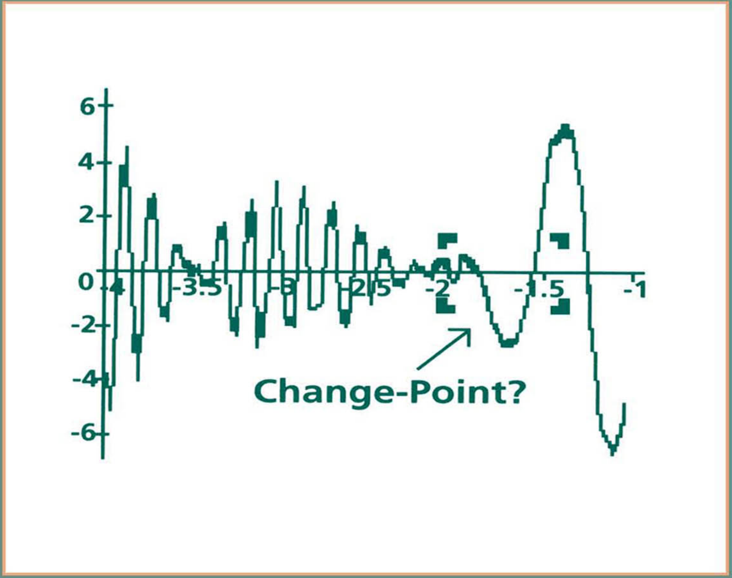
STATISTICS:
a DEKKER series of TEXTBOOKS and MONOGRAPHS

Applied Sequential Methodologies

Real-World Examples with Data Analysis



edited by

Nitis Mukhopadhyay
Sujay Datta
Saibal Chattopadhyay

Applied Sequential Methodologies

1. The Generalized Jackknife Statistic, *H. L. Gray and W. R. Schucany*
2. Multivariate Analysis, *Anant M. Kshirsagar*
3. Statistics and Society, *Walter T. Federer*
4. Multivariate Analysis: A Selected and Abstracted Bibliography, 1957–1972, *Kocherlakota Subrahmaniam and Kathleen Subrahmaniam*
5. Design of Experiments: A Realistic Approach, *Virgil L. Anderson and Robert A. McLean*
6. Statistical and Mathematical Aspects of Pollution Problems, *John W. Pratt*
7. Introduction to Probability and Statistics (in two parts), Part I: Probability; Part II: Statistics, *Narayan C. Giri*
8. Statistical Theory of the Analysis of Experimental Designs, *J. Ogawa*
9. Statistical Techniques in Simulation (in two parts), *Jack P. C. Kleijnen*
10. Data Quality Control and Editing, *Joseph I. Naus*
11. Cost of Living Index Numbers: Practice, Precision, and Theory, *Kali S. Banerjee*
12. Weighing Designs: For Chemistry, Medicine, Economics, Operations Research, Statistics, *Kali S. Banerjee*
13. The Search for Oil: Some Statistical Methods and Techniques, *edited by D. B. Owen*
14. Sample Size Choice: Charts for Experiments with Linear Models, *Robert E. Odeh and Martin Fox*
15. Statistical Methods for Engineers and Scientists, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
16. Statistical Quality Control Methods, *Irving W. Burr*
17. On the History of Statistics and Probability, *edited by D. B. Owen*
18. Econometrics, *Peter Schmidt*
19. Sufficient Statistics: Selected Contributions, *Vasant S. Huzurbazar (edited by Anant M. Kshirsagar)*
20. Handbook of Statistical Distributions, *Jagdish K. Patel, C. H. Kapadia, and D. B. Owen*
21. Case Studies in Sample Design, *A. C. Rosander*
22. Pocket Book of Statistical Tables, *compiled by R. E. Odeh, D. B. Owen, Z. W. Birnbaum, and L. Fisher*
23. The Information in Contingency Tables, *D. V. Gokhale and Solomon Kullback*
24. Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods, *Lee J. Bain*
25. Elementary Statistical Quality Control, *Irving W. Burr*
26. An Introduction to Probability and Statistics Using BASIC, *Richard A. Groeneveld*
27. Basic Applied Statistics, *B. L. Raktoe and J. J. Hubert*
28. A Primer in Probability, *Kathleen Subrahmaniam*
29. Random Processes: A First Look, *R. Syski*
30. Regression Methods: A Tool for Data Analysis, *Rudolf J. Freund and Paul D. Minton*
31. Randomization Tests, *Eugene S. Edgington*
32. Tables for Normal Tolerance Limits, Sampling Plans and Screening, *Robert E. Odeh and D. B. Owen*
33. Statistical Computing, *William J. Kennedy, Jr., and James E. Gentle*
34. Regression Analysis and Its Application: A Data-Oriented Approach, *Richard F. Gunst and Robert L. Mason*
35. Scientific Strategies to Save Your Life, *I. D. J. Bross*
36. Statistics in the Pharmaceutical Industry, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
37. Sampling from a Finite Population, *J. Hajek*
38. Statistical Modeling Techniques, *S. S. Shapiro and A. J. Gross*
39. Statistical Theory and Inference in Research, *T. A. Bancroft and C.-P. Han*
40. Handbook of the Normal Distribution, *Jagdish K. Patel and Campbell B. Read*
41. Recent Advances in Regression Methods, *Hrishikesh D. Vinod and Aman Ullah*
42. Acceptance Sampling in Quality Control, *Edward G. Schilling*
43. The Randomized Clinical Trial and Therapeutic Decisions, *edited by Niels Tygstrup, John M Lachin, and Erik Juhl*

44. Regression Analysis of Survival Data in Cancer Chemotherapy, *Walter H. Carter, Jr., Galen L. Wampler, and Donald M. Stablein*
45. A Course in Linear Models, *Anant M. Kshirsagar*
46. Clinical Trials: Issues and Approaches, *edited by Stanley H. Shapiro and Thomas H. Louis*
47. Statistical Analysis of DNA Sequence Data, *edited by B. S. Weir*
48. Nonlinear Regression Modeling: A Unified Practical Approach, *David A. Ratkowsky*
49. Attribute Sampling Plans, Tables of Tests and Confidence Limits for Proportions, *Robert E. Odeh and D. B. Owen*
50. Experimental Design, Statistical Models, and Genetic Statistics, *edited by Klaus Hinkelmann*
51. Statistical Methods for Cancer Studies, *edited by Richard G. Cornell*
52. Practical Statistical Sampling for Auditors, *Arthur J. Wilburn*
53. Statistical Methods for Cancer Studies, *edited by Edward J. Wegman and James G. Smith*
54. Self-Organizing Methods in Modeling: GMDH Type Algorithms, *edited by Stanley J. Farlow*
55. Applied Factorial and Fractional Designs, *Robert A. McLean and Virgil L. Anderson*
56. Design of Experiments: Ranking and Selection, *edited by Thomas J. Santner and Ajit C. Tamhane*
57. Statistical Methods for Engineers and Scientists: Second Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
58. Ensemble Modeling: Inference from Small-Scale Properties to Large-Scale Systems, *Alan E. Gelfand and Crayton C. Walker*
59. Computer Modeling for Business and Industry, *Bruce L. Bowerman and Richard T. O'Connell*
60. Bayesian Analysis of Linear Models, *Lyle D. Broemeling*
61. Methodological Issues for Health Care Surveys, *Brenda Cox and Steven Cohen*
62. Applied Regression Analysis and Experimental Design, *Richard J. Brook and Gregory C. Arnold*
63. Statpal: A Statistical Package for Microcomputers—PC-DOS Version for the IBM PC and Compatibles, *Bruce J. Chalmer and David G. Whitmore*
64. Statpal: A Statistical Package for Microcomputers—Apple Version for the II, II+, and IIe, *David G. Whitmore and Bruce J. Chalmer*
65. Nonparametric Statistical Inference: Second Edition, Revised and Expanded, *Jean Dickinson Gibbons*
66. Design and Analysis of Experiments, *Roger G. Petersen*
67. Statistical Methods for Pharmaceutical Research Planning, *Sten W. Bergman and John C. Gittins*
68. Goodness-of-Fit Techniques, *edited by Ralph B. D'Agostino and Michael A. Stephens*
69. Statistical Methods in Discrimination Litigation, *edited by D. H. Kaye and Mikel Aickin*
70. Truncated and Censored Samples from Normal Populations, *Helmut Schneider*
71. Robust Inference, *M. L. Tiku, W. Y. Tan, and N. Balakrishnan*
72. Statistical Image Processing and Graphics, *edited by Edward J. Wegman and Douglas J. DePriest*
73. Assignment Methods in Combinatorial Data Analysis, *Lawrence J. Hubert*
74. Econometrics and Structural Change, *Lyle D. Broemeling and Hiroki Tsurumi*
75. Multivariate Interpretation of Clinical Laboratory Data, *Adelin Albert and Eugene K. Harris*
76. Statistical Tools for Simulation Practitioners, *Jack P. C. Kleijnen*
77. Randomization Tests: Second Edition, *Eugene S. Edgington*
78. A Folio of Distributions: A Collection of Theoretical Quantile-Quantile Plots, *Edward B. Fowlkes*
79. Applied Categorical Data Analysis, *Daniel H. Freeman, Jr.*
80. Seemingly Unrelated Regression Equations Models: Estimation and Inference, *Virendra K. Srivastava and David E. A. Giles*

