

Chapman & Hall/CRC Interdisciplinary Statistics Series

Applied Directional Statistics: Modern Methods and Case Studies collects important advances in methodology and data analysis for directional statistics over the last two decades. All contributors are experts in the field and the book is meant to be useful to both researchers and practitioners. It is the companion book of the more theoretical treatment presented in *Modern Directional Statistics* (CRC Press, 2017).

The field of directional statistics has received a lot of attention over the past two decades, due to new demands from disciplines such as life sciences or machine learning, the availability of massive data sets requiring adapted statistical techniques, and technological advances. This book covers important progress in structural bioinformatics, biology, astrophysics, oceanography, environmental sciences, earth sciences, machine learning and social sciences.

Christophe Ley is professor of mathematical statistics at Ghent University. His research interests include semi-parametrically efficient inference, flexible modeling, directional statistics, sport statistics and the study of asymptotic approximations via Stein's Method. His achievements include the Marie-Jeanne Laurent-Duhamel prize of the Société Française de Statistique and an elected membership of the International Statistical Institute. He is associate editor for the journals *Computational Statistics & Data Analysis*, *Annals of the Institute of Statistical Mathematics*, and *Econometrics and Statistics*.

Thomas Verdebout is professor of mathematical statistics at Université libre de Bruxelles (ULB). His main research interests are semi-parametric statistics, high-dimensional statistics, directional statistics and rank-based procedures. He has won an annual prize of the Belgian Academy of Sciences and is an elected member of the International Statistical Institute. He is associate editor for the journals *Statistics and Probability Letters* and *Journal of Multivariate Analysis*.



CRC Press
Taylor & Francis Group
an informa business

www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K31852

ISBN: 978-1-138-62643-0



9 781138 626430

www.crcpress.com

Contributors	xi
Editors	xiii
Introduction	xv
1 Directional Statistics in Protein Bioinformatics	1
<i>Kanti V. Mardia, Jesper Illemaan Foldager, and Jes Frellsen</i>	
1.1 Introduction	1
1.2 Protein Structure	2
1.3 Protein Geometry	4
1.4 Structure Determination and Prediction	6
1.4.1 Markov Chain Monte Carlo Simulations of Proteins	8
1.5 Generative Models for the Polypeptide Backbone	10
1.5.1 Bivariate Angular Distributions	11
1.5.1.1 Bivariate von Mises	11
1.5.1.2 Histograms and Fourier Series	11
1.5.1.3 Mixture of von Mises	12
1.5.2 A Dynamical Bayesian Network Model: TorusDBN	12
1.6 Generative Models for Amino Acids Side Chains	15
1.7 Discussion	17
2 Orientations of Symmetrical Objects	25
<i>Richard Arnold and Peter Jupp</i>	
2.1 Ambiguous Rotations	25
2.1.1 Symmetry Groups	31
2.1.2 Symmetric Frames	32
2.2 From Symmetric Frames to Symmetric Arrays	33
2.3 Summary Statistics	35
2.4 Testing Uniformity	36
2.5 Distributions of Ambiguous Rotations	38
2.5.1 A General Class of Distributions on $SO(3)/K$	38
2.5.2 Concentrated Distributions	39
2.6 Tests of Location	40
2.6.1 One-Sample Tests	40
2.6.2 Two-Sample Tests	41
2.7 Further Developments	41

