

# Introduction to Probability and Statistics for Ecosystem Managers

Simulation and Resampling

A photograph of three coyotes standing in a field of tall grass. The coyotes are facing left, with the one in the foreground being the most prominent. They have reddish-brown fur with white underparts. The background is a soft-focus green field.

TIMOTHY C. HAAS

STATISTICS IN PRACTICE

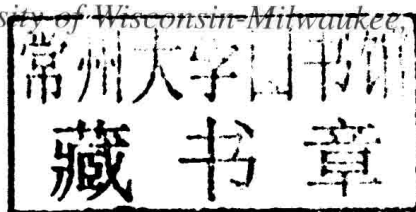
WILEY

# Introduction to Probability and Statistics for Ecosystem Managers

Simulation and Resampling

**Timothy C. Haas**

*Sheldon B. Lubar School of Business  
University of Wisconsin-Milwaukee, USA*



**WILEY**

This edition first published 2013  
© 2013 John Wiley & Sons, Ltd

*Registered office*

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at [www.wiley.com](http://www.wiley.com).

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

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, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

*Library of Congress Cataloging-in-Publication Data*

Haas, Timothy C.

Introduction to probability and statistics for ecosystem managers : simulation and resampling /

Timothy C. Haas, Sheldon B. Lubar.

pages cm

Includes bibliographical references and index.

ISBN 978-1-118-35768-2 (cloth)

I. Ecosystem management—Statistical methods. I. Lubar, Sheldon B. II. Title.

QH77.3.S73H33 2013

333.72 — dc23

2013002861

A catalogue record for this book is available from the British Library.

ISBN: 978-1-118-35768-2

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India

Printed and bound in Malaysia by Vivar Printing Sdn Bhd

# Introduction to Probability and Statistics for Ecosystem Managers

# STATISTICS IN PRACTICE

## *Series Advisors*

### **Human and Biological Sciences**

Stephen Senn

CRP-Santé, Luxembourg

### **Earth and Environmental Sciences**

Marian Scott

*University of Glasgow, UK*

### **Industry, Commerce and Finance**

Wolfgang Jank

*University of Maryland, USA*

### **Founding Editor**

Vic Barnett

*Nottingham Trent University, UK*

---

*Statistics in Practice* is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceuticals; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

*Dedicated to the next generation of ecosystem managers*

# Preface

This textbook will be useful for readers who are either in training for or are in positions having to do with the management of environmental systems and/or wildlife populations wherein one of the decreed management goals is the protection of some part of the ecosystem, for example, wildlife that is at threat from anthropogenic forces. Examples of such positions include being a member of a forestry, fish and game, national parks, or environmental protection agency – or a wildlife advocacy organization such as the African Wildlife Foundation or the World Wildlife Fund. The prerequisites needed for grasping the ideas presented in this textbook are some familiarity with natural resources and a precalculus course.

This textbook has the following pedagogical features:

1. Explanations of probabilistic and statistical concepts are intuitive. Probability is explained through simulations rather than mathematical derivations and statistics is presented through computer-based resampling methods rather than methods based on large sample approximations.
2. Almost all examples show how probability and statistics can be applied to ecosystem management challenges.
3. Exercises are plentiful and appear just after the associated content – making the textbook suitable for a lecture course in a natural resource and/or wildlife management department or as the main text in a program of self-study.
4. Detailed instructions for using the statistical program R are provided along with many complete R programs that generate the output of the textbook's many examples.
5. Enough mathematical programming details are given so that the reader can estimate statistical model parameters with minimum distance methods.
6. An introduction to Geographic Information Systems (GIS) appears that includes examples from quantum GIS (QGIS), a free GIS software package.
7. Spatial and spatio-temporal statistics are introduced and illustrated with examples from ecosystem management that make use of R's spatial statistics capabilities and JAVA<sup>®</sup> programs written by the author. The language of

vectors and matrices is introduced in enough detail to allow the reader to grasp spatial and spatio-temporal models that are expressed in this language.

8. A capstone case study is presented of how one might manage the rhino meta-population kept on private land in South Africa. This case study puts to use the textbook's material on probability, statistics, ecosystem stakeholder models, and individual-based models of wildlife populations.
9. An accompanying website ([www4.uwm.edu/people/haas/introttext](http://www4.uwm.edu/people/haas/introttext)) contains all R and JAVA codes used in this textbook. It also contains all datasets used in the textbook's examples, a web-based ecosystem management tool (EMT) developed by the author in his previous book, *Improving Natural Resource Management: Ecological and Political Models* (Wiley-Blackwell), and answers to all of the textbook's exercises.

