

Contents

<i>Preface</i>	xvii
<i>Contributors</i>	xix
A METHODOLOGICAL CHAPTERS	1
1 Two Challenges of Systems Biology	3
<i>William S. Hlavacek</i>	
1.1 Introduction	3
1.2 Cell signaling systems	4
1.3 The challenge of many moving parts	5
1.4 The challenge of parts with parts	7
1.5 Closing remarks	9
References	10
2 Introduction to Statistical Methods for Complex Systems	15
<i>Tristan Mary-Huard and Stéphane Robin</i>	
2.1 Introduction	15
2.2 Class comparison	16
2.2.1 Models for dependent data	16
2.2.2 Multiple testing	19
2.3 Class prediction	22
2.3.1 Building a classifier	22
2.3.2 Aggregation	25
2.3.3 Regularization	27
2.3.4 Performance assessment	29
2.4 Class discovery	31
2.4.1 Geometric methods	31
2.4.2 (Discrete) latent variable models	32
2.4.3 Inference	33
References	36
3 Bayesian Inference and Computation	39
<i>Christian P. Robert, Jean-Michel Marin and Judith Rousseau</i>	
3.1 Introduction	39
3.2 The Bayesian argument	39
3.2.1 Bases	39
3.2.2 Bayesian analysis in action	41

3.2.3	Prior distributions	42
3.2.4	Confidence intervals	46
3.3	Testing hypotheses	48
3.3.1	Decisions	48
3.3.2	The Bayes factor	48
3.3.3	Point null hypotheses	49
3.3.4	The ban on improper priors	50
3.3.5	The case of nuisance parameters	52
3.3.6	Bayesian multiple testing	54
3.4	Extensions	55
3.4.1	Prediction	55
3.4.2	Outliers	56
3.4.3	Model choice	56
3.5	Computational issues	57
3.5.1	Computational challenges	57
3.5.2	Monte Carlo methods	59
3.5.3	MCMC methods	61
3.5.4	Approximate Bayesian computation techniques	63
	Acknowledgements	64
	References	64
4	Data Integration: Towards Understanding Biological Complexity	66
	<i>David Gomez-Cabrero and Jesper Tegner</i>	
4.1	Storing knowledge: Experimental data, knowledge databases, ontologies and annotation	66
4.1.1	Data repositories	67
4.1.2	Knowledge Databases	68
4.1.3	Ontologies	70
4.1.4	Annotation	71
4.2	Data integration in biological studies	74
4.2.1	Integration of experimental data	74
4.2.2	Ontologies and experimental data	77
4.2.3	Networks and visualization software as integrative tools	77
4.3	Concluding remarks	77
	References	78
5	Control Engineering Approaches to Reverse Engineering Biomolecular Networks	83
	<i>Francesco Montefusco, Carlo Cosentino and Declan G. Bates</i>	
5.1	Dynamical models for network inference	83
5.1.1	Linear models	84
5.1.2	Nonlinear models	85
5.2	Reconstruction methods based on linear models	89
5.2.1	Least squares	89
5.2.2	Methods based on least squares	90
5.2.3	Dealing with noise: CTLS	91
5.2.4	Convex optimization methods	97

5.2.5	Sparsity pattern of the discrete-time model	100
5.2.6	Application examples	101
5.3	Reconstruction methods based on nonlinear models	104
5.3.1	Approaches based on polynomial and rational models	105
5.3.2	Approaches based on S-systems	107
5.3.3	A case-study	109
	References	111
6	Algebraic Statistics and Methods in Systems Biology	114
	<i>Carsten Wiuf</i>	
6.1	Introduction	114
6.2	Overview of chapter	115
6.3	Computational algebra	116
6.4	Algebraic statistical models	118
6.4.1	Definitions	118
6.4.2	Further examples	119
6.5	Parameter inference	122
6.6	Model invariants	124
6.7	Log-linear models	126
6.8	Reverse engineering of networks	129
6.9	Concluding remarks	130
	References	130
B	TECHNOLOGY-BASED CHAPTERS	133
7	Transcriptomic Technologies and Statistical Data Analysis	135
	<i>Elizabeth Purdom and Sach Mukherjee</i>	
7.1	Biological background	135
7.2	Technologies for genome-wide profiling of transcription	136
7.2.1	Microarray technology	136
7.2.2	mRNA expression estimates from microarrays	137
7.2.3	High throughput sequencing (HTS)	137
7.2.4	mRNA expression estimates from HTS	139
7.3	Evaluating the significance of individual genes	140
7.3.1	Common approaches for significance testing	140
7.3.2	Moderated statistics	141
7.3.3	Statistics for HTS	142
7.3.4	Multiple testing corrections	143
7.3.5	Filtering genes	145
7.4	Grouping genes to find biological patterns	147
7.4.1	Gene-set analysis	147
7.4.2	Dimensionality reduction	149
7.4.3	Clustering	150
7.5	Prediction of a biological response	153
7.5.1	Variable selection	153
7.5.2	Estimating the performance of a model	156
	References	157