81. Response Surfaces: Designs and Analyses, *Andre I. Khuri and John A. Cornell*
82. Nonlinear Parameter Estimation: An Integrated System in BASIC, *John C. Nash and Mary Walker-Smith*
83. Cancer Modeling, *edited by James R. Thompson and Barry W. Brown*
84. Mixture Models: Inference and Applications to Clustering, *Geoffrey J. McLachlan and Kaye E. Basford*
85. Randomized Response: Theory and Techniques, *Arijit Chaudhuri and Rahul Mukerjee*
86. Biopharmaceutical Statistics for Drug Development, *edited by Karl E. Peace*
87. Parts per Million Values for Estimating Quality Levels, *Robert E. Odeh and D. B. Owen*
88. Lognormal Distributions: Theory and Applications, *edited by Edwin L. Crow and Kunio Shimizu*
89. Properties of Estimators for the Gamma Distribution, *K. O. Bowman and L. R. Shenton*
90. Spline Smoothing and Nonparametric Regression, *Randall L. Eubank*
91. Linear Least Squares Computations, *R. W. Farebrother*
92. Exploring Statistics, *Damaraju Raghavarao*
93. Applied Time Series Analysis for Business and Economic Forecasting, *Sufi M. Nazem*
94. Bayesian Analysis of Time Series and Dynamic Models, *edited by James C. Spall*
95. The Inverse Gaussian Distribution: Theory, Methodology, and Applications, *Raj S. Chhikara and J. Leroy Folks*
96. Parameter Estimation in Reliability and Life Span Models, *A. Clifford Cohen and Betty Jones Whitten*
97. Pooled Cross-Sectional and Time Series Data Analysis, *Terry E. Dielman*
98. Random Processes: A First Look, Second Edition, Revised and Expanded, *R. Syski*
99. Generalized Poisson Distributions: Properties and Applications, *P. C. Consul*
100. Nonlinear L_p-Norm Estimation, *Rene Gonin and Arthur H. Money*
101. Model Discrimination for Nonlinear Regression Models, *Dale S. Borowiak*
102. Applied Regression Analysis in Econometrics, *Howard E. Doran*
103. Continued Fractions in Statistical Applications, *K. O. Bowman and L. R. Shenton*
104. Statistical Methodology in the Pharmaceutical Sciences, *Donald A. Berry*
105. Experimental Design in Biotechnology, *Perry D. Haaland*
106. Statistical Issues in Drug Research and Development, *edited by Karl E. Peace*
107. Handbook of Nonlinear Regression Models, *David A. Ratkowsky*
108. Robust Regression: Analysis and Applications, *edited by Kenneth D. Lawrence and Jeffrey L. Arthur*
109. Statistical Design and Analysis of Industrial Experiments, *edited by Subir Ghosh*
110. *U*-Statistics: Theory and Practice, *A. J. Lee*
111. A Primer in Probability: Second Edition, Revised and Expanded, *Kathleen Subrahmaniam*
112. Data Quality Control: Theory and Pragmatics, *edited by Gunar E. Liepins and V. R. R. Uppuluri*
113. Engineering Quality by Design: Interpreting the Taguchi Approach, *Thomas B. Barker*
114. Survivorship Analysis for Clinical Studies, *Eugene K. Harris and Adelin Albert*
115. Statistical Analysis of Reliability and Life-Testing Models: Second Edition, *Lee J. Bain and Max Engelhardt*
116. Stochastic Models of Carcinogenesis, *Wai-Yuan Tan*
117. Statistics and Society: Data Collection and Interpretation, Second Edition, Revised and Expanded, *Walter T. Federer*
118. Handbook of Sequential Analysis, *B. K. Ghosh and P. K. Sen*
119. Truncated and Censored Samples: Theory and Applications, *A. Clifford Cohen*
120. Survey Sampling Principles, *E. K. Foreman*
121. Applied Engineering Statistics, *Robert M. Bethea and R. Russell Rhinehart*
122. Sample Size Choice: Charts for Experiments with Linear Models: Second Edition, *Robert E. Odeh and Martin Fox*
123. Handbook of the Logistic Distribution, *edited by N. Balakrishnan*
124. Fundamentals of Biostatistical Inference, *Chap T. Le*
125. Correspondence Analysis Handbook, *J.-P. Benzécri*

126. Quadratic Forms in Random Variables: Theory and Applications, *A. M. Mathai and Serge B. Provost*
127. Confidence Intervals on Variance Components, *Richard K. Burdick and Franklin A. Graybill*
128. Biopharmaceutical Sequential Statistical Applications, *edited by Karl E. Peace*
129. Item Response Theory: Parameter Estimation Techniques, *Frank B. Baker*
130. Survey Sampling: Theory and Methods, *Arijit Chaudhuri and Horst Stenger*
131. Nonparametric Statistical Inference: Third Edition, Revised and Expanded, *Jean Dickinson Gibbons and Subhabrata Chakraborti*
132. Bivariate Discrete Distribution, *Subrahmaniam Kocherlakota and Kathleen Kocherlakota*
133. Design and Analysis of Bioavailability and Bioequivalence Studies, *Shein-Chung Chow and Jen-pei Liu*
134. Multiple Comparisons, Selection, and Applications in Biometry, *edited by Fred M. Hoppe*
135. Cross-Over Experiments: Design, Analysis, and Application, *David A. Ratkowsky, Marc A. Evans, and J. Richard Alldredge*
136. Introduction to Probability and Statistics: Second Edition, Revised and Expanded, *Narayan C. Giri*
137. Applied Analysis of Variance in Behavioral Science, *edited by Lynne K. Edwards*
138. Drug Safety Assessment in Clinical Trials, *edited by Gene S. Gilbert*
139. Design of Experiments: A No-Name Approach, *Thomas J. Lorenzen and Virgil L. Anderson*
140. Statistics in the Pharmaceutical Industry: Second Edition, Revised and Expanded, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
141. Advanced Linear Models: Theory and Applications, *Song-Gui Wang and Shein-Chung Chow*
142. Multistage Selection and Ranking Procedures: Second-Order Asymptotics, *Nitis Mukhopadhyay and Tumulesh K. S. Solanky*
143. Statistical Design and Analysis in Pharmaceutical Science: Validation, Process Controls, and Stability, *Shein-Chung Chow and Jen-pei Liu*
144. Statistical Methods for Engineers and Scientists: Third Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
145. Growth Curves, *Anant M. Kshirsagar and William Boyce Smith*
146. Statistical Bases of Reference Values in Laboratory Medicine, *Eugene K. Harris and James C. Boyd*
147. Randomization Tests: Third Edition, Revised and Expanded, *Eugene S. Edgington*
148. Practical Sampling Techniques: Second Edition, Revised and Expanded, *Ranjan K. Som*
149. Multivariate Statistical Analysis, *Narayan C. Giri*
150. Handbook of the Normal Distribution: Second Edition, Revised and Expanded, *Jagdish K. Patel and Campbell B. Read*
151. Bayesian Biostatistics, *edited by Donald A. Berry and Dalene K. Stangl*
152. Response Surfaces: Designs and Analyses, Second Edition, Revised and Expanded, *André I. Khuri and John A. Cornell*
153. Statistics of Quality, *edited by Subir Ghosh, William R. Schucany, and William B. Smith*
154. Linear and Nonlinear Models for the Analysis of Repeated Measurements, *Edward F. Vonesh and Vernon M. Chinchilli*
155. Handbook of Applied Economic Statistics, *Aman Ullah and David E. A. Giles*
156. Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators, *Marvin H. J. Gruber*
157. Nonparametric Regression and Spline Smoothing: Second Edition, *Randall L. Eubank*
158. Asymptotics, Nonparametrics, and Time Series, *edited by Subir Ghosh*
159. Multivariate Analysis, Design of Experiments, and Survey Sampling, *edited by Subir Ghosh*

160. Statistical Process Monitoring and Control, *edited by Sung H. Park and G. Geoffrey Vining*
161. Statistics for the 21st Century: Methodologies for Applications of the Future, *edited by C. R. Rao and Gábor J. Székely*
162. Probability and Statistical Inference, *Nitis Mukhopadhyay*
163. Handbook of Stochastic Analysis and Applications, *edited by D. Kannan and V. Lakshmikantham*
164. Testing for Normality, *Henry C. Thode, Jr.*
165. Handbook of Applied Econometrics and Statistical Inference, *edited by Aman Ullah, Alan T. K. Wan, and Anoop Chaturvedi*
166. Visualizing Statistical Models and Concepts, *R. W. Farebrother*
167. Financial and Actuarial Statistics: An Introduction, *Dale S. Borowiak*
168. Nonparametric Statistical Inference: Fourth Edition, Revised and Expanded, *Jean Dickinson Gibbons and Subhabrata Chakraborti*
169. Computer-Aided Econometrics, *edited by David E. A. Giles*
170. The EM Algorithm and Related Statistical Models, *edited by Michiko Watanabe and Kazunori Yamaguchi*
171. Multivariate Statistical Analysis: Second Edition, Revised and Expanded, *Narayan C. Giri*
172. Computational Methods in Statistics and Econometrics, *Hisashi Tanizaki*
173. Applied Sequential Methodologies: Real-World Examples with Data Analysis, *edited by Nitis Mukhopadhyay, Sujay Datta, and Saibal Chattopadhyay*

Additional Volumes in Preparation

Applied Sequential Methodologies

Real-World Examples with Data Analysis

edited by

Nitis Mukhopadhyay

University of Connecticut
Storrs, Connecticut, U.S.A.

Sujay Datta

Northern Michigan University
Marquette, Michigan, U.S.A.

