

Composite Sampling

Environmental and Ecological Statistics

Series Editors

G.P. Patil

Distinguished Professor of Mathematical Statistics
Director, Center for Statistical Ecology and Environmental Statistics
Editor-in-Chief, Environmental and Ecological Statistics
The Pennsylvania State University

Timothy G. Gregoire

J.P. Weyerhaeuser, Jr. Professor of Forest Management
School of Forestry and Environmental Studies
Yale University

Andrew B. Lawson

Division of Biostatistics & Epidemiology
Medical University of South Carolina

Barry D. Nussbaum

Chief Statistician
U.S. Environmental Protection Agency

The Springer series **Environmental and Ecological Statistics** is devoted to the cross-disciplinary subject area of environmental and ecological statistics discussing important topics and themes in statistical ecology, environmental statistics, and relevant risk analysis. Emphasis is focused on applied mathematical statistics, statistical methodology, data interpretation and improvement for future use, with a view to advance statistics for environment, ecology, and environmental health, and to advance environmental theory and practice using valid statistics.

Each volume in the **Environmental and Ecological Statistics** series is based on the appropriateness of the statistical methodology to the particular environmental and ecological problem area, within the context of contemporary environmental issues and the associated statistical tools, concepts, and methods.

Previous Volumes in this Series

VOLUME 1

Landscape Pattern Analysis for Assessing Ecosystem Condition
Glen D. Johnson and Ganapati Patil

VOLUME 2

Pattern-Based Compression of Multi-Band Image Data for Landscape Analysis
Wayne Myers and Ganapati Patil

VOLUME 3

Modeling Demographic Processes In Marked Populations
Edited by David L. Thomson, Evan G. Cooch and Michael J. Conroy

VOLUME 4

Composite Sampling: A Novel Method to Accomplish Observational Economy in Environmental Studies
Ganapati P. Patil, Sharad D. Gore and Charles Taillie (Deceased)



For further volumes:

<http://www.springer.com/series/7506>

Ganapati P. Patil · Sharad D. Gore ·
Charles Taillie (Deceased)

Composite Sampling

A Novel Method to Accomplish
Observational Economy in
Environmental Studies



Springer

Ganapati P. Patil
Center for Statistical Ecology
and Environmental Statistics
The Pennsylvania State University
University Park, PA, USA
gpp@stat.psu.edu

Sharad D. Gore
Department of Statistics
University of Pune
Pune, India
sdgore@stats.unipune.ac.in

Charles Taillie
(Deceased)
Center for Statistical Ecology
and Environmental Statistics
Penn State University
University Park, PA, USA

ISBN 978-1-4419-7627-7 e-ISBN 978-1-4419-7628-4
DOI 10.1007/978-1-4419-7628-4
Springer New York Dordrecht Heidelberg London

© Springer Science+Business Media, LLC 2011

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper.

Springer is part of Springer Science+Business Media (www.springer.com)

To
Lalit, Parimal, Parag, Pawan, Priya
Rekha, Archis, Reshma
Marcella
With love and affection

Acknowledgements

This monograph provides, for the first time, a most comprehensive statistical account of composite sampling as an ingenious environmental sampling method to help accomplish observational economy in a variety of environmental studies. It is primarily based on a 4-year effort on composite sampling in theory and practice under a cooperative research agreement CR-821531 of the Penn State Center for Statistical Ecology and Environmental Statistics with what was known as the Statistical Analysis and Computing Branch of the Division of Environmental Statistics and Information in the Office of Policy, Planning, and Evaluation of the United States Environmental Protection Agency. Several unpublished results and applications appear for the first time in this monograph.

Our appreciation goes to the members of the Statistical Analysis and Computing Branch for their technical interest and administrative support. Our thanks are particularly due to Dr. N. Phillip Ross, the division director, and his distinguished staff statisticians, Herbert Lacayo, Robert O'Brien, John Warren, and Barry Nussbaum, the branch chief, and now the chief statistician of EPA. We should also express our sincere thanks to several of our friends and colleagues for discussions regarding various issues and aspects of composite sampling in environmental studies. They are Marilyn Boswell, George Flatman, Richard Gilbert, Glen Johnson, Henry Kahn, Gianfranco Lovison, Royal Nadeau, Dean Neptune, Charles Rohde, David Schaeffer, Llew Williams, and a few others.

And, finally, our immense appreciation to Barbara Freed and Janet Huff for taking care of it all at the Penn State Center for Statistical Ecology and Environmental Statistics.