8	Statistical Data Analysis in Metabolomics	163
	<i>Timothy M. D. Ebbels and Maria De Iorio</i>	
8.1	Introduction	163
8.2	Analytical technologies and data characteristics	164
	8.2.1 Analytical technologies	164
	8.2.2 Preprocessing	166
8.3	Statistical analysis	169
	8.3.1 Unsupervised methods	169
	8.3.2 Supervised methods	171
	8.3.3 Metabolome-wide association studies	172
	8.3.4 Metabolic correlation networks	173
	8.3.5 Simulation of metabolic profile data	176
8.4	Conclusions	178
	Acknowledgements	178
	References	178
9	Imaging and Single-Cell Measurement Technologies	181
	<i>Yu-ichi Ozaki and Shinya Kuroda</i>	
9.1	Introduction	181
	9.1.1 Intracellular signal transduction	181
	9.1.2 Lysate-based assay and single-cell assay	182
	9.1.3 Live cell and fixed cell	182
9.2	Measurement techniques	182
	9.2.1 Western blot analysis	183
	9.2.2 Immunocytochemistry	183
	9.2.3 Flow cytometry	184
	9.2.4 Fluorescent microscope	185
	9.2.5 Live cell imaging	187
	9.2.6 Fluorescent probes for live cell imaging	188
	9.2.7 Image cytometry	190
	9.2.8 Image processing	191
9.3	Analysis of signal cell measurement data	194
	9.3.1 Time series (mean, variation, correlation, localization)	194
	9.3.2 Bayesian network modeling with single-cell data	197
	9.3.3 Quantifying sources of cell-to-cell variation	197
9.4	Summary	197
	Acknowledgements	199
	References	199
10	Protein Interaction Networks and Their Statistical Analysis	200
	<i>Waqar Ali, Charlotte Deane and Gesine Reinert</i>	
10.1	Introduction	200
10.2	Proteins and their interactions	201
	10.2.1 Protein structure and function	201
	10.2.2 Protein–protein interactions	202

10.2.3	Experimental techniques for interaction detection	202
10.2.4	Computationally predicted data-sets	203
10.2.5	Protein interaction databases	204
10.2.6	Error in PPI data	204
10.2.7	The interactome concept and protein interaction networks	205
10.3	Network analysis	205
10.3.1	Graphs	205
10.3.2	Network summary statistics	206
10.3.3	Network motifs	207
10.3.4	Models of random networks	207
10.3.5	Parameter estimation for network models	209
10.3.6	Approximate Bayesian Computation	209
10.3.7	Threshold behaviour in graphs	210
10.4	Comparison of protein interaction networks	211
10.4.1	Network comparison based on subgraph counts	211
10.4.2	Network alignment	213
10.4.3	Using functional annotation for network alignment	215
10.5	Evolution and the protein interaction network	217
10.5.1	How evolutionary models affect network alignment	217
10.6	Community detection in PPI networks	218
10.6.1	Community detection methods	219
10.6.2	Evaluation of results	220
10.7	Predicting function using PPI networks	221
10.8	Predicting interactions using PPI networks	223
10.8.1	Tendency to form triangles	224
10.8.2	Using triangles for predicting interactions	224
10.9	Current trends and future directions	226
10.9.1	Dynamics	226
10.9.2	Integration with other networks	227
10.9.3	Limitations of models, prediction and alignment methods	227
10.9.4	Biases, error and weighting	228
10.9.5	New experimental sources of PPI data	228
	References	228
C	NETWORKS AND GRAPHICAL MODELS	235
11	Introduction to Graphical Modelling	237
	<i>Marco Scutari and Korbinian Strimmer</i>	
11.1	Graphical structures and random variables	237
11.2	Learning graphical models	241
11.2.1	Structure learning	242
11.2.2	Parameter learning	246
11.3	Inference on graphical models	246
11.4	Application of graphical models in systems biology	247
11.4.1	Correlation networks	247
11.4.2	Covariance selection networks	248