Saibal Chattopadhyay

Indian Insitute of Management
Calcutta, India

MARCEL DEKKER

MARCEL DEKKER, INC. NEW YORK • BASEL

Transferred to Digital Printing 2005

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

ISBN: 0-8247-5395-X

Headquarters
Marcel Dekker, Inc., 270 Madison Avenue, New York, NY 10016, U.S.A.
tel: 212-696-9000; fax: 212-685-4540

Distribution and Customer Service
Marcel Dekker, Inc., Cimarron Road, Monticello, New York 12701, U.S.A.
tel: 800-228-1160; fax: 845-796-1772

Eastern Hemisphere Distribution
Marcel Dekker AG, Hutgasse 4, Postfach 812, CH-4001 Basel, Switzerland
tel: 41-61-260-6300; fax: 41-61-260-6333

World Wide Web
http://www.dekker.com

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the headquarters address above.

Current printing (last digit):

10 9 8 7 6 5 4 3 2 1

In celebration of the brilliant career of Professor Anis Mukhopadhyay, my elder brother and teacher, and in recognition of his recent retirement from the Indian Statistical Institute, Calcutta, this volume is presented to him with love and affection.

Nitis Mukhopadhyay

In the memory of my late father, in recognition of my mother's lifelong dedication to the well-being of her son, with deepest gratitude to my ever-inspiring wife, and in loving acknowledgement of many sweet distractions from my two-year old daughter.

Sujay Datta

In loving memories of my late sister and father, in admiration of the untiring effort of my mother toward my upbringing, and in recognition of the support and encouragement of my wife and children.

Saibal Chattopadhyay

APPRECIATION

To those colleagues who most kindly offered to help and freely shared their expertise and vision at various junctures of editing this volume, the Co-Editors express their sincerest gratitude and appreciation.

Many colleagues helped tremendously in the editorial process by diligently sharing the burden of refereeing one or more articles. What a difference each individual has made! The Co-Editors thank each referee for showing unselfish dedication and unmistakable enthusiasm.

The Co-Editors consider it a privilege on their part
to mention all referees by name:

Douglas A. Abraham
Makoto Aoshima
Uttam Bandyopadhyay
Tathagata Banerjee
Michael I. Baron
Atanu Biswas
Arup Bose
Benzion Boukai
Probal Chaudhuri
Pinyuen Chen
Sam Efromovich
Joseph Glaz
Robert W. Keener
Subrata Kundu
Wei Liu
Rahul Mukerjee
Adam T. Martinsek
Madhuri Mulekar
Connie Page
William F. Rosenberger
Pranab K. Sen
Sugata Sen Roy
Tumulesh K. S. Solanky
Alexander G. Tartakovsky
Rand R. Wilcox
Peter Willett
Linda J. Young
Shelemyahu Zacks

Preface

Since the publication of Abraham Wald's classic text, *Sequential Analysis*, in 1947, a particularly impressive list of monographs has appeared in this field. These have led to an enormous growth in research methodologies. Some monographs boldly charted the research-track and created an indelible mark of tradition that is so well-known in sequential analysis. It is a fitting tribute to the authors that these volumes continue to serve as flag-bearers and resource guides in this field. We cite some of these influential volumes here:

Bechhofer, R.E., Kiefer, J. and Sobel, M. (1968). *Sequential Identification and Ranking Procedures.* University of Chicago Press: Chicago.

Berry, D.A. and Fristedt, B. (1985). *Bandit Problems.* Chapman & Hall: New York.

Chernoff, H. (1972). *Sequential Analysis and Optimal Design.* CBMS #8. SIAM: Philadelphia.

Chow, Y.S., Robbins, H. and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin: *Boston.*

Ghosh, B.K. (1970). *Sequential Tests of Statistical Hypotheses.* Addison-Wesley: Reading.

Ghosh, B.K. and Sen, P.K. (1991). *Handbook of Sequential Analysis*, edited volume. Marcel Dekker: New York.

Ghosh, M., Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation.* Wiley: New York.

Govindarajulu, Z. (1981). *The Sequential Statistical Analysis.* American Sciences Press: Columbus.

Gut, Allan (1988). *Stopped Random Walks: Limit Theorems and Applications.* Springer-Verlag: New York.

Mukhopadhyay, N. and Solanky, T.K.S. (1994). *Multistage Selection and Ranking Procedures.* Marcel Dekker: New York.

Sen, P.K. (1981). *Sequential Nonparametrics.* Wiley: New York.

Sen, P.K. (1985). *Theory and Applications of Sequential Nonparametrics.* CBMS #49. SIAM: Philadelphia.

Shiryaev, A.N. (1978). *Optimal Stopping Rules.* Springer-Verlag: New York.

Wald, A. (1947). *Sequential Analysis.* Wiley: New York.

Wetherill, G.B. (1975). *Sequential Methods in Statistics*, 2nd ed. Chapman & Hall: London.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* CBMS #39. SIAM: Philadelphia.

While one continues to draw inspiration from these exclusive publications, we may add that some leading 'non-sequential' books have included important material from this area too. We cite, for example, the following monographs:

Bechhofer, R.E., Santner, T.J. and Goldsman, D.M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. Wiley: New York.
Gibbons, J.D., Olkin, I. and Sobel, M. (1977). *Selecting and Ordering Populations*. Wiley: New York.
Gupta, S.S. and Panchapakesan, S. (1979). *Multiple Decision Theory*. Wiley: New York.
Rao, C.R. (1973). *Linear Statistical Inference*, 2nd ed. Wiley: New York.
Zacks, S. (1971). *The Theory of Statistical Inference*. Wiley: New York.

Wald's monograph was unique in its style in 1947 and in many ways it still remains unique largely because Wald's elegantly original mathematical and statistical contributions played a fundamental role in solving practical problems of real-life importance at the time. We surmise, however, that over the years these other volumes have pointed more toward theoretical advancements. Directly or indirectly, purely theoretical contributions have received more encouragement from many quarters and hence the theory of sequential analysis has indeed become very rich. Unfortunately, at the same time, real applications have taken serious hits.

We are personally convinced that this field can and should interface with every conceivable applied area of statistics. But since this field has not been accessible to practitioners for widespread real-world applications, we believe that its popularity among statisticians has dwindled. Real-life experimental data are rarely presented or discussed in sequential books and journal articles. This frustrating situation amounting to what may be viewed as a 'death sentence' has developed over many decades and sadly, this otherwise attractive field with such great promise has alienated itself nearly completely from most practitioners in statistical sciences.

One notable exception, in our view, is the area of clinical trials which has continued to be the major beneficiary of some of the basic research in sequential methodologies. Again, we cite some influential volumes in this area:

Armitage, P. (1975). *Sequential Medical Trials*, 2nd ed. Blackwell Scientific Publications: Oxford.
Jennison, C. and Turnbull, B.W. (1999). *Group Sequential Methods with Applications in Clinical Trials*. Chapman & Hall: London.

Rosenberger, W.F. and Lachin, J.M. (2002). *Randomization in Clinical Trials: Theory and Practice*. Wiley: New York.

But, a specialized field such as sequential analysis cannot be expected to thrive solely on applications in just one area of statistics. It is time for everyone involved to join an aggressive pursuit of real applications of sequential methodologies in as many contemporary and interesting problems of statistics as possible. We believe that time is quickly running out for purely theoretical researchers in this field to continue building newer levels of ivory towers and living in them!

Together, we all must make sequential analysis accessible to all practitioners in statistics. The idea that it is all right for sequential analysis to remain esoteric since a practitioner can seek assistance from a sequential analyst whenever needed remains as far-fetched as ever. That attitude has not worked in the last fifty years and is certainly not about to work now. We urge sequential analysts to take the initiative to vigorously 'market' their methodologies themselves — someone else can hardly ever be expected to do that for us. The field has survived thus far largely by perpetuating the idea of potential applications in the sense that somebody else may eventually use sequential methodologies somewhere in solving real-life problems some day! But, when we look at the bigger picture today, it becomes abundantly clear that sequential analysis has nearly lost its deserving place in the realm of applied statistics. This field has been ignored by nearly every practicing statistician. This is why we strongly feel that it is incumbent on all researchers in sequential analysis to try to rebuild this field's image and market their products themselves. It may eventually mean the difference between 'life' and 'death' of our wonderful field.

We urge everyone to energetically engage in turning the situation around in a positive way because there is still a great deal of hope out there. We believe that the spectrum of applications of sequential methodologies is much broader than what one finds in some of the so-called mainstream statistical monographs and journals. A variety of interesting and important real applications already exist. We dare to dream that the present volume will help in narrowing the unhealthy gap that has existed far too long between the theory and practice of sequential methodologies in problems and issues of broader interest. For us it will indeed be a dream come true if this volume serves as a catalyst to raise the level of consciousness of all sequential analysts about the current status of the field and to inspire them to fine-tune the focus of their initiatives appropriately from time to time so that sequential analysis becomes more relevant to contemporary statistical applications.

