2.1 The number of descendants of an edge	397
2.2 Coalescence times in a subtree	397
2.3 Branches in the subtree below a mutation	398
3 Mutations in the tree	399
3.1 The distribution of the number of segregating sites	400

3.2	The distribution of pairwise differences	400
3.3	Tajima's $D$	400
4	Frequency spectra	401
5	Distribution of the age of a mutation	402
6	Coalescence times in a subtree	403
7	Further mutations in subtrees	405
7.1	A simulation algorithm	405
7.2	The age of the mutation	405
7.3	An example of simulation of ages	406
7.4	Simulation of Tajima's $D$ in a subtree	407
8	Sampling under a mutation in the population	407
8.1	Results for the standard coalescent	408
9	Other sampling schemes	409
9.1	Sampling from the disease population	409
9.2	Case-control sampling	410
10	Conclusion	411
	Acknowledgements	411
	References	412
<b>13A</b>	<b>Linked versus unlinked DNA data – a comparison based on ancestral inference</b>	414
	<i>Gunter Weiss</i>	
1	Estimation of a scaled mutation rate parameter	414
2	Estimation of a growth rate parameter	415
3	An illustrative example	416
<b>13B</b>	<b>The age of a rare mutation</b>	417
	<i>Carsten Wiuf</i>	
1	Introduction	417
2	Defining the problem	419
3	The likelihood approach	419
4	The coalescent approach	420
5	Conclusion	421
	Additional references in discussion	421
<b>14</b>	<b>HSSS model criticism</b>	423
	<i>Anthony O'Hagan</i>	
1	Introduction	423
1.1	The role of model criticism	423
1.2	Model adequacy and sensitivity analysis	424
1.3	Model criticism	425
1.4	The origins and structure of this article	427
2	Predictive diagnostic techniques	428
2.1	Comparing data with model predictions	428
2.2	Predictive diagnostics	430

3	Node-based methods	431
3.1	Residuals	432
3.2	Conflict	434
3.3	Example	436
3.4	Interpretation of conflict	438
3.5	Some related issues	439
4	Sensitivity to prior distributions	440
4.1	Diagnosing potential sensitivity	440
4.2	Example	442
5	Discussion	443
	Acknowledgements	443
	References	443
<b>14A</b>	<b>Which ‘base’ distribution for model criticism?</b>	<b>445</b>
	<i>M. J. Bayarri</i>	
1	Basic issues in model criticism	445
2	Distributions for the diagnostic measure	446
3	An example (Bayarri and Morales 2003)	447
4	Concluding remarks	448
<b>14B</b>	<b>Some comments on model criticism</b>	<b>449</b>
	<i>Alan E. Gelfand</i>	
	Additional references in discussion	452
<b>15</b>	<b>Topics in non-parametric Bayesian statistics</b>	<b>455</b>
	<i>Nils Lid Hjort</i>	
1	Introduction and summary	455
1.1	What is it, and why?	455
1.2	On the present article	457
2	The Dirichlet process prior and some extensions	457
2.1	The Dirichlet process	457
2.2	The Dirichlet as a limit	458
2.3	An extension via an infinite sum representation	459
3	Random Dirichlet means	459
3.1	Transform identities	459
3.2	Stochastic equations and the full moment sequence	460
4	Quantile pyramid processes	461
4.1	Quantile inference with the Dirichlet process	461
4.2	An automatic non-parametric density estimator	462
4.3	Quantile pyramids	463
4.4	Posterior quantile pyramids	463
5	Bayesian density estimation and consistency issues	464
6	Lévy frailty processes and proportional hazards	466
7	Beta processes in a linear hazard model	468
8	Local Bayesian regression	469

9	Random shapes with smoothed Dirichlets	471
10	Non-parametric envelopes around parametric models	472
10.1	A semiparametric Bayes model	472
10.2	Model fitting with control sets	473
10.3	Randomness around a parametric survival data model	473
11	Concluding remarks	474
	Acknowledgements	474
	References	475
<b>15A</b>	<b>Asymptotics of non-parametric posteriors</b>	<b>479</b>
	<i>Aad van der Vaart</i>	
1	Introduction	479
2	Questions	479
2.1	Semiparametric versus non-parametric	479
2.2	Really high-dimensional problems	480
2.3	Proper and fixed	481
2.4	Model selection	481
3	Rates of convergence	481
<b>15B</b>	<b>A predictive point of view on Bayesian non-parametrics</b>	<b>483</b>
	<i>Sonia Petrone</i>	
1	An example	483
2	A predictive approach	484
2.1	Non-parametric priors for high-dimensional data	484
2.2	Characterization of the prior via predictive assumptions	484
3	Goodness of fit	485
4	Computational issues and asymptotic behaviour	485
5	Problems of dependence	486
	Additional references in discussion	486
	<b>Author index</b>	<b>489</b>
	<b>Subject index</b>	<b>503</b>