11.4.3	Bayesian networks	250
11.4.4	Dynamic Bayesian networks	250
11.4.5	Other graphical models	250
	References	251
12	Recovering Genetic Network from Continuous Data with Dynamic Bayesian Networks	255
	<i>Gaëlle Lelandais and Sophie Lèbre</i>	
12.1	Introduction	255
12.1.1	Regulatory networks in biology	255
12.1.2	Objectives and challenges	256
12.2	Reverse engineering time-homogeneous DBNs	256
12.2.1	Genetic network modelling with DBNs	256
12.2.2	DBN for linear interactions and inference procedures	259
12.3	Go forward: how to recover the structure changes with time	261
12.3.1	ARTIVA network model	262
12.3.2	ARTIVA inference procedure and performance evaluation	263
12.4	Discussion and Conclusion	267
	References	268
13	Advanced Applications of Bayesian Networks in Systems Biology	270
	<i>Dirk Husmeier, Adriano V. Werhli and Marco Grzegorzczak</i>	
13.1	Introduction	270
13.1.1	Basic concepts	270
13.1.2	Dynamic Bayesian networks	272
13.1.3	Bayesian learning of Bayesian networks	273
13.2	Inclusion of biological prior knowledge	273
13.2.1	The ‘energy’ of a network	274
13.2.2	Prior distribution over network structures	275
13.2.3	MCMC sampling scheme	276
13.2.4	Practical implementation	277
13.2.5	Empirical evaluation on the Raf signalling pathway	277
13.3	Heterogeneous DBNs	281
13.3.1	Motivation: Inferring spurious feedback loops with DBNs	281
13.3.2	A nonlinear/nonhomogeneous DBN	282
13.3.3	MCMC inference	284
13.3.4	Simulation results	284
13.3.5	Results on Arabidopsis gene expression time series	285
13.4	Discussion	287
	Acknowledgements	288
	References	288
14	Random Graph Models and Their Application to Protein–Protein Interaction Networks	290
	<i>Desmond J. Higham and Nataša Pržulj</i>	
14.1	Background and motivation	290
14.2	What do we want from a PPI network?	293
14.3	PPI network models	294

14.3.1	Lock and key	294
14.3.2	Geometric networks	297
14.4	Range-dependent graphs	301
14.5	Summary	305
	References	306
15	Modelling Biological Networks via Tailored Random Graphs	309
	<i>Anthony C. C. Coolen, Franca Fraternali, Alessia Annibale, Luis Fernandes and Jens Kleinjung</i>	
15.1	Introduction	309
15.2	Quantitative characterization of network topologies	310
15.2.1	Local network features and their statistics	310
15.2.2	Examples	311
15.3	Network families and random graphs	312
15.3.1	Network families, hypothesis testing and null models	312
15.3.2	Tailored random graph ensembles	313
15.4	Information-theoretic deliverables of tailored random graphs	315
15.4.1	Network complexity	315
15.4.2	Information-theoretic dissimilarity	316
15.5	Applications to PPINs	317
15.5.1	PPIN assortativity and wiring complexity	320
15.5.2	Mapping PPIN data biases	320
15.6	Numerical generation of tailored random graphs	323
15.6.1	Generating random graphs via Markov chains	323
15.6.2	Degree-constrained graph dynamics based on edge swaps	324
15.6.3	Numerical examples	325
15.7	Discussion	325
	References	327
D	DYNAMICAL SYSTEMS	331
16	Nonlinear Dynamics: A Brief Introduction	333
	<i>Alessandro Moura and Celso Grebogi</i>	
16.1	Introduction	333
16.2	Sensitivity to initial conditions and the Lyapunov exponent	334
16.3	The natural measure	334
16.4	The Kolmogorov–Sinai entropy	335
16.5	Symbolic dynamics	336
16.6	Chaos in biology	338
	References	338
17	Qualitative Inference in Dynamical Systems	339
	<i>Fatihcan M. Atay and Jürgen Jost</i>	
17.1	Introduction	339
17.2	Basic solution types	343

17.3	Qualitative behaviour	346
17.4	Stability and bifurcations	347
17.5	Ergodicity	353
17.6	Timescales	354
17.7	Time series analysis	356
	References	357
18	Stochastic Dynamical Systems	359
	<i>Darren J. Wilkinson</i>	
18.1	Introduction	359
18.2	Origins of stochasticity	359
	18.2.1 Low copy number	359
	18.2.2 Other sources of noise and heterogeneity	360
18.3	Stochastic chemical kinetics	360
	18.3.1 Reaction networks	360
	18.3.2 Markov jump process	360
	18.3.3 Diffusion approximation	364
	18.3.4 Reaction rate equations	365
	18.3.5 Modelling extrinsic noise	365
18.4	Inference for Markov process models	366
	18.4.1 Likelihood-based inference	366
	18.4.2 Partial observation and data augmentation	367
	18.4.3 Data augmentation MCMC approaches	368
	18.4.4 Likelihood-free approaches	369
	18.4.5 Approximate Bayesian computation	370
	18.4.6 Particle MCMC	371
	18.4.7 Iterative filtering	371
	18.4.8 Stochastic model emulation	371
	18.4.9 Inference for stochastic differential equation models	372
18.5	Conclusions	372
	Acknowledgements	373
	References	373
19	Gaussian Process Inference for Differential Equation Models of Transcriptional Regulation	376
	<i>Neil Lawrence, Magnus Rattray, Antti Honkela and Michalis Titsias</i>	
19.1	Introduction	376
	19.1.1 A simple systems biology model	377
19.2	Generalized linear model	379
	19.2.1 Fitting basis function models	380
	19.2.2 An infinite basis	383
	19.2.3 Gaussian processes	385
	19.2.4 Sampling approximations	387
19.3	Model based target ranking	387
19.4	Multiple transcription factors	391
19.5	Conclusion	393
	References	394

