

Hurricane Climatology

Hurricane Climatology

A Modern Statistical Guide Using R

James B. Elsner and Thomas H. Jagger

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

© Oxford University Press 2013

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data
Elsner, James B., author.
Hurricane climatology : a modern statistical guide using R / James B. Elsner and Thomas
H. Jagger.
pages ; cm
Includes bibliographical references and index.
ISBN 978-0-19-982763-3 (hardcover : alk. paper)
1. Hurricanes--Forecasting--Statistical methods.
2. R (Computer program language) I. Jagger, Thomas H., author. II. Title.
QC944.E47 2013
551.55'20285555--dc23
2012019317

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

CONTENTS

Preface	xiii
Part One: Data, Statistics, and Software	
1. Hurricanes, Climate, and Statistics	3
1.1. Hurricanes	3
1.2. Climate	5
1.3. Statistics	6
1.4. R	8
1.5. Organization	8
2. R Tutorial	10
2.1. Introduction	10
2.1.1. What Is R?	10
2.1.2. Get R	11
2.1.3. Packages	12
2.1.4. Calculator	12
2.1.5. Functions	13
2.1.6. Warnings and Errors	14
2.1.7. Assignments	14
2.1.8. Help	15
2.2. Data	15
2.2.1. Small Amounts	15
2.2.2. Functions	16
2.2.3. Vectors	17
2.2.4. Structured Data	20
2.2.5. Logic	21
2.2.6. Imports	23
2.3. Tables and Plots	25
2.3.1. Tables and Summaries	25
2.3.2. Quantiles	27
2.3.3. Plots	28
Bar Plots	28
Scatter Plots	29

3. Classical Statistics	33
3.1. Descriptive Statistics	33
3.1.1. Mean, Median, and Maximum	34
3.1.2. Quantiles	36
3.1.3. Missing Values	36
3.2. Probability and Distributions	37
3.2.1. Random Samples	38
3.2.2. Combinatorics	39
3.2.3. Discrete distributions	39
3.2.4. Continuous Distributions	41
3.2.5. Distributions	42
3.2.6. Density	42
3.2.7. Cumulative Distribution Functions	44
3.2.8. Quantile Functions	45
3.2.9. Random Numbers	46
3.3. One-Sample Test	47
3.4. Wilcoxon Signed-Rank Test	52
3.5. Two-Sample Test	54
3.6. Statistical Formula	56
3.7. Two-Sample Wilcoxon Test	57
3.8. Compare Variances	57
3.9. Correlation	58
3.9.1. Pearson's Product-Moment Correlation	59
3.9.2. Spearman's Rank and Kendall's τ Correlation	61
3.9.3. Bootstrap Confidence Intervals	62
3.9.4. Causation	63
3.10. Linear Regression	64
3.11. Multiple Linear Regression	70
3.11.1. Predictor Choice	73
3.11.2. Cross-validation	75
4. Bayesian Statistics	76
4.1. Learning about the Proportion of Landfalls	76
4.2. Inference	80
4.3. Credible Interval	81
4.4. Predictive Density	82
4.5. Is Bayes's Rule Needed?	85
4.6. Bayesian Computation	85
4.6.1. Time-to-Acceptance	86
4.6.2. Markov Chain Monte Carlo approach	90
4.6.3. JAGS	91
4.6.4. WinBUGS	94

5. Graphs and Maps	98
5.1. Graphs	98
5.1.1. Box Plot	99
5.1.2. Histogram	100
5.1.3. Density Plot	102
5.1.4. Q–Q Plot	104
5.1.5. Scatter Plot	106
5.1.6. Conditional Scatter Plot	107
5.2. Time Series	108
5.2.1. Time Series Graph	109
5.2.2. Autocorrelation	110
5.2.3. Dates and Times	111
5.3. Maps	114
5.3.1. Boundaries	114
5.3.2. Data Types	116
Point Data	116
Areal Data	120
Field Data	122
5.4. Coordinate Reference Systems	123
5.5. Export	126
5.6. Other Graphic Packages	128
5.6.1. lattice	128
5.6.2. ggplot2	129
5.6.3. ggmap	132
6. Data Sets	133
6.1. Best-Tracks Data	133
6.1.1. Description	133
6.1.2. Import	135
6.1.3. Intensification	137
6.1.4. Interpolation	138
6.1.5. Regional Activity	140
6.1.6. Lifetime Maximum Intensity	141
6.1.7. Regional Maximum Intensity	142
6.1.8. Tracks by Location	143
6.1.9. Attributes by Location	145
6.2. Annual Aggregation	146
6.2.1. Annual Cyclone Counts	146
6.2.2. Environmental variables	147
6.3. Coastal County Winds	152
6.3.1. Description	152
6.3.2. Counts and Magnitudes	154
6.4. NetCDF Files	156

