ECOLOGIA

ATTI DEL PRIMO CONGRESSO NAZIONALE DELLA SOCIETÀ ITALIANA DI ECOLOGIA
Salsomaggiore Terme (Parma): 21-24 ottobre 1980

ESTRATTO
Some perspectives in statistical ecology

G.P. Patil
Department of Statistics, Pennsylvania State University,
University Park, PA. 16802 USA

Based on an invited Inaugural Address on Quantitative Ecology at the first Congress of the Italian Ecological Society held in October 1980 and on a Presidential Invited Address at the Spring Meetings of the Biometric Society (ENAR) and the American Statistical Association held in March 1980.

1. Introduction

At the invitation of Professors Moroni and Rossi, a six-week research and training institute on statistical ecology was held at the University of Parma in 1978. Over one hundred participants from within Italy and abroad participated in one capacity or the other. It was a very successful program. It is a great pleasure and honor for me to be back, this time to be with you at the First Congress of Ecology of the Italian Ecological Society. My best wishes to you and to the Society in your timely efforts to strengthen research and training activities in quantitative ecology. As you are aware, quantitative ecology involves a study of mathematical, statistical, and computational aspects of ecology both in theory and in practice. Today I would like to share with you some of my perspectives of statistical ecology - a subject area that has acquired importance in view of the current ecological issues. We shall attempt to envision statistical ecology in a variety of perspectives, such as, historical, organizational, scientific, research training, and finally research itself.

Russel Train (1973) comes to mind when he said: «For top management and general public policy development, monitoring data must be shaped into easy-to-understand indices that aggregate data into understandable forms. I am convinced that much greater effort must be placed on the development of better monitoring systems and indices than we have in the past. Failure to do so will result in suboptimum achievement of goals at much greater expense».

Also, John Skellam (1972) comes to mind when he said: «Mathematical ecology is moving out of its classical phase carrying with it untold promise for the future, but, as H. A. L. Fisher, the historian, remarks, progress is not a law of nature. 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 workshop for the new electronic gods. It is up to all of us to ensure that this does not happen».

John Skellam was a friend, philosopher, and guide for statistical ecology. Russel Train made his comments when he was the chief of the Environmental Protection Agency of the United States of America. Let us hope and trust that it would be possible for statistical ecology to chart out its progress in Italy with the words of Skellam and Train in constant view.

2. Historical perspectives

Statistical ecology, as we know, is now in its teens! It was born of a proposal I made to the Ford Foundation in support of the First International Symposium on what I perceived to be Statistical Ecology. It is now interesting to know that a paper had appeared as early as in 1907 with «statistical ecology» in its title. The term statistical ecology was used for the collection and description of ecological data. The author of the article was Stephen A. Forbes who later became known as the Father of American Limnology! No subsequent reference appears in print until 1952 when we again see «statistical ecology» in the title of a paper.

I am indebted to Dr. Robert Keen for bringing this reference to my attention.
Here the term of statistical ecology was used as a research theme. The author was none other than John Skellam. Next in 1971 came three volumes on statistical ecology based on the First International Symposium on Statistical Ecology. Since then any study of ecology that involves statistical aspects is routinely considered a part of statistical ecology.

Several institutions have courses, programs, and groups to cover statistical ecology. A series of publications on statistical ecology is also in print of which thirteen volumes have appeared. (See references at the end.) Today the subject area of statistical ecology seems to have acquired a well deserved focus and identity. It is sure to have challenging opportunities to adequately serve the needs of science, technology, and society in the days ahead.

3. Organizational perspectives

The First International Symposium on Statistical Ecology was held in 1969 at Yale University with support from the Ford Foundation and the US Forest Service. The three symposium co-chairmen (G. P. Patil, E. C. Pielou, and W. E. Waters) represented the fields of statistics, theoretical ecology, and applied ecology. The program was well attended, and it provided a broad picture of where statistics and ecology stood relative to each other. While effort was apparent, communication between the two disciplines was inadequate.

