STATISTICAL ECOLOGY AND ENVIRONMENTAL STATISTICS

Ganapati P. Patil
Department of Statistics, Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, USA

Keywords: Biodiversity measurement and comparison, change detection, diversity profiles, echelon analysis, ecological sampling, ecometrics, environmental indicators and their integration, environmental monitoring and assessment, environmental statistics, environmetrics, hierarchical classified map simulation model, hierarchical Markov transition matrix models, landscape ecology, landscape fragmentation profiles, multiscale assessment, multispectral environmental change detection, observational economy, risk assessment, statistical ecology

Contents

Summary

Ecology is undergoing some major changes in response to changing times of societal concerns coupled with remote sensing information and computer technology. Both theoretical and applied ecology are using more of statistical thought processes and procedures with advancing software and hardware to satisfy public policy and research, variously incorporating sample survey data, intensive site-specific data, and remote sensing image data. Statistical ecology and environmental statistics have numerous challenges and opportunities in the waiting for the twenty-first century. This chapter shares some of the highlights in statistical ecology, environmental statistics, and ecological assessment in this connection.

1. Introduction

Statistical ecology and environmental statistics are in a take-off stage both for reasons of societal challenge and statistical opportunity. It is becoming clear that statistical ecology and environmental statistics are calling for more and more of non-traditional statistical approaches. This is partly because ecological and environmental studies involve space, time, and innovative sampling and monitoring. Also, statistical ecology and environmental statistics must satisfy public policy responsibility in addition to disciplinary and interdisciplinary research. It is only appropriate to attempt a perspective of statistical ecology and environmental statistics within a forward-looking context.

The year 1994 marked the 25th year of statistical ecology and related ecological statistics with reference to the First International Symposium on Statistical Ecology held at Yale in 1969 with G. P. Patil, E. C. Pielou, and W. E. Waters as three co-chairmen representing the fields of statistics, theoretical ecology, and applied ecology respectively. Over the past 25 years, statistical ecology has had a major impact on the collection, analysis, and interpretation of data on various fields of application and their theory. While much progress has been made in the past, the future promises even more rapid developments as sophisticated computing technology is utilized to apply newly developed statistical methods to increasingly detailed data bases in both space and time.

It is no wonder that the Statistical Ecology Section of the International Association for Ecology and the related Liaison Committee on Statistical Ecology of the International Association for Ecology, the International Statistical Institute, and the International Biometric Society have been around since their inception in 1969. And now the Ecological Society of America has a Statistical Ecology Section, and the American Statistical Association has a Section on Statistics and the Environment. The International Biometric Society and the American Statistical Association have together initiated a new *Journal of Agricultural, Biological, and Environmental Statistics*. For five years now, we have had the International Environmetrics Society with its journal, *Environmetrics*. And 1994 saw the creation of the Committee on Environmental Statistics at the International Statistical Institute and the launching of a new cross-disciplinary journal, *Environmental and Ecological Statistics*, published by Kluwer Academic Publishers. Also, an international initiative entitled SPRUCE (Statistics in Public Resources, Utilities, and in Care of the Environment) is now operational.

This space-limited overview necessarily has to be short and subjective. In this review chapter, we share some of the highlights and experiences in statistical ecology, environmental statistics, and risk assessment, giving references at the end for further reading. For purposes of organization, the sections are titled: simple stories but challenging concerns; ecological sampling and statistical inference; biodiversity measurement and comparison; environmental data and cost-effective acquisition; landscape ecology and multiscale assessment; echelon analysis for multispectral environmental change detection; statistics as an instrument to deal with environmental and ecological crisis; future areas of concern and challenge; looking ahead.