Part Two: Models and Methods

7. Frequency Models	161
7.1. Counts	161
7.1.1. Poisson Process	162
7.1.2. Inhomogeneous Poisson Process	164
7.2. Environmental Variables	165
7.3. Bivariate Relationships	167
7.4. Poisson Regression	167
7.4.1. Limitation of Linear Regression	168
7.4.2. Poisson Regression Equation	169
7.4.3. Method of Maximum Likelihood	169
7.4.4. Model Fit	170
7.4.5. Interpretation	171
7.5. Model Predictions	172
7.6. Forecast Skill	174
7.6.1. Metrics	174
7.6.2. Cross-Validation	175
7.7. Nonlinear Regression Structure	177
7.8. Zero-Inflated Count Model	179
7.9. Machine Learning	181
7.10. Logistic Regression	183
7.10.1. Exploratory Analysis	185
7.10.2. Logit and Logistic Functions	186
7.10.3. Fit and Interpretation	187
7.10.4. Prediction	188
7.10.5. Fit and Adequacy	189
7.10.6. Receiver Operating Characteristics	191
8. Intensity Models	194
8.1. Lifetime Highest Intensity	194
8.1.1. Exploratory Analysis	194
8.1.2. Quantile Regression	200
8.2. Fastest Hurricane Winds	204
8.2.1. Exploratory Analysis	205
8.2.2. Return Periods	205
8.2.3. Extreme Value Theory	207
8.2.4. Generalized Pareto Distribution	208
8.2.5. Extreme Intensity Model	210
8.2.6. Intensity and Frequency Model	211
8.2.7. Confidence Intervals	212
8.2.8. Threshold Intensity	213
8.3. Categorical Wind Speeds by County	214
8.3.1. Marked Poisson Process	215
8.3.2. Return Levels	216

8.3.3. Covariates	217
8.3.4. Miami-Dade	219
9. Spatial Models	221
9.1. Track Hexagons	221
9.1.1. Spatial Points Data Frame	221
9.1.2. Hexagon Tessellation	223
9.1.3. Overlays	224
9.1.4. Maps	226
9.2. SST Data	228
9.3. SST and Intensity	230
9.4. Spatial Autocorrelation	231
9.4.1. Moran's I	231
9.4.2. Spatial Lag Variable	232
9.4.3. Statistical Significance	233
9.5. Spatial Regression Models	234
9.5.1. Linear Regression	236
9.5.2. Geographically Weighted Regression	237
9.5.3. Model Fit	241
9.6. Spatial Interpolation	243
9.6.1. Preliminaries	243
9.6.2. Sample Variogram	245
9.6.3. Variogram Model	247
9.6.4. Kriging	248
9.6.5. Uncertainty	250
10. Time Series Models	253
10.1. Time Series Overlays	254
10.2. Discrete Time Series	254
10.2.1. Count Variability	255
10.2.2. Moving Average	257
10.2.3. Seasonality	258
10.3. Change Points	260
10.3.1. Counts	261
10.3.2. Covariates	263
10.4. Continuous Time Series	264
10.5. Time-Series Network	268
10.5.1. Time Series Visibility	268
10.5.2. Network Plot	270
10.5.3. Degree Distribution and Anomalous Years	271
10.5.4. Global Metrics	272
11. Cluster Models	275
11.1. Time Clusters	275