20	Model Identification by Utilizing Likelihood-Based Methods	395
	<i>Andreas Raue and Jens Timmer</i>	
20.1	ODE models for reaction networks	396
	20.1.1 Rate equations	397
20.2	Parameter estimation	398
	20.2.1 Sensitivity equations	399
	20.2.2 Testing hypothesis	400
	20.2.3 Confidence intervals	401
20.3	Identifiability	403
	20.3.1 Structural nonidentifiability	403
	20.3.2 Practical nonidentifiability	404
	20.3.3 Connection of identifiability and observability	405
20.4	The profile likelihood approach	405
	20.4.1 Experimental design	406
	20.4.2 Model reduction	407
	20.4.3 Observability and confidence intervals of trajectories	407
	20.4.4 Application	408
20.5	Summary	413
	Acknowledgements	414
	References	415
E	APPLICATION AREAS	417
21	Inference of Signalling Pathway Models	419
	<i>Tina Toni, Juliane Liepe and Michael P. H. Stumpf</i>	
21.1	Introduction	419
21.2	Overview of inference techniques	420
21.3	Parameter inference and model selection for dynamical systems	422
	21.3.1 Model selection	424
21.4	Approximate Bayesian computation	425
21.5	Application: Akt signalling pathway	426
	21.5.1 Exploring different distance functions	428
	21.5.2 Posteriors	430
	21.5.3 Parameter sensitivity through marginal posterior distributions	430
	21.5.4 Sensitivity analysis by principal component analysis (PCA)	430
21.6	Conclusion	435
	References	436
22	Modelling Transcription Factor Activity	440
	<i>Martino Barenco, Daniel Brewer, Robin Callard and Michael Hubank</i>	
22.1	Integrating an ODE with a differential operator	441
22.2	Computation of the entries of the differential operator	443
	22.2.1 Taking into account the nature of the biological system being modelled	443
	22.2.2 Bounds choice for polynomial interpolation	445

22.3	Applications	447
22.4	Estimating intermediate points	449
	Acknowledgements	450
	References	450
23	Host–Pathogen Systems Biology	451
	<i>John W. Pinney</i>	
23.1	Introduction	451
23.2	Pathogen genomics	453
23.3	Metabolic models	453
23.4	Protein–protein interactions	455
23.5	Response to environment	457
23.6	Immune system interactions	458
23.7	Manipulation of other host systems	459
23.8	Evolution of the host–pathogen system	460
23.9	Towards systems medicine for infectious diseases	462
23.10	Concluding remarks	462
	Acknowledgements	463
	References	463
24	Bayesian Approaches for Mass Spectrometry-Based Metabolomics	467
	<i>Simon Rogers, Richard A. Scheltema, Michael Barrett and Rainer Breitling</i>	
24.1	Introduction	467
24.2	The challenge of metabolite identification	468
24.3	Bayesian analysis of metabolite mass spectra	469
24.4	Incorporating additional information	471
24.5	Probabilistic peak detection	472
24.6	Statistical inference	473
24.7	Software development for metabolomics	474
24.8	Conclusion	475
	References	475
25	Systems Biology of microRNAs	477
	<i>Doron Betel and Raya Khanin</i>	
25.1	Introduction	477
25.2	Current approaches in microRNA Systems Biology	477
25.3	Experimental findings and data that guide the developments of computational tools	478
25.4	Approaches to microRNA target predictions	479
25.5	Analysis of mRNA and microRNA expression data	482
	25.5.1 Identifying microRNA activity from mRNA expression	482
	25.5.2 Modeling combinatorial microRNA regulation from joint microRNA and mRNA expression data	484
25.6	Network approach for studying microRNA-mediated regulation	485
25.7	Kinetic modeling of microRNA regulation	486
	25.7.1 A basic model of microRNA-mediated regulation	487

25.7.2 Estimating fold-changes of mRNA and proteins in microRNA transfection experiments	488
25.7.3 The influence of protein and mRNA stability on microRNA function	489
25.7.4 microRNA efficacy depends on target abundance	489
25.7.5 Reconstructing microRNA kinetics	489
25.8 Discussion	490
References	491
<i>Index</i>	495