Several items are original to this textbook:

1. A new function to transform non-normal data to near-normality.
2. R codes to compute a spatial median filter, spatial cumulants, and spatial neural networks; along with codes that implement a probabilistic model of the spatial diffusion of an invasive species and an algorithm for constrained random search.
3. A learning algorithm that models how ecosystem stakeholders learn from experience as they reach ecosystem-affecting decisions.
4. Complete coverage of how to build and evaluate an individual-based model of a wildlife population that is to be managed.
5. An Online Intelligent Tutoring System (OITS) tied to the text that uses a learned model of a reader to deliver explanations that are focused on just those topics the reader is having difficulty with. This tutoring system can be found at the above-mentioned website.



# Acknowledgments

The author appreciates comments made on an early version of Chapter 10 by the participants of the 7th International Wildlife Ranching Symposium, Kimberley, South Africa, October 10–13, 2011.

# List of abbreviations

<i>m</i> -NN	<i>m</i> -nearest-neighbor
AER	actual error rate
ANOVA	analysis of variance
BLUP	best linear unbiased predictor
CDF	cumulative distribution function
CPT	conditional probability table
CRS	coordinate reference system
CSR	complete spatial randomness
d.o.f.	degrees of freedom
DAG	directed acyclic graph
DL	description length
DM-group	decision-making group
EMAT	ecosystem management actions taxonomy
EMT	ecosystem management tool
EPA	Environment Protection Agency
ESA/NASA	European Space Agency/National Aeronautics and Space Administration
ESTDM	event-based spatio-temporal data model
FBLGDMD	feedback-based learning for group decision-making diagrams
FFT	fast Fourier transform
GIS	geographic information system
GLS	generalized least squares
GUI	graphical user interface
i.i.d.	independently and identically distributed
IBM	individual-based model
ICBEMP	Interior Columbia Basin Ecosystem Management Project
INTERCALV	intercalving interval
IQR	inter-quartile range
IUCN	International Union for Conservation of Nature
KECs	Key environmental correlates
LE	life expectancy
LOMAP	Local Model And Predictor
MA	maturation age
MCSTK	moving cylinder spatio-temporal kriging

MDLEP	minimum description length-evolutionary programming
MDL	minimum description length
MHD	minimum Hellinger distance
MLE	maximum likelihood estimate
MPEMP	most practical ecosystem management plan
MSHD	minimum simulated Hellinger distance
MSL	maximum simulated likelihood
MWRCK	moving window residual cokriging
MWRRK	moving window, regression, residual kriging
NEMBA	National Environmental Management: Biodiversity Act
NEMPAA	National Environmental Management: Protected Areas Act
NGO	nongovernmental organization
NP-hard	non-polynomial time hard
OC	operating characteristic
OGA	overall goal attainment
OITS	online intelligent tutoring system
OK	ordinary kriging
OLS	ordinary least squares
PAC	protected area complex
PA	protected area
PDF	probability density function
PDPF	probability density-probability function
PMF	probability mass function
POM	pattern-oriented modeling
Q-Q	Quantile-Quantile
QGIS	quantum GIS
SPPP	spatial Poisson point process
SSE	error sum of squares
SSR	regression sum of squares
SST	total sum of squares
TOPS	Threatened and Protected Species
USDA	US Department of Agriculture
WKT	well-known text

# Contents

List of figures	xiii
List of tables	xvii
Preface	xix
Acknowledgments	xxi
List of abbreviations	xxiii
<b>1 Introduction</b>	<b>1</b>
1.1 The textbook's purpose	1
1.1.1 The textbook's focus on ecosystem management	2
1.1.2 Reader level, prerequisites, and typical reader jobs	3
1.2 The textbook's pedagogical approach	4
1.2.1 General points	4
1.2.2 Use of this textbook for self-study	4
1.2.3 Learning resources	5
1.3 Chapter summaries	7
1.4 Installing and running R Commander	9
1.4.1 Running R	9
1.4.2 Starting an R Commander session	9
1.4.3 Terminating an R Commander session	10
1.5 Introductory R Commander session	10
1.6 Teaching probability through simulation	13
1.6.1 The frequentist statistical inference paradigm	14
1.7 Summary	15
<b>2 Probability and simulation</b>	<b>17</b>
2.1 Introduction	17
2.2 Basic probability	17
2.2.1 Definitions	17
2.2.2 Independence	20
2.3 Random variables	22