The contributing authors for the present volume of collected papers were earnestly requested to adhere to a set of general guidelines including the following:

> *"Every article should discuss clearly at least one substantive applied problem and the appropriate sequential method(s). Tangential references to potential applications are strongly discouraged. A specific application should remain in focus and guide throughout the development and/or implementation of a methodology. That is, each article should justify the relevance, importance, and usefulness of sequential methodology by highlighting an application and the associated gains with the help of real data. Theoretical developments, specific to a problem on hand, will be most welcome but their practical usefulness should be demonstrated.*
>
> *Real applications are encouraged rather than potential applications. An article will preferably include the data or refer its readers to the source of the data or provide a web-site-address if appropriate. Each article will be anonymously refereed.*
>
> *The exposition should be such that any interested reader may readily appreciate the importance of the practical problem(s) discussed and the conclusions drawn. The idea is that the variety of applied problem(s) considered in this volume will ultimately entice readers to take a look at the methodologies even if they do not consider themselves as sequential analysts. We hope to demonstrate that mathematical sophistication and complexity need not deter enthusiastic practitioners to take a look at this field which has plenty to offer in terms of everyday statistics and as it turns out, sequential methodologies are indeed often essential for solving today's challenging practical problems. It is our belief that with sufficient care, the technical coverage can be judiciously blended with lucidity of presentation so that the volume may remain accessible to many users including graduate students and budding researchers, statisticians or otherwise, looking for exposure to this area. At the same time, some hard-core researchers in sequential analysis would be expected to benefit significantly from seeing real-world applications of our craft."*

We had a modest set of goals. We clearly understood that we simply could not continue doing business as usual. We wanted to present the material in such a way that sequential analysts would get a taste of real-life problem-solving which could, in turn, inspire more methodological work in the near future. At the same time, we wanted to make sure that those scientists who were not thoroughly familiar with sequential analysis would also benefit from this volume by observing sequential methodologies at work in the real world. We thank the authors for trying their very best to address our seemingly unending list of demands like these and others.

In the early stages of planning, we invited a number of leading scientists in many substantive areas of applications including Agricultural Statistics, Animal Abundance, Bayesian Strategies, Biometry, Clinical Trials, Computer Simulation, Data Mining, Ecology, Engineering, Finance, Fisheries, Genetics, Multiple Comparisons, Multivariate Analysis, Nonparametrics, Psychology, Sonar Detection, Tracking, and Time Series to contribute specially prepared articles. However, on account of tight deadlines set by us or due to other commitments, some invited authors could not participate in this project. We deeply regret this and their contributions are sorely missed. To those who have kindly participated in our crusade to revive our field's relevance and image in today's statistical world, we remain eternally grateful.

To our true delight, we report that this volume includes interesting methodological articles on:

- *passive acoustic detection of marine mammals* (Abraham)
- *selecting the best component* (Aoshima, Aoki, and Kai)
- *randomization tests* (Banerjee and Ghosh)
- *multistate processes* (Barón)
- *adaptive designs for clinical trials with longitudinal responses* (Biswas and Dewanji)
- *data mining* (Chang and Martinsek)
- *approximations for moving sums of discrete random variables* (Chen and Glaz)
- *measurement-error model* (Datta and Chattopadhyay)
- *density estimation of wool fiber diameter* (de Silva and Mukhopadhyay)
- *financial applications of nonparametric curve estimation* (Efromovich)
- *interim and terminal analyses of clinical trials* (Lai)
- *tests for target tracking* (Li and Solanky)
- *multiple comparisons* (Liu)
- *designing computer simulations* (Mukhopadhyay and Cicconetti)
- *estimation in the agricultural sciences* (Mulekar and Young)
- *contrasting group-sequential and time-sequential interim analysis in clinical trials* (Sen)
- *change-point detection in multichannel and distributed systems* (Tartakovsky and Veeravalli)
- *two-stage multiple comparison procedures in Psychology* (Wilcox)
- *testing in the agricultural sciences* (Young), and
- *ordering genes* (Zacks and Rogatko).

We believe that this is quite an impressive list indeed.

More than one colleague refereed each paper anonymously. The authors revised their manuscripts diligently by taking into account all constructive suggestions and criticisms from the referees. What one finds in this volume is a direct result of total commitment as well as unending patience and support from all parties involved. We remain indebted to this enthusiastic group of colleagues.

We admit that the twenty articles included here are not all written at the same level and personally, we view this disparity positively. Seeing this unevenness in some places, the readers will probably come to realize more that routine phrases such as "applied statistics" and "usefulness of a methodology" are also subject to interpretation.

It is our belief that many students, researchers, or practitioners will find in this volume some important and interesting material. The volume can be used both as reference material as well as a solo textbook. One can also use it as a companion with another book while offering a senior undergraduate or graduate level course in sequential methods. An experienced teacher may also discover a number of hidden or not-so-hidden ideas on conducting hands-on practical experiments to gather real or realistic data that would make a traditional offering of a course in sequential analysis more interesting, lively, and above all, relevant.

Even though this is a substantial volume in itself on applications of sequential methodologies, in no way do we claim that this represents all types of applications. To be truthful, it is far from it. We hope to have other opportunities in the future to be more inclusive and capture a greater diversity of applications. If the present volume makes readers realize that this is a field with a great deal of promise for both intra-disciplinary work in statistics as well as for all types of inter-disciplinary work, then that will be our most gratifying reward.

It has been a real pleasure to work with the editorial and production staff at Marcel Dekker in planning and completing this project. We specially mention Ms. Maria Allegra and Ms. Helen Paisner and thank them both. Without their patience and constant support, this project could not have reached this stage. We are also thankful to several technical experts who helped us at various stages of the compilation process, especially to Dr. Andrew A. Poe from the Department of Mathematics and Computer Science, Northern Michigan University and Professor Uttam Sarkar from the Indian Institute of Management Calcutta. In addition, two of us (Datta and Chattopadhyay) gratefully acknowledge the support received in the form of a faculty grant from Northern Michigan University and a research grant from the Center for Management Development Studies, Indian Institute of Management

Calcutta, respectively. One of us (Mukhopadhyay) gratefully acknowledges the support received through a sabbatical leave from the University of Connecticut in the fall semester 2003 and expresses heartfelt gratitude to Dr. Tom F. Babor, Professor and Head of the Department of Community Medicine, University of Connecticut Health Center, for gracious hospitalities and use of their facilities.

Both collectively and individually, we express indebtedness to our colleagues, students and staff at our home institutions. We must apologize, however, for not mentioning them by name. Mr. Ranjan Mukhopadhyay and Ms. Cathy Brown have rendered invaluable help during the final preparation of a camera-ready copy. Our sincerest thanks go to both Ranjan and Cathy.

Finally, we express our deepest sense of gratitude to our families for their never-ending encouragement and love that gave us the ultimate courage to shoulder this challenging project in the first place.

Nitis Mukhopadhyay
Sujay Datta
Saibal Chattopadhyay

Contents

Contributing Authors

Douglas A. Abraham

Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, PA 16804, U.S.A. E-mail: daa10@psu.edu

Mitsuru Aoki

Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan. E-mail: aoki@math.tsukuba.ac.jp

Makoto Aoshima

Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan. E-mail: aoshima@math.tsukuba.ac.jp

Tathagata Banerjee

Department of Statistics, Calcutta University, 35 Ballygunge Circular Road, Kolkata 700019, India. E-mail: btathaga@yahoo.com

Michael I. Barón

Department of Statistics, University of Texas at Dallas, Richardson, TX 75080, U.S.A. E-mail: mbaron@utdallas.edu

Atanu Biswas

Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, India. E-mail: atanu@isical.ac.in

Yuan-Chin Ivan C. Chang

Institute of Statistical Science, Academia Sinica, 128, Sec. 2, Academia Road, Taipei, Taiwan 11529. E-mail: ycchang@sinica.edu.tw

Saibal Chattopadhyay

Operations Management Group, Indian Institute of Management Calcutta, Joka, D.H. Road, PO Box 16757, Alipore, Kolkata 700027, INDIA. E-mail: chattopa@hotmail.com

Jie Chen

Department of Computing Services, University of Massachusetts, 100 Morrissey Boulevard, Boston, MA 02125, U.S.A. E-mail: jie.chen @umb.edu

Greg Cicconetti

Department of Mathematical Sciences, Muhlenberg College, Allentown, PA 18104, U.S.A. E- Mail: cicconet@muhlenberg.edu

Sujay Datta

Department of Mathematics, Statistics and Computer Science, Northern Michigan University, 1401 Presque Isle Avenue, Marquette, MI 49855, U.S.A. E-mail: sdatta@euclid.acs.nmu.edu

Basil M. de Silva

Department of Mathematics and Statistics, RMIT University City Campus, GPO Box 2476V, Melbourne, Victoria 3001, Australia. E-Mail: desilva@rmit.edu.au

Anup Dewanji

Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, India. E-mail: dewanjia@isical.ac.in

Sam Efromovich

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131-1141, U.S.A. E-mail:efrom@math.unm.edu

Onkar P. Ghosh

Directorate General of Commercial Intelligence and Statistics, Govt. of India, 1 Council House Street, Kolkata 700001, India. E-mail: onkar-ghosh@yahoo.co.uk

Joseph Glaz

Department of Statistics, UBox 4120, University of Connecticut, Storrs, CT 06269-4120, U.S.A. E-Mail: glaz@uconnvm.uconn.edu

Masaki Kai

Insurance Distribution, Fidelity Investments Japan Limited, Tokyo 104-0033, Japan. E-mail: masaki.kai@fid-intl.com

Tze L. Lai

Department of Statistics, Stanford University, Stanford, CA 94305-4065, U.S.A. E-mail: lait@stat.stanford.edu

X. Rong Li

Department of Electrical Engineering, University of New Orleans, New Orleans, LA 70148, U.S.A. E-mail: xli@uno.edu

Wei Liu

Department of Mathematics, University of Southampton, SO17 1BJ, U.K. E-mail: W.Liu@maths.soton.ac.uk