Contents

1	Introduction	1
2	Classifying Individual Samples into One of Two Categories	9
2.1	Introduction	9
2.2	Presence/Absence Measurements	11
2.2.1	Exhaustive Retesting	12
2.2.2	Sequential Retesting	15
2.2.3	Binary Split Retesting	18
2.2.4	Curtailed Exhaustive Retesting	23
2.2.5	Curtailed Sequential Retesting	27
2.2.6	Curtailed Binary Split Retesting	31
2.2.7	Entropy-Based Retesting	33
2.2.8	Exhaustive Retesting in the Presence of Classification Errors	38
2.2.9	Other Costs	40
2.3	Continuous Response Variables	41
2.3.1	Quantitatively Curtailed Exhaustive Retesting	45
2.3.2	Binary Split Retesting	46
2.3.3	Entropy-Based Retesting	49
2.4	Cost Analysis of Composite Sampling for Classification	49
2.4.1	Introduction	49
2.4.2	General Cost Expression	49
2.4.3	Effect of False Positives and False Negatives on Composite Sample Classification	50
2.4.4	Presence/Absence Measurements	51
2.4.5	Continuous Measurements	53
3	Identifying Extremely Large Observations	55
3.1	Introduction	55
3.2	Prediction of the Sample Maximum	56
3.3	The Sweep-Out Method to Identify the Sample Maximum	58

3.4	Extensive Search of Extreme Values	59
3.5	Application	60
3.6	Two-Way Composite Sampling Design	68
3.7	Illustrative Example	70
3.8	Analysis of Composite Sampling Data Using the Principle of Maximum Entropy	76
3.8.1	Introduction	76
3.8.2	Modeling Composite Sampling Using the Principle of Maximum Entropy	77
3.8.3	When Is the Maximum Entropy Model Reasonable in Practice?	78
4	Estimating Prevalence of a Trait	81
4.1	Introduction	81
4.2	The Maximum Likelihood Estimator	82
4.3	Alternative Estimators	84
4.4	Comparison Between \hat{p} and \tilde{p}	85
4.5	Estimation of Prevalence in Presence of Measurement Error	85
5	A Bayesian Approach to the Classification Problem	87
5.1	Introduction	87
5.2	Bayesian Updating of p	90
5.3	Minimization of the Expected Relative Cost	93
5.4	Discussion	95
6	Inference on Mean and Variance	97
6.1	Introduction	97
6.2	Notation and Basic Results	98
6.2.1	Notation	98
6.2.2	Basic Results	99
6.3	Estimation Without Measurement Error	101
6.4	Estimation in the Presence of Measurement Error	103
6.5	Maintaining Precision While Reducing Cost	104
6.6	Estimation of σ_x^2 and σ_ϵ^2	105
6.7	Estimation of Population Variance	106
6.8	Confidence Interval for the Population Mean	109
6.9	Tests of Hypotheses in the Population Mean	110
6.9.1	One-Sample Tests	110
6.9.2	Two-Sample Tests	111
6.10	Applications	112
6.10.1	Comparison of Arithmetic Averages of Soil pH Values with the pH Values of Composite Samples	112

- 6.10.2 Comparison of Random and Composite Sampling Methods for the Estimation of Fat Contents of Bulk Milk Supplies 112
- 6.10.3 Optimization of Sampling for the Determination of Mean Radium-226 Concentration in Surface Soil . . 113

- 7 Composite Sampling with Random Weights 115**
 - 7.1 Introduction 115
 - 7.2 Expected Value, Variance, and Covariance of Bilinear Random Forms 116
 - 7.3 Models for the Weights 118
 - 7.3.1 Assumptions on the First Two Moments 119
 - 7.3.2 Distributional Assumptions 119
 - 7.4 The Model for Composite Sample Measurements 121
 - 7.4.1 Subsampling a Composite Sample 121
 - 7.4.2 Several Composite Samples 124
 - 7.4.3 Subsampling of Several Composite Samples 125
 - 7.4.4 Measurement Error 126
 - 7.5 Applications 128
 - 7.5.1 Sampling Frequency and Comparison of Grab and Composite Sampling Programs for Effluents 128
 - 7.5.2 Theoretical Comparison of Grab and Composite Sampling Programs 128
 - 7.5.3 Grab vs. Composite Sampling: A Primer for the Manager and Engineer 129
 - 7.5.4 Composite Samples Overestimate Waste Loads 129
 - 7.5.5 Composite Samples for Foliar Analysis 132
 - 7.5.6 Lateral Variability of Forest Floor Properties Under Second-Growth Douglas-Fir Stands and the Usefulness of Composite Sampling Techniques 133