2.3.1	Definitions	22
2.3.2	Simulating random variables	26
2.3.3	A random variable's expected value (mean) and variance	26
2.3.4	Details of the normal (Gaussian) distribution	28
2.3.5	Distribution approximations	30
2.4	Joint distributions	31
2.4.1	Definition	31
2.4.2	Mixed variables	32
2.4.3	Marginal distribution	32
2.4.4	Conditional distributions	33
2.4.5	Independent random variables	34
2.5	Influence diagrams	34
2.5.1	Definitions	34
2.5.2	Example of a Bayesian network in ecosystem management	36
2.5.3	Modeling causal relationships with an influence diagram	38
2.6	Advantages of influence diagrams in ecosystem management	40
2.7	Two ecosystem management Bayesian networks	41
2.7.1	Waterbody eutrophication	41
2.7.2	Wildlife population viability	41
2.8	Influence diagram sensitivity analysis	41
2.9	Drawbacks to influence diagrams	42
<b>3</b>	<b>Application of probability: Models of political decision making in ecosystem management</b>	<b>43</b>
3.1	Introduction	43
3.2	Influence diagram models of decision making	43
3.2.1	Ecosystem status perception nodes	44
3.2.2	Image nodes	44
3.2.3	Economic, militaristic, and institutional goal nodes	45
3.2.4	Audience effect nodes	45
3.2.5	Resource nodes	46
3.2.6	Action and target nodes	46
3.2.7	Overall goal attainment node	47
3.2.8	How a group influence diagram reaches a decision	47
3.2.9	An advantage of this decision-making architecture	47
3.2.10	Evaluation dimensions	47
3.3	Rhino poachers: A simplified model	50
3.4	Policymakers: A simplified model	57
3.5	Conclusions	59
<b>4</b>	<b>Statistical inference I: Basic ideas and parameter estimation</b>	<b>61</b>
4.1	Definitions of some fundamental terms	61
4.2	Estimating the PDF and CDF	62
4.2.1	Histograms	62
4.2.2	Ogive	64

4.3	Measures of central tendency and dispersion	64
4.4	Sample quantiles	65
4.4.1	Sample quartiles	65
4.4.2	Sample deciles and percentiles	65
4.5	Distribution of a statistic	65
4.5.1	Basic setup in statistics	65
4.5.2	Sampling distributions	66
4.5.3	Normal quantile–quantile plot	66
4.6	The central limit theorem	68
4.7	Parameter estimation	68
4.7.1	Bias, variance, and efficiency	69
4.8	Interval estimates	70
4.8.1	A confidence interval for $\mu$ when $\sigma^2$ is known	70
4.9	Basic regression analysis	71
4.9.1	Definitions and fundamental characteristics	71
4.9.2	The regression model	72
4.9.3	Correlation	74
4.9.4	Sampling distributions	75
4.9.5	Prediction and estimation	76
4.9.6	Misuse of regression models	76
4.10	General methods of parameter estimation	79
4.10.1	Maximum likelihood	79
4.10.2	Minimum Hellinger distance	80
4.10.3	Consistency analysis	80
<b>5</b>	<b>Statistical inference II: Hypothesis tests</b>	<b>83</b>
5.1	Introduction	83
5.2	Hypothesis tests: General definitions and properties	83
5.2.1	Definitions and procedure	83
5.2.2	Confidence intervals and hypothesis tests	85
5.2.3	Types of mistakes	85
5.2.4	One way to set the test's level	86
5.2.5	The $z$ -test for hypotheses about $\mu$	89
5.2.6	$p$ -Values	91
5.3	Power	92
5.3.1	Power curves	93
5.4	$t$ -Tests and a test for equal variances	95
5.4.1	The $t$ -test	95
5.4.2	Two-sample $t$ -tests	95
5.4.3	Tests for paired data	96
5.4.4	Testing for equal variances	98
5.5	Hypothesis tests on the regression model	98
5.5.1	Prediction and estimation confidence intervals	103
5.5.2	Multiple regression	104
5.5.3	Original scale prediction in regression	106