Adam T. Martinsek

Department of Statistics, University of Illinois, 725 South Wright Street, Urbana, IL 61801, U.S.A. E-mail: martins@stat.uiuc.edu

Nitis Mukhopadhyay

Department of Statistics, UBox 4120, University of Connecticut, Storrs, CT 06269-4120, U.S.A. E-Mail: mukhop@uconnvm.uconn.edu

Madhuri Mulekar

Department of Mathematics and Statistics, University of South Alabama, 307 University Boulevard, ILB 325, Mobile, AL 36688-0002, U.S.A. E-mail: mmulekar@jaguar1.usouthal.edu

André Rogatko

Department of Biostatistics, Fox Chase Cancer Center, Philadelphia, PA 19111-2497. E-mail: A_Rogatko@fccc.edu

Pranab K. Sen

Department of Biostatistics, 3105E McGavran-Greenberg, CB #7420, School of Public Health, University of North Carolina, Chapel Hill, NC 27599-7400, U.S.A. E-mail: pksen@bios.unc.edu

Tumulesh K. S. Solanky

Department of Mathematics, University of New Orleans, New Orleans, LA 70148, U.S.A. E-mail: tsolanky@uno.edu

Alexander G. Tartakovsky

Center for Applied Mathematical Sciences, University of Southern California, 1042 Downey Way, DRB-155, Los Angeles, CA 90089-1113, U.S.A. E-mail: tartakov@math.usc.edu

Venugopal V. Veeravalli

Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 128 Coordinated Science Laboratory, 1308 West Main Street, Urbana, IL 61801, U.S.A. E-mail: vvv@ uiuc.edu

Rand R. Wilcox

Department of Psychology, University of Southern California, Los Angeles, CA 90089-1061, U.S.A. E-mail: rwilcox@usc.edu

Linda J. Young

Department of Biostatistics, College of Medicine, P.O. Box 100212, Gainesville, FL 32610, U.S.A. E-mail: lyoung@biostat.ufl.edu

Shelemyahu Zacks

Department of Mathematical Sciences, Binghamton University, Bingham -ton, NY 13902-6000, U.S.A. E-mail: Shelly@math.Binghamton.edu.

Chapter 1

Passive Acoustic Detection of Marine Mammals Using Page's Test

DOUGLAS A. ABRAHAM
SACLANT Undersea Research Centre, La Spezia, Italy

1.1 INTRODUCTION

Sonar signal processing is a subset of signal processing related to the analysis of acoustic signals measured underwater. Applications of sonar signal processing lie in diverse fields such as oil exploration, marine mammal study and naval warfare. The fundamental objectives in sonar signal processing are the detection, classification, and localization of sounds that are heard under water. Sonar systems can be either active or passive in their use of sound (Burdic (1984), Urick (1983)). Passive sonar systems only process signals that are recorded on underwater microphones called hydrophones. A typical sonar system will use many hydrophones that are held together physically in what is called an array. An active sonar system transmits a signal using an underwater loud speaker and processes the subsequently heard reflections.

Many of the signal processing algorithms that are used in sonar systems were developed using methodologies from statistical decision theory. Detectors and classifiers may be formulated as binary and multiple hypothesis tests. Localization is simply the estimation of the physical location of the sound emitter or reflector. Sequential methodologies have primarily been used for the detection of sounds, but have also seen use in localization procedures such as target tracking (Lerro and Bar-Shalom (1993)). This paper will focus on the application of the cumulative summation type of sequential procedure called Page's test. Page (1954) developed the procedure to determine the time at which a change in the distribution of sequentially obtained data had occurred. It has seen significant applications in areas such as fault detection (Basseville and Nikiforov (1993)), but has also been used for the detection of unknown but finite duration signals (Han et al. (1999)). It is this latter application that is useful in both active and passive sonar signal processing. Page's test, when it utilizes the log-likelihood ratio, is known to be optimal in the sense of minimizing the average delay before detection while constraining the average time between false alarms (Lorden (1971), Moustakides (1986)). Analysis of its performance at detecting finite duration signals is not trivial, though accurate approximations exist for simple configurations (Han et al. (1999)).

As an example of an active sonar signal processing application, Page's test has been used to detect target echoes when they are corrupted by propagation through a shallow water environment (Abraham and Willett (2002), Abraham (1996b)). In shallow water environments, the sound travels from the source to the target and then from the target to the receiving hydrophones through many paths. As the time it takes to traverse each path differs, the received signal appears to be spread over time. Owing to the difficulty in accurately modeling the ocean propagation, Page's test was applied to obtain adequate detection performance over a wide range of environmental conditions.

An example of a passive sonar signal processing application, and the focus of this paper, may be found in the detection of marine mammals by their acoustic emissions. Such non-invasive detection of marine mammals is necessary for the study of their habits as well as to aid in ensuring their absence prior to any potentially harmful activity such as oil exploration or naval testing. In this passive sonar application, the sounds generated by marine mammals are recorded on hydrophone arrays and must be discerned from all background noises. As marine mammals vary considerably in size and vocal characteristics, the detector must be quite flexible and able to detect both short and long duration signals with potentially widely varying frequency content. In the following sections the problem will be described in more detail along with the sequential detector that was implemented and some results from the analysis of real data. A more detailed description of the analysis may be found in the SACLANT Undersea Research Centre report (Abraham (2000)) from which most of this article is derived.

1.2 ACOUSTIC DETECTION OF MARINE MAMMALS

Detection of the presence of marine mammals is crucial in the vicinity of a research vessel carrying out operations involving the projection of acoustic energy into the local ocean environment. The following sections detail an analysis of passive acoustic data garnered from data recorded during the SACLANT Undersea Research Centre's SWAC4 sea-trial in Kyparissiakos Gulf off the coast of Greece during May 1996. Specifically, the automatic detection processing that was implemented will be described and examples of the data processing results presented.

The automatic detection of marine mammal acoustic emissions is a difficult task. The data available are in the form of passive sonar recordings from a hydrophone array that have been "beamformed" to emphasize signals coming from certain directions. This process increases the signal-to-noise ratio and aids in the localization of the sound source by dictating on which bearing it is heard. Unfortunately, it also means that the detection algorithm must be applied to the data from each beam (direction). The data from the SWAC4 sea-trial were beamformed to point to 120 directions spanning from forward to aft. Designing the detection algorithm is difficult owing to the wide variety of frequencies, bandwidths, and time characteristics of marine mammal acoustic emissions from different species. Take, for example, the sperm whale clicks shown in Figure 1.2.1. These data are in the frequency range of 750–1500 Hertz owing to the limitations imposed by the experimental conditions and acquisition systems. It is known that marine mammals emit sound over a much wider frequency band. Nevertheless, even with this limited band, it is seen that the spectrum is not necessarily flat and that the time series exhibits the effects of multipath propagation, spreading the signal energy in time. As other types of waveforms are expected (for example, whistles or sweeps) a detector is desired that is flexible both in the time duration of the waveform and its frequency content.

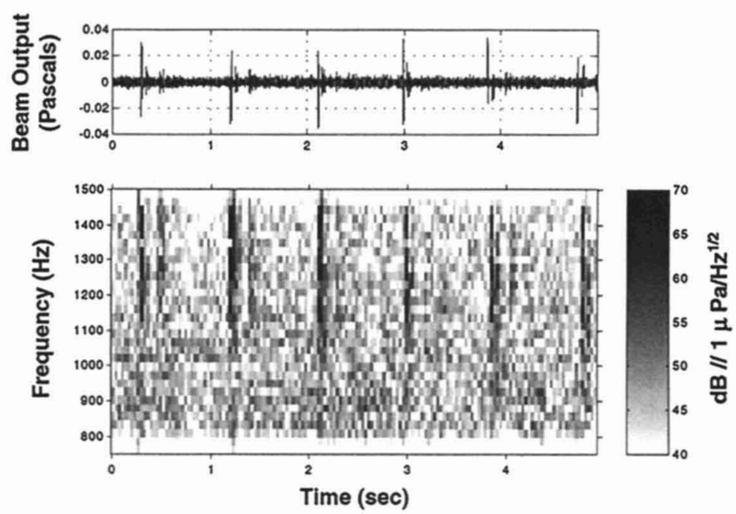


Figure 1.2.1: Time Series and Frequency Spectrum of Sperm Whale Click Train Limited to 750–1500 Hertz Frequency Band.