Besides a broad spectrum of several important papers focused on a variety of relevant issues and topics in Environmental and Ecological Statistics, witness several timely special issues devoted to select themes, such as: Environmental Monitoring and Assessment, A. R. Olsen, Guest Editor; Space-Time Processes in Environmental and Ecological Studies, Peter Guttorp, Guest Editor; Statistical Design and Analysis with Ranked Set Samples, N. Phillip Ross and Lynne Stokes, Guest Editors; Statistical Toxicology and Toxicological Statistics, Wolfgang Urfer, Guest Editor; and Statistical Ecology and Forest Biometry, Timothy G. Gregoire and Michael Kohl, Guest Editors. A couple more special issues are in the making for the near future on some exciting themes, such as: Spatial Statistics for Production Ecology, Alfred Stein, Guest Editor; Adaptive Sampling, Steven K. Thompson, Guest Editor; Statistical Design and Analysis with Composite Samples, Richard O. Gilbert and Barry D. Nussbaum, Guest Editors; Classified Raster Map Analysis and Cellular Automation for Environmental and Ecological Statistics, Wayne Myers and Charles Taillie, Guest Editors; Regional Environmental Indicators and Their Integration, N. Phillip Ross and Ashbindu Singh, Guest Editors.

2. Simple Stories but Challenging Concerns

2.1. Introduction

Statistical methods were initially developed for use in basic and applied sciences, and later in engineering and management. While basic statistical science is common to all areas, there are specific techniques developed to answer specific questions in each area. Statistical ecology and environmental statistics are relatively new and need some of their own special methodologies. Statistical thinking is an aid to the collection and interpretation of data. It may help clarify seeming confusion. It may help confuse seeming clarity. The statistical approach is expected to contribute to the overall balance, insight and perspective of the substantive issue and its resolution in the light of the evidence on hand, be it in the nature of empirical data, literature-assembled data, expert opinion data, or a combination thereof.

Is statistical ecology a science, technology or art? It is more of a combination of all these. What is the future of statistical ecology and environmental statistics? The future is in cross-disciplinary communication. There will be more emphasis on understanding environmental and ecological data and extracting all the available information rather than answering some routine questions. Statistics will be more a way of thinking or reasoning rather than a tool for beating data to yield answers. An environmental and ecological statistician without any knowledge of ecology and environmental science is like a doctor who has specialized in principles of surgery, but cannot decide where and when surgery is needed for a patient. Science strives for the discovery of significant scientific truth. It is statistics that takes care of the uncertainty of the scientific method consisting of design, analysis and interpretation, and even the assessment of significance. And, the society in which we live has chosen to fully use statistics as a legislative and educational instrument to deal with societal crises, whether they be related to environment, education, economy, energy, engineering or excellence.

2.2. Life and Death with Averages and Variability

‘Average’ is a word which sounds simple and pedestrian in literal usage. It has a status of being optimal in the least squares. In the world of statistics, it is one’s expectation! In the world of quality it sounds mediocre. The statistical average is useless without the qualifying support of variability.

(a) Happy Hunter:

First shot, one inch on the left of the animal; second shot, one inch on the right of the animal. So, on the average, shot on the spot; a perfect average shot!

(b) Tourist:

I wish to cross the river. I cannot swim. Can you help?
Native: Certainly! Average depth of this river around here is known to be well below three feet.
You look to be six.
Tourist: You are encouraging, and yet not quite helpful. Depth is usually uneven. Variability sure is a matter of life and death.

(c) Birds:

Concerned about the typical direction in which disoriented birds of a certain species fly, someone goes out in an open field, stands facing north, and observes a bird vanish at the horizon at an angle of 10 degrees. A little later, he finds a second bird vanish at the horizon at an angle of 350 degrees. What can be said of the typical direction based on the evidence.

After submitting these data to a computer and requesting the average direction, the software returns a value of (10 + 350)/2 = 180 degrees. The report concludes that, on average, the birds are flying south. Of course, the exact opposite is true, demanding correct and appropriate software.