2.8	Analysis of Examples	42
2.8.1	Analysis of Example 1	42
2.8.2	Analysis of Example 2	42
2.8.3	Analysis of Example 3	42
3	Correlated Cylindrical Data	45
	<i>Francesco Lagona</i>	
3.1	Correlated Cylindrical Data	47
3.2	Cylindrical Hidden Markov Models	49
3.2.1	The Abe–Ley Density	49
3.2.2	Modeling a Cylindrical Time Series	49
3.2.3	Modelling a Cylindrical Spatial Series	51
3.3	Identification of Sea Regimes	52
3.4	Segmentation of Current Fields	54
3.5	Outline	55
4	Toroidal Diffusions and Protein Structure Evolution	61
	<i>Eduardo García-Portugués, Michael Golden, Michael Sørensen, Kanti V. Mardia, Thomas Hamelryck, and Jotun Hein</i>	
4.1	Introduction	62
4.1.1	Protein Structure	62
4.1.2	Protein Evolution	64
4.1.3	Toward a Generative Model of Protein Evolution	66
4.2	Toroidal Diffusions	67
4.2.1	Toroidal Ornstein–Uhlenbeck Analogues	70
4.2.2	Estimation for Toroidal Diffusions	72
4.2.3	Empirical Performance	75
4.3	ETDBN: An Evolutionary Model for Protein Pairs	77
4.3.1	Hidden Markov Model Structure	77
4.3.2	Site-Classes: Constant Evolution and Jump Events	80
4.3.3	Model Training	81
4.3.4	Benchmarks	83
4.4	Case Study: Detection of a Novel Evolutionary Motif	87
4.5	Conclusions	90
5	Noisy Directional Data	95
	<i>Thanh Mai Pham Ngoc</i>	
5.1	Introduction	95
5.2	Some Preliminaries about Harmonic Analysis on $SO(3)$ and S^2	96
5.3	Model and Assumptions	98
5.3.1	Null and Alternative Hypotheses	98
5.3.2	Noise Assumptions	99
5.4	Test Constructions	100
5.5	Numerical Illustrations	102
5.5.1	The Testing Procedures	102

5.5.2	Alternatives	103
5.5.3	Simulations	104
5.5.4	Real Data: Paleomagnetism	107
5.5.5	Real Data: UHECR	108

6 On Modeling of $SE(3)$ Objects111

Louis-Paul Rivest and Karim Oualkacha

6.1	Introduction	111
6.2	The One Axis Model in $SE(3)$	113
6.2.1	Rotation Matrices and Cardan Angles in $SO(3)$	113
6.2.2	A Geometric Construction of the One Axis Model	113
6.3	Modeling Data from $SE(3)$	115
6.4	Estimation of the Parameters	116
6.4.1	The Rotation Only Estimator of the Rotation Axis A_3 and B_3	116
6.4.2	The Translation Only Estimator of the Parameters	117
6.4.3	The Rotation-Translation Estimator of the Parameters	118
6.5	Numerical Examples	119
6.5.1	Simulations	119
6.5.2	Data Analysis	121
6.6	Discussion	123

7 Spatial and Spatio-Temporal Circular Processes with Application to Wave Directions129

Giovanna Jona-Lasinio, Alan E. Gelfand, and Gianluca Mastrantonio

7.1	Introduction	130
7.2	The Wrapped Spatial and Spatio-Temporal Process	131
7.2.1	Wrapped Spatial Gaussian Process	132
7.2.2	Wrapped Spatio-Temporal Process	133
7.2.3	Kriging and Forecasting	133
7.2.4	Wave Data for the Examples	134
7.2.5	Wrapped Skewed Gaussian Process	136
7.2.5.1	Space-Time Analysis Using the Wrapped Skewed Gaussian Process	140
7.3	The Projected Gaussian Process	141
7.3.1	Univariate Projected Normal Distribution	141
7.3.2	Projected Gaussian Spatial Processes	142
7.3.3	Model Fitting and Inference	145
7.3.4	Kriging with the Projected Gaussian Processes	146
7.3.4.1	An Example Using the Projected Gaussian Process	147
7.3.5	The Space-Time Projected Gaussian Process	149
7.3.6	A Separable Space-Time Wave Direction Data Example	149
7.4	Space-Time Comparison of the WN and PN Models	151
7.5	Joint Modeling of Wave Height and Direction	153