1.3 METHODOLOGY AND ANALYSIS

In hypothesis testing, there must be some difference between the null and alternative hypotheses in order for the Type-I and Type-II errors to be small. In terms of detector design, this translates into requiring that there be some means for distinguishing the signal to be detected from the background noise and any interferences. As this situation calls for a very general detector, not tuned to any specific characteristics of individual marine mammal emissions, the only distinction from the background or interferences is time duration. Thus, a detector is desired that finds short duration signals that are not similar to the more slowly changing background (for example, ambient ocean noise from the ocean surface or distant shipping) or interferences (for example, near-by surface vessels including the ship towing the hydrophone array). Additionally, the detector should be robust to varying signal duration and frequency content. A detector with these characteristics was proposed by Abraham and Stahl (1996) by combining Page's (1954) test with the power-law processor of Nuttall (1994) for the combination of *discrete Fourier transform* (DFT) bin outputs. A block diagram of this detector structure is shown in the left half of Figure 1.3.1. The time series data are transformed into the frequency domain by overlapping DFTs. The duration of the DFT should be near or less than the duration of the shortest signal that may be encountered. Define the magnitude squared of the DFT bins of interest for the k^{th} DFT as $\{X_{k,1}, X_{k,2}, \ldots, X_{k,m}\}$. For the data presented in this paper the bins of interest are those in the 750–1500 Hertz band. Let the estimated background power in the j^{th} DFT bin at time k be $\lambda_{k,j}$ for $j = 1, \ldots, m$. As will soon be described, the background power levels in each DFT bin need to be estimated from previously observed data. These estimates are used to form normalized DFT bin data,

$$Y_{k,j} = \frac{X_{k,j}}{\lambda_{k,j}}, \tag{1.3.1}$$

which are combined into a single test statistic by a power-law non-linearity,

$$Z_k = \left(\frac{1}{m} \sum_{j=1}^{m} Y_{k,j}^p \right)^{\frac{1}{p}}. \tag{1.3.2}$$

The power-law non-linearity provides robustness against varying signal bandwidth or frequency structure and can have the effect of either picking the maximum DFT bin output (high power law) or summing all the DFT bins together (power law equal to one) as in an energy detector. This one statistic from each DFT is then used in Page's test to detect the onset of a signal. The update for Page's test has the form

$$W_{k+1} = \max\{0, W_k + Z_{k+1} - b_0\} \tag{1.3.3}$$

where b_0 is a bias and a detection is declared when W_k exceeds a threshold. The bias is most easily chosen through Dyson's (1986) method which uses the average value of Z_k under the null and alternative hypotheses. In certain cases it is possible to choose the bias to optimize an asymptotic performance measure (Abraham (1996a)). The threshold may be chosen according to a desired average time between false alarms through the standard Wald or Siegmund approximations or quantization based approximations (Basseville and Nikiforov (1993), Brook and Evans (1972), Abraham and Stahl (1996)). Assuming perfect normalization, the performance of a power-law non-linearity feeding a Page test is examined in terms of the average sample numbers (average time between false alarms and average delay before detection) in Abraham and Stahl (1996). A more relevant measure of detection performance is the probability of detection, which may be obtained from quantization based methods (Han et al. (1999)). Theoretical analyses such as those found in Abraham and Stahl (1996), Han et al. (1999) or Nuttall (1994) typically assume that the normalization is perfect and hence that the DFT bin data are exponentially distributed with unit mean when no signal is present. It is often assumed that the signal manifests itself as either a change in the scale or as a constant additive component to the complex DFT data, yielding a non-central chi-squared distribution for the DFT power data. Though these theoretical analyses help evaluate the performance of various detector structures under ideal conditions, the processing of real data introduces many difficulties ranging from data quality issues (for example, data glitches or drop-outs that can occur with varying frequency) to inadequate modeling (simplifications in the modeling that allow analysis but only approximately represent the real data). Thus, in practice, signal strength based parameters such as the bias in (1.3.3) are usually set for a minimum detectable signal level and thresholds are typically chosen so that the operator and any post-detection processing are not overloaded with false alarms.

In this application, Page's test provides robustness to the unknown duration of the signal compared with the use of a sliding fixed block detector that would be tuned to a single signal duration. Robustness in this sense means that the detector provides adequate, though perhaps sub-optimal, performance over a wide range of signal durations. As estimates of both the start and end times of the signal are desired, the alternating-hypothesis form of Page's test (Streit (1995), Abraham and Willett (2002)) must be implemented. The update structure for this form is similar to (1.3.3) and is described in the following section. In the alternating-hypothesis configuration, the start and end times of a detected event are estimated by the most recent reset of the Page test to its null state as described in Abraham (1997).

Necessary to the implementation of the detector is the estimation of the background noise and interference power at the output of each DFT bin. As these may be considered nuisance parameters (that is, parameters that need to be estimated but are not used to describe the signal), the scheme proposed in Abraham (1996c) which exploits the structure of the Page test to isolate data believed to be signal-free is appropriate. This detector structure uses data prior to the most recent

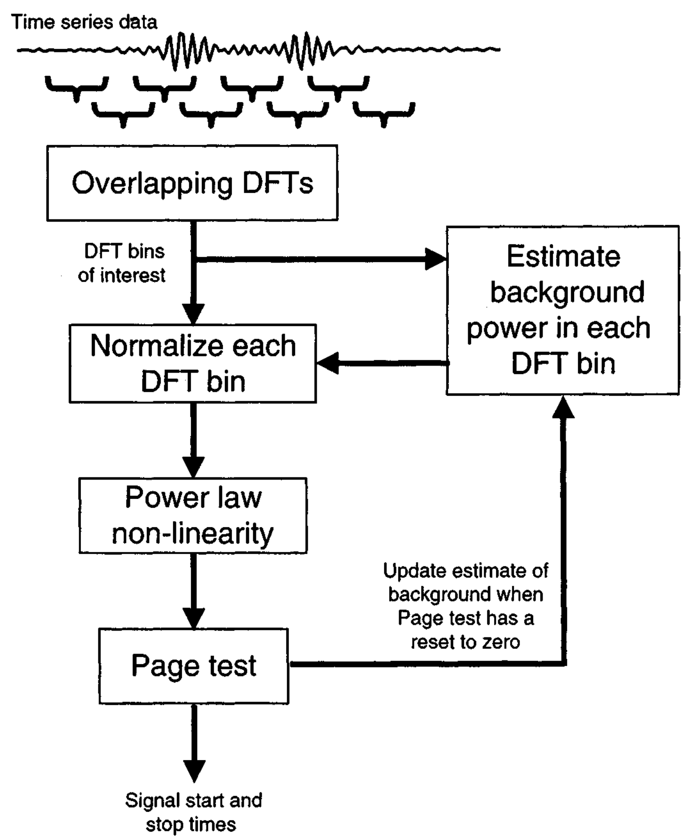


Figure 1.3.1: Block Diagram of Detection Scheme.

reset of the Page test to estimate the nuisance parameters. The form of the background estimator may be of a sliding block or exponentially averaged type. In this case, the latter is chosen because it is (marginally) easier to implement. The exponential averager simply applies a single pole filter to the time series; thus, updating the background estimate results in

$$\lambda_{k+1,j} = \alpha\lambda_{k,j} + (1-\alpha)\, X_{k,j} \tag{1.3.4}$$

where $\alpha \in (0,1)$ is usually near one to provide a long averaging time. Each time the Page test has a reset to zero, the background estimate is updated as in (1.3.4) but using only the non-signal data between the previous reset to zero and the current reset to zero (that is, if a signal was detected in between, those data would not be included). The estimates should be initialized with either a fixed block estimate or an exponential averager that has been running for a short time.

1.3.1 Algorithm Specifics

Following are the algorithm (in pseudocode) and definitions of the variables used to implement the detector along with their values when they require setting. The dependence of the variables on DFT snapshot k is suppressed as it is not necessary when implementing the algorithm.

Detection Algorithm

(1) Initialization

- Set $W = 0$, $k = 1$, $i_0 = 1$, $i_0' = 1$
- Form initial estimate of $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$

(2) Normalization and power-law

- Form normalized DFT bin outputs

  $Y_j = \frac{X_{k,j}}{\lambda_j}$ for $j = 1, \ldots, m$
- Apply power-law to normalized data

  $Z = \left[\frac{1}{m}\sum_{j=1}^{m} Y_j^p\right]^{\frac{1}{p}}$

(3) If $W < h_0$,

- Set $W = \max\{0, W + Z - b_0\}$
- If $W \geq h_0$,
  - The leading edge of a signal has been detected
  - An estimate of the starting time index is i_0
  - Set $W = h_0 + h_1$ and $i_1 = k$
- Else if $W = 0$,
  - A reset to zero has occurred, update background estimate

    for $i = i_0'$ to k and $j = 1, \ldots, m$

    $\lambda_j = \alpha\lambda_j + (1 - \alpha) X_{i,j}$

    end
  - Set $i_0 = k$ and $i_0' = k$

(4) If $W \geq h_0$,

- Set $W = \min\{h_0 + h_1, W + Z - b_1\}$
- If $W \leq h_0$,
  - The lagging edge of a signal has been detected
  - An estimate of the stopping time index is i_1
  - Set $W = 0$, $i_0 = k$, and $i'_0 = i_1$
- If $W = h_0 + h_1$, set $i_1 = k$

(5) Set $k = k + 1$ and go to **(2)**

Description of variables

p - power law ($p \geq 1$, $p = 1$ was used)

h_0 - threshold for signal onset detection ($h_0 = 12$)

b_0 - Page test bias for signal onset detection ($b_0 = 2.5$)

h_1 - threshold for signal termination detection ($h_1 = 10$)

b_1 - Page test bias for signal termination detection ($b_1 = 5$)

α - time constant for exponential averager ($0 < \alpha < 1$, $\alpha = 0.95$ was used)

N_{fft} - size of DFT block ($N_{fft} = 128$)

N_{off} - offset from one DFT block to next ($N_{off} = 32$)

W - Page test statistic

i_0 - index to most recent reset to zero

i'_0 - index for updating background power estimates

i_1 - index to most recent reset to $h_1 + h_0$ (signal present state)

It should be noted that the indices for the starting and stopping times are in terms of DFT blocks and must be converted to time samples based on the DFT size and amount of overlap. Also, the power-law was kept at unity because of the small bandwidth of the data being processed relative to the potential bandwidth of the marine mammal acoustic emissions. In general, data used for the acoustic detection of marine mammals would have a higher bandwidth and may exhibit tonal emissions. In this more common situation, a higher power law would improve detection performance.