2.3. Innovative Statistical Mind Sets

An important question in ecosystem health assessment is “What type of risk is at stake?”. Are we concerned with the average exposure of the population at risk, or the maximum exposed individual?
Furthermore, are we addressing risks associated with chronic or acute effects of a substance? Still another big question is “How is the contaminant(s) distributed over the site, both spatially and through a variety of media including plant and animal members of the food chain?”.
Such questions determine whether sampling should be designed to estimate average or median concentrations, or to identify “hot spots”, or both. In order to address these questions and satisfy the needs of affected parties, sampling can become very extensive before and after remediation of a site. For this reason, site managers stand to economize greatly by adopting more innovative methods of statistical sampling and analysis. The following may be insightful.

2.4. Comprehensive vs. Comprehensible

Once a hazardous waste site is discovered, we are presented with a situation that we need to clearly comprehend. Often this situation presents a dilemma:
1. For lack of information, we do not quite comprehend the situation.
2. We therefore collect information, tending to collect comprehensive information.
3. Because the information is comprehensive, we do not quite comprehend it.

4. Therefore we summarize the information through a set of indices (statistics) so that it would be comprehensible.
5. Now, however, we do not comprehend quite what the indices exactly mean.
6. Therefore we do not quite comprehend the situation.
7. Thus, without (all) information, or with (partial) information, or with summarized information, we do not quite comprehend a situation!

This dilemma is not to suggest a bleak picture for one's ability to understand, predict, or manage a situation in the face of uncertainty. It is more to suggest a need to clearly state the purpose, formulation and solution for the study under consideration, in line of Data Quality Objectives.

2.5. Space Age/Stone Age

Great effort is made these days to obtain very accurate measurements on the environment at different scales, whether organic chemical concentrations are measured by Gas Chromatography coupled with a double Mass Spectrometer (GC/MS-MS) or landscape level measurements are obtained by Multispectral Scanners (MSS) aboard satellites. When such Space Age data is available, it would certainly be a shame to apply Stone Age analysis for drawing inference. On the same token, applying space age analysis to Stone Age data could be equally in vain.

The goal of environmental researchers should be to maximize the mining of information from the ore of data by matching space age analysis with Space Age data, at least to the extent required by Data Quality Objectives. In this direction, research should continue to merge statistical theory with computing technology, such as for innovative spatial analysis via geographic information systems (GIS) and the incorporation of probabilistic uncertainty with expert systems.

2.6. Cycle of No Information, New Information, and Non Information

Surveys for monitoring changes and trends in our environment and its resources involve some unusual conceptual and methodological issues pertaining to the observer, the observed and the observational process. Problems that are not typical of current theory and practice arise. Everyone concerned needs to find innovative ways and means of not contributing to, but breaking into, the burdensome and unaffordable cycle of no information, new information, and non information.

2.7. Mechanization/Computerization

The potential danger of model misspecification was brought out by J.G. Skellam, who said “Without enlightenment and eternal vigilance on the part of both ecologists and mathematicians there always lurks the danger that mathematical ecology might enter a dark age of barren formalism, fostered by an excessive faith in the magic of mathematics, blind acceptance of methodological dogma and worship of the new electronic gods.”

A similar message is eloquently carried forward by J. C. Bailar III in a fairly recent exposition on environmental statistics where he said “What is needed is not cookbook understanding; not technique, but scientific wisdom; not increased access to computer programs, but more role models in statistical thinking."

2.8. Normality, Lognormality and Beyond Lognormality

As scientific inquiry ventures into environmental systems, it soon becomes obvious that non-traditional statistical methods often are needed. While most environmental and ecological measurements are lognormally distributed, being more skewed towards high values than a normal (bell-shaped) distribution, chemical concentrations at a hazardous waste site are typically skewed even more extremely. Furthermore, high values are also often clustered in spatial proximity.

The concepts of simplicity, efficiency, and economy within the context of science, technology, and society are becoming critical to realize the achievable and available mandates and guidelines with the statistical, computational, and logistical technologies around. The age of means, medians, modes, quantiles, and relationships continues, but with emphasis on maps, contours, and improved geospatial-temporal visuals wherever applicable.