11.1.1. Cluster Detection	276
11.1.2. Conditional Counts	277
11.1.3. Cluster Model	278
11.1.4. Parameter Estimation	279
11.1.5. Model Diagnostics	280
11.1.6. Forecasts	283
11.2. Spatial Clusters	284
11.2.1. Point Processes	286
11.2.2. Spatial Density	288
11.2.3. Second-Order Properties	290
11.2.4. Models	292
11.3. Feature Clusters	292
11.3.1. Dissimilarity and Distance	292
11.3.2. K-Means Clustering	295
11.3.3. Track Clusters	297
11.3.4. Track Plots	298
12. Bayesian Models	301
12.1. Long-Range Outlook	301
12.1.1. Poisson-Gamma Conjugate	301
12.1.2. Prior Parameters	302
12.1.3. Posterior Density	304
12.1.4. Predictive Distribution	305
12.2. Seasonal Model	306
12.3. Consensus Model	310
12.3.1. Bayesian Model Averaging	310
12.3.2. Data Plots	313
12.3.3. Model Selection	313
12.3.4. Consensus Hindcasts	319
12.4. Space-Time Model	320
12.4.1. Lattice Data	320
12.4.2. Local Independent Regressions	325
12.4.3. Spatial Autocorrelation	329
12.4.4. BUGS Data	330
12.4.5. MCMC Output	331
12.4.6. Convergence and Mixing	333
12.4.7. Updates	336
12.4.8. Relative Risk Maps	338
13. Impact Models	340
13.1. Extreme Losses	340
13.1.1. Exploratory Analysis	341
13.1.2. Conditional Losses	342
13.1.3. Industry Loss Models	343

13.2. Future Wind Damage	344
13.2.1. Historical Catalogue	344
13.2.2. Gulf of Mexico Hurricanes and SST	347
13.2.3. Intensity Changes with SST	348
13.2.4. Stronger Hurricanes	350
Appendix A. R Functions	352
Appendix B. R Packages	361
Appendix C. Data sets	362
Bibliography	363
Index	370

PREFACE

“The goal is to provide analytical tools that will last students a lifetime.”
—Edward Tufte


A hurricane is nature’s most destructive storm. Violent wind, flooding rain, and powerful surge pose hazards to life and property. Changes in hurricane activity could have significant societal consequences. Hurricanes are the leading source of insured losses from natural catastrophes worldwide and global warming could make the strongest hurricanes even stronger.

Most of what we know about hurricanes comes from past storms. Hurricane climatology is the study of hurricanes as a collection of past events. It answers questions about when, where, and how often. This book is an argument that there is much more remaining to learn. The goal is to show you how to analyze, model, and predict hurricane climate using data. It shows you how to create statistical models from hurricane data that are accessible and explanatory.

The book is didactic. It teaches you how to learn about hurricane climate from data. It uses statistics. Statistics is the science of organizing and interpreting data. Statistics is best done using R. The text is written around code that when copied to an R session reproduces the graphs, tables, and maps presented. The approach is different from other books that use R. Rather than showing you how to use R for a variety of purposes, this book focuses on a single topic and shows you how to better understand it by using R.

The first five chapters provide background material on R and doing statistics. This material is appropriate for an undergraduate course on statistical methods in the environmental sciences. Chapter 6 presents details on the data sets that are used in the later chapters. Chapters 7, 8, and 9 lay out the building blocks of models for hurricane climate research. This material is appropriate for graduate-level courses in climatology. Chapters 10–13 give examples from our more recent research that could be used in a seminar on methods and models for hurricane climate analysis and prediction.

The book benefited from the research conducted with our students including Robert Hodges, Jill Trepanier, and Kelsey Scheitlin. Editorial assistance came from Laura Michaels with additional help from Sarah Strazzo. We thank Ian Elsner for the book’s Web site design. Our sincere thanks go to the R Core Development Team

for maintaining  software and to the many package authors who enrich the software environment. We thank Richard Murnane and Tony Knap of the Risk Prediction Initiative for their unwavering support over the years.

Gratitude goes to Svetla for her wry humor.

James B. Elsner and Thomas H. Jagger
Tallahassee, Florida
August 2012

Hurricane Climatology