1.3.2 Detection Results

The algorithm of Section 1.3.1 was applied to data from Run 9[1] of the SWAC4 data set to obtain a series of start and stop times for every detection. From these start and stop times, the total signal energy was estimated and tabulated along with the current average noise power estimate (here it is assumed that the detected signal is an energy signal and that the background noise is a power signal). Additionally, an estimate of the noise background after removal of the detected signals was formulated every 12 seconds. The detection processing results are then displayed in Figure 1.3.2 as the total energy-to-noise power ratio (ENR) detected on each beam over 6 second intervals for the nearly three hour Run 9. This display only shows the results of the detection processing for short duration signals. Thus, when detections occur over an extended period, as seen in the figure, then the time domain signal has persistent non-stationary components as illustrated by the sperm whale clicks in Figure 1.2.1. As previously mentioned, the beams span from forward to aft and are spaced equally in the cosine space of bearing so that there are more beams broadside to the hydrophone array (which was in the shape of a line) than near the forward or aft directions.

The detection results are grouped into events by visual association over beam and time, as indicated by the numbers in Figure 1.3.2. Event time series are then formed by choosing the beam containing the largest ENR over each 12 second period. These time series were then submitted to Prof. G. Pavan of the University of Pavia, Pavia, Italy for classification. Those shown in Figure 1.3.2 were all classified as sperm whale click trains. The detector also found many signals associated with surface vessels, particularly those from the Research Vessel Alliance (the ship towing the hydrophone array) in the forward beams, and a plethora of isolated detections that could be marine mammal, fish, man-made or false alarms. It is possible to associate some of the detections with surface vessels by overlaying the detection results on the estimated background noise, as shown in Figure 1.3.3 for the first 20 minutes of Run 9. In this figure the background power is shown in gray as indicated by the scale on the right. The ENRs of detected short time duration events are overlaid in black. The surface vessels are clearly visible in the background power estimate as slowly moving lines in the gray scale. Short duration signal detections that overlay

[1]As will be seen in Section 1.3.3, Run 9 is unique in that the data allow for localization of two sperm whales.

the surface vessels in bearing and time are most likely originating there as well (though this is not necessarily so). The sperm whale click trains (events 1, 2 and 3) arrive on quiet beams (that is, there is no surface ship in the background), additionally supporting their classification as marine mammal. Event 21 was eventually classified as acoustic emissions from fish of unknown type. It may also be surmised that, of the two sperm whales detected during the first several minutes, event 1 is nearer than event 2, assuming they both produced approximately the same source levels. As will be discussed in the following section, the towed hydrophone array was completing a turn previously carried out by the RV Alliance so localization of these two events is possible, including resolution of the left/right ambiguity inherent in the array signal processing.

1.3.3 Localization

Passive listening of acoustic emissions inherently only provides bearing information. Additionally, owing to the straight line shape of the hydrophone array, there exists a cone of ambiguity; that is, the sound arriving at the array sounds the same if it arrives from anywhere on a cone axially aligned with the towed line array. All of the runs analyzed were such that the tow ship (RV Alliance) was on a constant bearing. Thus, triangulating detections observed over extended periods of time still results in an ambiguity to the left or right side of the array. However, during the first 10 minutes of Run 9, the array was still completing a turn the tow-ship had made prior to commencing the run. From the array heading information (which is quite noisy) it was possible to localize the two sperm whale click trains detected as shown in Figures 1.3.4 and 1.3.5. Lines along the bearing of the detected events from the position of the RV Alliance are shown for events 1, 4, 10 and 15 of Run 9 in Figure 1.3.4. These events are believed to originate from the same whale, though there is no proof of this other than approximate coincidence in space and time. Each line is 15 nautical miles long and when taken in conjunction with the others form a locus where the sperm whale might have been, effectively localizing the whale in range, bearing and resolving the left/right ambiguity. The lines from event 1 (the lighter ones) are shown assuming that the whale was on either the left or right side of the array. The diverging lines seen to the east of the track illustrate an incorrect localization. Figure 1.3.5 contains the localization of events 2, 3, 5, 6, 8, 12, 13 and 17 of Run 9. The ranging

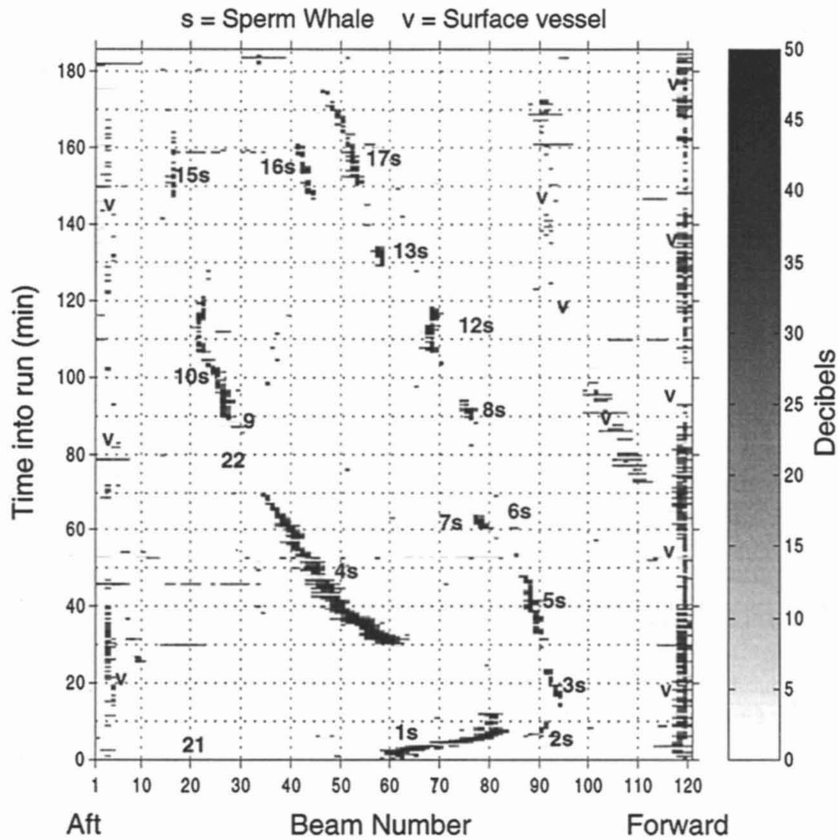


Figure 1.3.2: Signal Energy-to-Noise Power Ratio (ENR) of Detected Short Time Duration Events from Run 9, Combined Over Every 6 Seconds.

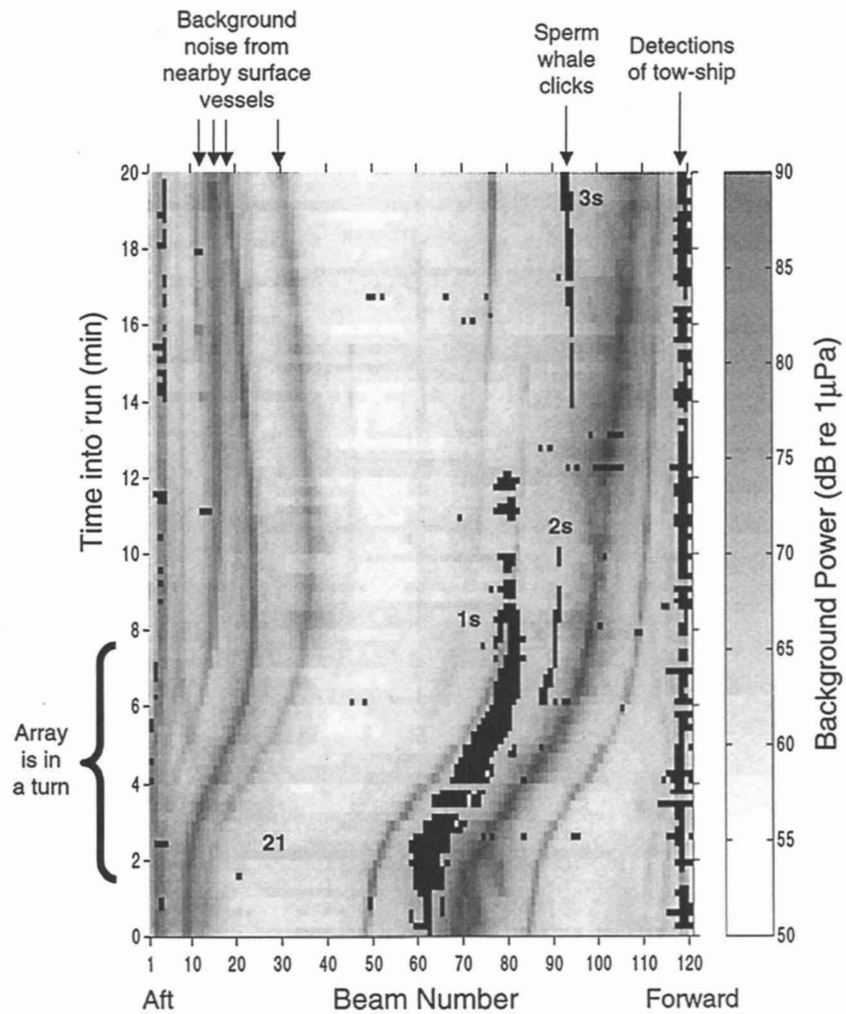


Figure 1.3.3: ENR of Detected Events from First 20 Minutes of Run 9 (Black) Overlaid on Background Noise and Interference Power Estimates (Gray Scale).