5.6	Brief introduction to vectors and matrices	106
5.6.1	Basic definitions	106
5.6.2	Inverse of a matrix	108
5.6.3	Random vectors and random matrices	108
5.7	Matrix form of multiple regression	109
5.7.1	Generalized least squares	111
5.8	Hypothesis testing with the delete- <i>d</i> jackknife	111
5.8.1	Background	111
5.8.2	A one-sample delete- <i>d</i> jackknife test	111
5.8.3	Testing classifier error rates	114
5.8.4	Important points about this test	115
5.8.5	Parameter confidence intervals	115
<b>6</b>	<b>Introduction to spatial statistics</b>	<b>117</b>
6.1	Overview	117
6.1.1	Types of spatial processes	118
6.2	Spatial statistics and GIS	118
6.2.1	Types of spatial data	118
6.3	QGIS	121
6.3.1	Capabilities	122
6.3.2	Installing QGIS	122
6.3.3	Documentation and tutorials	122
6.3.4	Installing plugins	123
6.3.5	How to convert a text file to a shapefile	123
6.4	Continuous spatial processes	125
6.4.1	Definitions	125
6.4.2	Graphical tools for exploring continuous spatial data	127
6.4.3	Third- and fourth-order cumulant minimization	132
6.4.4	Best linear unbiased predictor	132
6.4.5	Kriging variance	134
6.4.6	Model-fitting diagnostics	136
6.4.7	Kriging within a window	137
6.5	Spatial point processes	138
6.5.1	Definitions	138
6.5.2	Marked spatial point processes	149
6.5.3	Conclusions	150
6.6	Continuously valued multivariate processes	151
6.6.1	Fitting multivariate covariance functions	151
6.6.2	Cokriging: The MWRCK procedure	155
<b>7</b>	<b>Introduction to spatio-temporal statistics</b>	<b>159</b>
7.1	Introduction	159
7.2	Representing time in a GIS	159
7.2.1	The QGIS Time Manager plugin	160
7.2.2	A Clifford algebra-based spatio-temporal data structure	163

7.2.3	A raster- and event-based spatio-temporal data model	163
7.2.4	Application of ESTDM to a land cover study	166
7.3	Spatio-temporal prediction: MCSTK	166
7.3.1	Algorithms	166
7.3.2	Covariogram model and its estimator	169
7.4	Multivariate processes	174
7.4.1	Definitions	175
7.4.2	Transformations	175
7.4.3	Covariograms and cross-covariograms	180
7.4.4	Parameter estimation	181
7.4.5	Prediction algorithms	182
7.4.6	Cross-validation	183
7.4.7	Summary	190
7.5	Spatio-temporal point processes	190
7.6	Marked spatio-temporal point processes	195
7.6.1	A mark semivariogram estimator	196
<b>8</b>	<b>Application of statistical inference: Estimating the parameters of an individual-based model</b>	<b>199</b>
8.1	Overview	199
8.2	A simple IBM and its estimation	200
8.2.1	Simple IBM	200
8.2.2	Parameter estimation	201
8.3	Fitting IBMs with MSHD	204
8.3.1	Ergodicity	206
8.3.2	Observable random variables from IBM output	207
8.4	Further properties of parameter estimators	207
8.4.1	Consistency	207
8.4.2	Robustness	208
8.5	Parameter confidence intervals for a nonergodic model	209
8.6	Rhino-supporting ecosystem influence diagram	209
8.6.1	Spatial effects on poaching	210
8.6.2	IBM variables	213
8.6.3	Initial conditions and hypothesis values of parameters	214
8.6.4	Mapping functions	215
8.6.5	Realism of ecosystem influence diagram output	217
8.7	Estimation of rhino IBM parameters	219
8.7.1	Parameter confidence intervals	220
<b>9</b>	<b>Guiding an influence diagram's learning</b>	<b>223</b>
9.1	Introduction	223
9.2	Online learning of Bayesian network parameters	224
9.2.1	Basic algorithm using simulation	224
9.2.2	Updating influence diagrams	225
9.3	Learning an influence diagram's structure	229



9.3.1	Minimum description length score function	229
9.3.2	Description length of an edge	229
9.3.3	Random generation of DAGs	230
9.3.4	Algorithm to detect and delete cycles	230
9.3.5	Mutate functions	231
9.3.6	MDLEP algorithm	232
9.3.7	Using MDLEP to learn influence diagram structure	232
9.4	Feedback-based learning for group decision-making diagrams	233
9.4.1	Definitions and algorithm	233
9.5	Summary and conclusions	234
<b>10</b>	<b>Fitting and testing a political–ecological simulator</b>	<b>235</b>
10.1	Introduction	235
10.1.1	Background on rhino poaching	236
10.1.2	Scenarios wherein rhino poaching is reduced	237
10.2	EMT simulator construction	237
10.2.1	Modeled groups	237
10.2.2	Rhino-supporting ecosystem influence diagram	248
10.3	Consistency analysis estimates of simulator parameters	248
10.4	MPEMP computation	251
10.4.1	Setup	251
10.4.2	Solution	253
10.5	Conclusions	254
	<b>Appendix</b>	<b>257</b>
	Simpson’s rule in two dimensions	257
	<b>References</b>	<b>263</b>
	<b>Index</b>	<b>275</b>