7.6	Concluding Remarks	157
8	Cylindrical Distributions and Their Applications to Biological Data	163
	<i>Toshihiro Abe and Ichiro Ken Shimatani</i>	
8.1	Introduction	164
8.2	Example: Commonly Observed Patterns for Cylindrical Data	165
8.3	A Brief Review of the Univariate Probability Distribution . .	165
8.3.1	Probability Distributions on $[0, \infty)$	165
8.3.2	Circular Distributions	167
8.3.3	Sine-Skewed Perturbation to the Symmetric Circular Distributions	168
8.4	Cylindrical Distributions	168
8.4.1	The Johnson–Wehrly Distribution	168
8.4.2	The Weibull–von Mises Distribution	169
8.4.3	Gamma–von Mises Distribution	170
8.4.4	Generalized Gamma–von Mises Distribution	171
8.4.5	Sine-Skewed Weibull–von-Mises Distribution	172
8.4.6	Parameter Estimation	173
8.5	Application 1: Quantification of the Speed/Turning Angle Patterns of a Flying Bird	173
8.6	Application of Cylindrical Distributions 2: How Trees Are Expanding Crowns	174
8.6.1	Crown Asymmetry in Boreal Forests	176
8.6.2	Crown Asymmetry Model	177
8.6.3	Results of the Cylindrical Models	180
8.7	Concluding Remarks	184
9	Directional Statistics for Wildfires	187
	<i>Jose Ameijeiras-Alonso, Rosa M. Crujeiras, and Alberto Rodríguez Casal</i>	
9.1	Introduction to Wildfire modeling	187
9.2	Fires’ Seasonality	188
9.2.1	Landscape Scale	189
9.2.2	Global Scale	195
9.3	Fires’ Orientation	197
9.3.1	Main Spread on the Orientation of Fires	199
9.3.2	Orientation–Size Joint Distribution	200
9.3.3	Orientation–Size Regression Modeling	205
9.4	Open Problems	206
10	Bayesian Analysis of Circular Data in Social and Behavioral Sciences	211
	<i>Irene Klugkist, Jolien Cremers, and Kees Mulder</i>	
10.1	Introduction	212

10.2	Introducing Two Approaches Conceptually	213
10.2.1	Intrinsic	214
10.2.2	Embedding	215
10.3	Bayesian Modeling	216
10.3.1	Intrinsic	217
10.3.2	Embedding	218
10.4	The Development of Spatial Cognition	218
10.4.1	The Data	218
10.4.2	Bayesian Inference	219
10.4.3	Inequality Constrained Hypotheses	225
10.5	Basic Human Values in the European Social Survey	227
10.5.1	The Data	228
10.5.2	The Model	229
10.5.3	Variable Selection	230
10.5.4	Bayesian Inference	231
10.5.5	Comparison of Approaches	234
10.6	Discussion	236
11	Nonparametric Classification for Circular Data	241
	<i>Marco Di Marzio, Stefania Fensore, Agnese Panzera,</i> <i>and Charles C. Taylor</i>	
11.1	Density Estimation on the Circle	243
11.2	Classification via Density Estimation	245
11.3	Local Logistic Regression	246
11.3.1	Binary Regression via Density Estimation	246
11.3.2	Local Polynomial Binary Regression	248
11.4	Numerical Examples	250
11.5	Classification of Earth's Surface	251
11.6	Conclusion	253
12	Directional Statistics in Machine Learning: A Brief Review	259
	<i>Suvrit Sra</i>	
12.1	Introduction	259
12.2	Basic Directional Distributions	260
12.2.1	Uniform Distribution	260
12.2.2	The von Mises-Fisher Distribution	261
12.2.3	Watson Distribution	261
12.2.4	Other Distributions	262
12.3	Related Work and Applications	263
12.4	Modeling Directional Data: Maximum-Likelihood Estimation	263
12.4.1	Maximum-Likelihood Estimation for vMF	264
12.4.2	Maximum-Likelihood Estimation for Watson	265
12.5	Mixture Models	266
12.5.1	EM Algorithm	267
12.5.2	Limiting Versions	268

12.5.3	Application: Clustering Using movMF	269
12.5.4	Application: Clustering Using moW	271
12.6	Conclusion	272
13	Applied Directional Statistics with R: An Overview	277
	<i>Arthur Pewsey</i>	
13.1	Introduction	277
13.2	The Circular Package	278
13.3	Packages that Use the Circular Package	280
13.4	Other Packages for Circular Statistics	281
13.5	The Directional Package	281
13.6	Other Packages for Directional Statistics	283
13.7	Unsupported Directional Statistics Methodologies	284
13.8	Conclusions	285
	Index	291