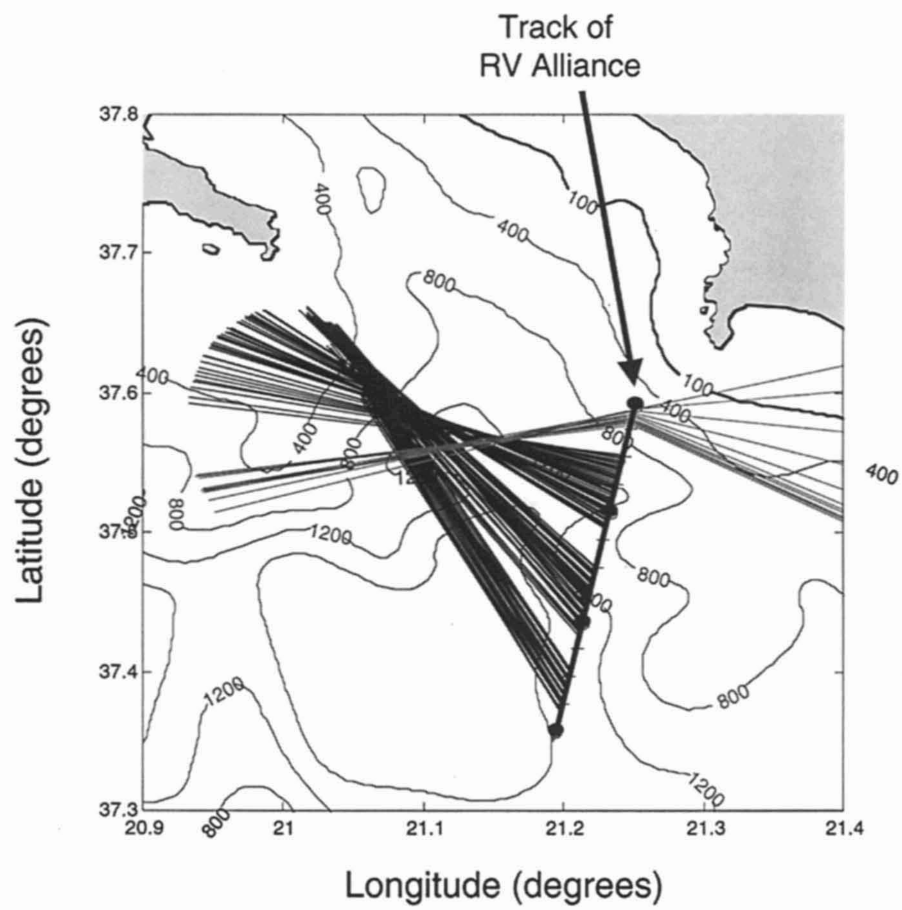


Figure 1.3.4: Lines Along Bearings of Detections for Events 1, 4, 10 and 15 of Run 9. Each Line Is 15 Nautical Miles in Length. The Lighter Lines Are from Event 1 and Illustrate Localization of a Sperm Whale to the West Side of the Track of RV Alliance.

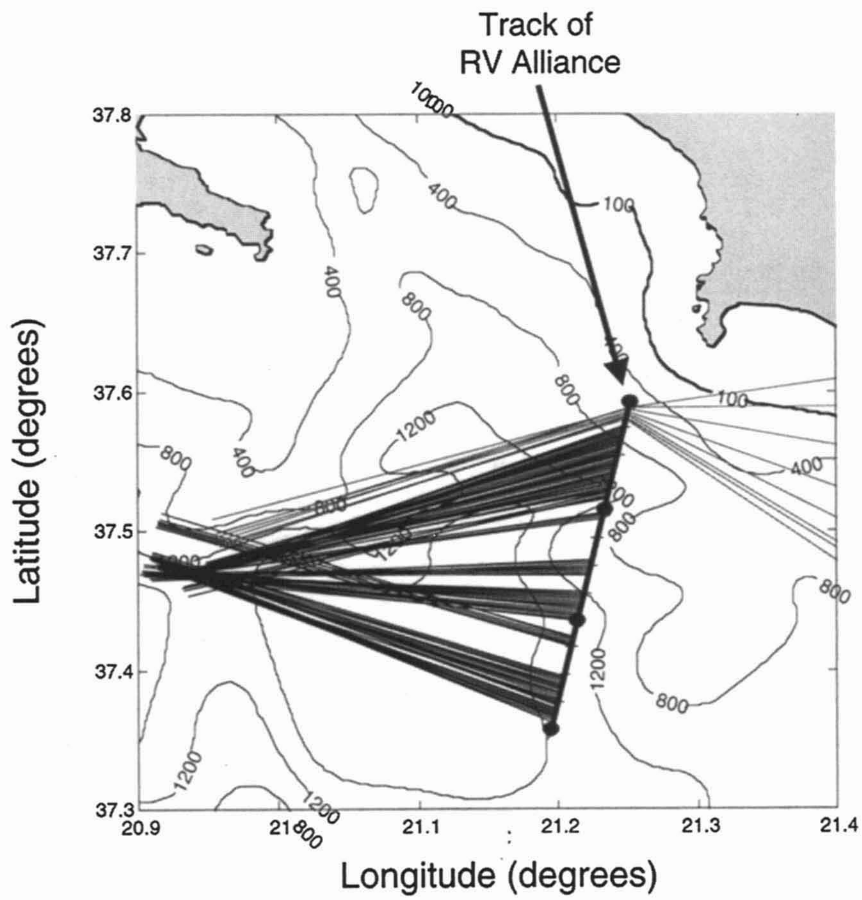


Figure 1.3.5: Lines Along Bearings of Detections for Events 2, 3, 5, 6, 8, 12, 13 and 17 of Run 9. Each Line Is 15 Nautical Miles in Length. The Lighter Lines Are from Event 2 and Illustrate Localization of a Sperm Whale to the West Side of the Track of RV Alliance.

information garnered from Figures 1.3.4 and 1.3.5, that event 2 is farther away than event 1, is corroborated by the ENR levels observed in Figure 1.3.2 where event 2 is weaker than event 1.

1.4 CONCLUDING REMARKS

This paper has illustrated the use of sequential methodologies in both active and passive sonar signal processing. It was shown in detail how Page's test may be applied to the detection of marine mammal acoustic emissions having unknown time and frequency structure. The sequential procedure allows for the detection and segmentation of signals with unknown and potentially widely varying time structure. The algorithm described was effective in estimating the background noise and interference power and then detecting departures from that norm, which in this case were either acoustic emissions from marine mammals or non-stationary acoustic emissions from nearby surface ships. In the data presented, sperm whale sounds were detected, along with numerous other detections not analyzed further owing to limited time and resources. Localization of some of the sperm whale sounds was possible through triangulation.

ACKNOWLEDGMENT

This work was performed while the author was with the SACLANT Undersea Research Centre, La Spezia, Italy. Without the efforts of Walter Zimmer, this work would not have been possible. He designed and implemented a passive sonar system on SACLANT Undersea Research Centre's real-time system, enabling the analysis of large amounts of data and provided the filtered and beamformed data for the automatic detection processing. The author would also like to acknowledge the efforts of Angela D'Amico and Ettore Capriulo who both contributed to the work presented in this paper.

REFERENCES

[1] Abraham, D.A. (1996a). Asymptotically optimal bias for a general non-linearity in Page's test. *IEEE Trans. Aerosp. Elec. Sys.*, **32**, 360–367.

[2] Abraham, D.A. (1996b). A nonparametric Page test applied to active sonar detection. In *Oceans 96 MTS/IEEE Conf. Proc., The Coastal Ocean–Prospects for the 21st Century*, 908–913. IEEE: Piscataway.

[3] Abraham, D.A. (1996c). A Page test with nuisance parameter estimation. *IEEE Trans. Inform. Theory*, **42**, 2242–2252.

[4] Abraham, D.A. (1997). Analysis of a signal starting time estimator based on the Page test statistic. *IEEE Trans. Aerosp. Elec. Sys.*, **33**, 1225–1234.

[5] Abraham, D.A. (2000). Passive acoustic detection of marine mammals. SM-351, SACLANT Undersea Research Centre, La Spezia, Italy.

[6] Abraham, D.A. and Stahl, R.J. (1996). Rapid detection of signals with unknown frequency content using Page's test. In *Proc. 1996 Conf. Inform. Sci. Sys.* (S. Kulkarni and M. Orchard, co-chairs), 809–814. Princeton University: Princeton.

[7] Abraham, D.A. and Willett, P.K. (2002). Active sonar detection in shallow water using the Page test. *IEEE J. Ocean. Eng.*, **27**, 35–46.

[8] Basseville, M. and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Applications.* Prentice-Hall: Englewood Cliffs.

[9] Brook, D. and Evans, D.A. (1972). An approach to the probability distribution of cusum run length. *Biometrika*, **59**, 539–549.

[10] Burdic, W.S. (1984). *Underwater Acoustic System Analysis.* Prentice Hall: Englewood Cliffs.

[11] Dyson, T. (1986). *Topics in Nonlinear Filtering and Detection.* Ph.D. thesis, Princeton University, Princeton.

[12] Han, C., Willett, P.K. and Abraham, D.A. (1999). Some methods to evaluate the performance of Page's test as used to detect transient signals. *IEEE Trans. Signal Proc.*, **47**, 2112–2127.

[13] Lerro, D. and Bar-Shalom, Y. (1993). Interacting multiple model