2.9. Triad

A traditional approach to environmental monitoring has been pairwise interaction among the research scientist, statistical scientist and the resource manager, while interaction between the resource manager and the statistical scientist has been minimal. Many of us have witnessed the limitation of this approach for the emergence of useful information. We feel that a triad approach of simultaneous working interaction among the three parties is the way for useful information to emerge in the days ahead. Just like a three-legged stool, full functionality depends on support of all three legs, otherwise the stool collapses. Maintaining the triad is a primary thrust of the "Total Quality Management" concept.

2.10. Follow-up

The need for environmental remediation and protection is often lost in the shambles of adversarial proceedings. Science can become part of the problem, and often enough the "patsy". As quoted from Hennemuth and Patil, "It seems far easier to use science to obfuscate rather than to clarify." However, one should also remember a perceptive comment by Frederick Mosteller, who said "While it is easy to lie with figures, it is easier to lie without them!". Sound environmental science should therefore be strongly defended for the sake of public decision making. We should welcome improvement and innovation that "break" the conventional see-saw type balance between uncertainty and cost. What is now "Best Possible Statistical Technology" may eventually be "Best Available Statistical Technology".

3. Ecological Sampling and Statistical Inference

3.1. Encounter Sampling

Surveys for monitoring changes and trends in our environment and its resources involve some unusual conceptual and methodological issues pertaining to the observer, the observed, and the observational process. Problems that are not typical of current statistical theory and practice arise.

Traditional statistical theory and practice have been occupied largely with statistics involving randomization and replication. But in ecological and environmental work, observations most often fall in the non-experimental, non-replicated, and non-random categories. Additionally, the problems of model specification and data interpretation acquire special importance and great concern. In statistical ecology and environmental statistics, the theory of weighted distributions provides a perceptive and unifying approach for the problems of model specification and data interpretation within the context of encounter sampling.

Weighted distributions take into account the observer-observed interface, i.e. the method of ascertainment, by adjusting the probabilities of actual occurrence of events to arrive at a specification of the probabilities of those events as observed and recorded. Appropriate statistical modeling approaches help accomplish unbiased inference in spite of the biased data and, at times, even provide a more informative and economic setup.

3.2. Adaptive Sampling

Several ecological and environmental populations are spatially distributed in a clumped manner.

They are not very efficiently sampled by conventional probability based sampling designs.

Adaptive sampling is therefore introduced as a multistage design in which only the initial sample is obtained using a conventional probability based procedure. When the variable of interest for a sampling unit satisfies a given criterion, however, additional units in the neighborhood are selected in the next sampling stage. This procedure is repeated until no new units satisfy the criterion, or the conditions of a stopping rule are satisfied.

With the recent growth of geographic information systems (GIS), spatial data for landscapes are becoming universal. This information provides a powerful aid to adaptive sampling and needs to be exploited.

-
-
-

Bibliography

Gore, S. D. and Patil, G. P. (1994). Identifying extremely large values using composite sample data. *Environmental and Ecological Statistics* **1**, 227-245. [This paper provides a breakthrough for more use of composite sampling in practice, because the methodology now makes it possible to recover the extreme values otherwise composited].

Gove, J. H., Patil, G. P., and Taillie, C. (1994). A mathematical programming model for maintaining structural diversity in uneven-aged forest stands with implications to other formulations. *Ecological Modelling* **79**, 11-19. [This paper provides a nicely worked out example giving explicit comparisons of diversity profiles of different ecological communities].

Hennemuth, R. C., and Patil, G. P. (1983). Implementing statistical ecology initiatives to cope with global resource impacts. In *Proceedings of the Conference on Renewable Resource Inventories for Monitoring Changes and Trends*, (eds. J. Bell and T. Atterbury), 374-378. Corvallis, USA: Oregon State University. [An important paper making a strong pitch in the initial formative stages of statistical ecology within the context of global resource impacts].