- 8 A Linear Model for Estimation with Composite Sample Data 135**
 - 8.1 Introduction 135
 - 8.2 Motivation for a Unified Model 136
 - 8.3 The Model 137
 - 8.4 Discussion of the Assumptions 139
 - 8.4.1 The Structural/Sampling Submodel 139
 - 8.4.2 The Compositing/Subsampling Submodel 140
 - 8.4.3 The Structure of the Matrices W , M_W , and Σ_W 140
 - 8.5 Moments of x and y 146
 - 8.6 Complex Sampling Schemes Before Compositing 146
 - 8.6.1 Segmented Populations 147
 - 8.6.2 Estimating the Mean in Segmented Populations 147

- 8.6.3 Estimating Variance Components in Segmented Populations 150
- 8.7 Estimating the Effect of a Binary Factor 153
 - 8.7.1 Fully Segregated Composites 157
 - 8.7.2 Fully Confounded Composites 161
- 8.8 Elementary Matrices and Kronecker Products 164
 - 8.8.1 Decomposition of Block Matrices 165
- 8.9 Expectation and Dispersion Matrix When Both W and x Are Random 168
 - 8.9.1 The Expectation of Wx 168
 - 8.9.2 Variance/Covariance Matrix of Wx 172
- 9 Composite Sampling for Site Characterization and Cleanup** 175
 - Evaluation** 175
 - 9.1 Data Quality Objectives 175
 - 9.2 Optimal Composite Designs 178
 - 9.2.1 Cost of a Sampling Program 179
 - 9.2.2 Optimal Allocation of Resources 179
 - 9.2.3 Power of a Test and Determination of Sample Size 180
 - 9.2.4 Algorithms for Determination of Sample Size 181
- 10 Spatial Structures of Site Characteristics and Composite Sampling** . . 183
 - 10.1 Introduction 183
 - 10.2 Models for Spatial Processes 183
 - 10.2.1 Composite Sampling 187
 - 10.3 Application to Two Superfund Sites 190
 - 10.3.1 The Two Sites 190
 - 10.3.2 Methods 191
 - 10.3.3 Results 192
 - 10.3.4 Discussion 195
 - 10.4 Compositing by Spatial Contiguity 198
 - 10.4.1 Introduction 198
 - 10.4.2 Retesting Strategies 199
 - 10.4.3 Composite Sample-Forming Schemes 200
 - 10.5 Compositing of Ranked Set Samples 201
 - 10.5.1 Ranked Set Sampling 201
 - 10.5.2 Relative Precision of the RSS Estimator of a Population Mean Relative to Its SRS Estimator ... 204
 - 10.5.3 Unequal Allocation of Sample Sizes 205
 - 10.5.4 Formation of Homogeneous Composite Samples 206
- 11 Composite Sampling of Soils and Sediments** 209
 - 11.1 Detection of Contamination 209
 - 11.1.1 Detecting PCB Spills 209
 - 11.1.2 Compositing Strategy for Analysis of Samples 211

- 11.2 Estimation of the Average Level of Contamination 213
 - 11.2.1 Estimation of the Average PCB Concentration on the Spill Area 213
 - 11.2.2 Onsite Surface Soil Sampling for PCB at the Armagh Site 214
 - 11.2.3 The Armagh Site 215
 - 11.2.4 Simulating Composite Samples 218
 - 11.2.5 Locating Individual Samples with High PCB Concentrations 221
- 11.3 Estimation of Trace Metal Storage in Lake St. Clair Post-settlement Sediments Using Composite Samples 222

- 12 Composite Sampling of Liquids and Fluids 227**
 - 12.1 Comparison of Random and Composite Sampling Methods for the Estimation of Fat Content of Bulk Milk Supplies 227
 - 12.1.1 Experiment 227
 - 12.1.2 Estimation Methods 228
 - 12.1.3 Results 228
 - 12.1.4 Composite Compared with Yield-Weighted Estimate of Fat Percentage 229
 - 12.2 Composite Sampling of Highway Runoff 229
 - 12.3 Composite Samples Overestimate Waste Loads 232

- 13 Composite Sampling and Indoor Air Pollution 235**
 - 13.1 Household Dust Samples 235

- 14 Composite Sampling and Bioaccumulation 239**
 - 14.1 Example: National Human Adipose Tissue Survey 241
 - 14.2 Results from the Analysis of 1987 NHATS Data 241

- Glossary and Terminology 243**

- Bibliography 249**

- Index 267**