Johnson, G., and Patil, G. P. (2004). *Landscape Pattern Analysis for Assessing Ecosystem Condition.* Boston, MA, USA: Kluwer Academic Publishers. (To appear). [A contemporary monograph providing lucid exposition of frontier methods and tools for landscape pattern analysis for assessing ecosystem condition].

Myers, W. L. and Patil, G. P. (2004). *Doubly Segmented Images and Landscape Indicators for GIS Analysis: With Emphasis on Investigation of Landscape Change*. Boston, MA, USA: Kluwer Academic Publishers. (To appear). [A timely monograph full of novel and innovative concepts and methods of doubly segmented images and landscape indicators for GIS analysis with emphasis on investigations of landscape change].

Patil, G. P. (1991). Encountered data, statistical ecology, environmental statistics, and weighted distribution methods. *Environmetrics* **2**, 377-423. [One of the basic and fundamental papers capturing the spirit, essence, and tool for encounter sampling in ecological work].

Patil, G. P. (1995). Statistical ecology and related ecological statistics-25 years. *Environmental and Ecological Statistics* **2**, 81-89. [This article provides a perceptive historical sketch of the initial quarter century of the foundational period of statistical ecology].

Patil, G. P. and Taillie, C. (1979a). An overview of diversity. In *Ecological Diversity in Theory and Practice* **5**, (eds. J. F. Grassle, G. P. Patil, W. K. Smith and C. Taillie), 3-27. Fairland, Maryland, USA: International Co-operative Publishing House. [This provides a fundamental paper for diversity measurement and comparison as a concept and method].

Patil, G. P., and Taillie, C. (1979b). A study of diversity profiles and orderings for a bird community in the vicinity of Colstrip, Montana. In *Contemporary Quantitative Ecology and Related Ecometrics*, (eds. Patil, G. P. and Rosenzweig, M.), 23-48. Burtonsville, MD, USA: International Co-operative Publishing House. [This provides a first fully worked out paper for diversity profiles and orderings].

Patil, G. P., and Taillie, C.(2001). A multiscale hierarchical Markov transition matrix model for generating and analyzing thematic raster maps. Technical Report 2000-0603. University Park, PA., USA: Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University. Also *Environmental and Ecological Statistics*, 8(1), 71—84. [This provides a foundational paper for multiscale advanced raster map analysis system].

Biographical Sketch

For everyone **Professor Patil** is GP. He is Distinguished Professor of Mathematical and Environmental Statistics in the Department of Statistics at the Pennsylvania State University, and is a former Visiting Professor of Biostatistics at Harvard University in the Harvard School of Public Health.

He has a Ph.D. in Mathematics, D.Sc. in Statistics, one Honorary Degree in Biological Sciences, and another in Letters. GP is a Fellow of American Statistical Association, Fellow of American Association of Advancement of Science, Fellow of Institute of Mathematical Statistics, Elected Member of the International Statistical Institute, Founder Fellow of the National Institute of Ecology and the Society for Medical Statistics in India.

GP has been a founder of Statistical Ecology Section of International Association for Ecology and Ecological Society of America, a founder of Statistics and Environment Section of American Statistical Association, and a founder of the International Society for Risk Analysis. He is founding editor-in-chief of the international journal, *Environmental and Ecological Statistics* and founding director of the Penn State Center for Statistical Ecology and Environmental Statistics. He has published thirty volumes and three hundred research papers. GP has received several distinguished awards which include: Distinguished Statistical Ecologist Award of the International Association for Ecology, Distinguished Achievement Medal for Statistics and the Environment of the American Statistical Association, Distinguished Twentieth Century Service Award for Statistical Ecology and Environmental Statistics of the Ninth Lukacs Symposium, Best Paper Award of the American Fisheries Society, and lately, the Best Paper Award of the American Water Resources Association, among others.

Currently, GP is principal investigator of a multi-year NSF grant for surveillance geoinformatics for hotspot detection and prioritization across geographic regions and networks for digital government in the 21st Century. The project has a dual disciplinary and cross-disciplinary thrust. You are invited.