Clearly, a focal forum was necessary to discuss and develop a constructive interface between quantifiable problems in ecology and relevant quantitative methods. As a partial solution to meet this need at a professional organizational level, the director of the symposium (G. P. Patil) made certain recommendations to the Presidents of the International Association for Ecology (A. D. Hasler), the International Statistical Institute (W. G. Cochran), and the International Biometric Society (P. Armitage). The International Association for Ecology (INTECOL) took a timely step in creating a new section in the organization, the statistical ecology section. The three societies together set up a liaison committee on statistical ecology. The INTECOL Section and the Liaison Committee together developed the International Statistical Ecology Program. Since its inception in 1970, INTECOL (as it has come to be known) has put emphasis on identifying the interdisciplinary needs of statistics and ecology at advanced instructional levels, and also at research conference and workshop levels.

The First Advanced Institute on Statistical Ecology in the United States was organized at the Pennsylvania State University for six weeks in 1972 with support from the US National Science Foundation, the US Forest Service, and the Mathematical Social Sciences Board. The participants of the institute have all enjoyed the benefits of their fruitful participation. With support from the UNESCO program of Man and Biosphere, a six month program was held in Venezuela for participants from Latin America in 1974 under the direction of Jorge Rabinovich, also a participant in the 1972 Institute. With some initiatives from ISEP, special statistical ecology sessions have been held at the international conferences of the International Statistical Institute and the Biometric Society.

While plans were being made for the Second International Congress of Ecology, the then Secretary General and current President of INTECOL (G. A. Knox), and the ISEP chairman (G. P. Patil) discussed the need for and the timeliness of a program in statistical ecology. From this emerged the Satellite Program in Statistical Ecology.

Satellite A at College Station, Texas, and at Berkeley, California covered a four week period in 1977 and had 125 participants. Satellite B at Parma, Italy had 130 participants spread over a six week period in 1978. Satellite C at Jerusalem, Israel had 35 participants. Approximately, one-third of the participants were graduate students, one-half were university faculty, and one-third were agency scientists. Approximately, one-third of the participants had an affiliation with one mathematical science or another, one-half an affiliation with one environmental science or the other, and one-quarter had an affiliation with one environmental management program or another. Thus the group was a good mixture of great variety contributing to the effectiveness of the program. Not only was what one heard enlightening, but, what one over-heard, was also equally enlightening!

We wish there was no further need for a program of this nature and dimension. It would be ideal if the needs of an interdisciplinary program were satisfactorily met in the existing institutions. Unfortunately, universities and governmental agencies have not been able to find effective ways to foster healthy interdisciplinary programs. The individuals attempting to do something in this direction tend to feel disheartened or disillusioned.

The satellite-like programs help create and sustain the enthusiasm, inner strength, and working efficiency in those who desire to meet a contemporary social need in the form of interdisciplinary work. It would only be proper and rewarding for everyone involved, that such programs be planned from time to time.

Plans are being made for sessions on statistical ecology at the Third International Congress of Ecology to be held in Warsaw, Poland, in 1982.

The program includes topics of current interest and concern, such as community ecology and quantitative approach, dynamics of exploited populations, compartmental models analysis, sampling designs for environmental impact studies, quantitative natural resource management, statistical ecology software, and statistical ecology bibliography. I hope and trust that many of you will be able to participate in this forthcoming program.

4. Scientific perspectives

The perceptions of Skellam and Train described ear-
lier encapsulate the cautions, inspirations, and objectives responsible for a fruitful scientific development. The rigorous formulation of a quantitative scientific concept requires and in a sense creates empirically measurable quantities. Conversely, the scientific validity of the concept is totally dependent upon the measured values of those quantities. This is a capsule version of the feedback process known as the "scientific method". In crude modern terms, we might label the first procedure "modeling" and the second "curve-fitting". For reasons of complexity, historical accident, or whatever, these mutually dependent, complementary components have never become firmly integrated within ecology and its application to environmental studies.

Both procedures involve forms of mathematics: In the "modeling" process, the mathematics is used relationally — as a system of logic to ensure rigor and clarity of reasoning. This is in the historical tradition of Volterra and Lotka. Validation is most often based upon qualitative agreement: the right trend, or the correct shape of the curve. This is as it should be, especially for broad general theories: it is the ideas and understanding that count, not so much the quantitative detail.

In the "curve-fitting" tradition, the mathematics is used numerically — as a system for precise measurement and prediction. Validation is most often based on quantitative agreement: n-digit accuracy, or minimum uncertainty. This is also as it should be, especially for application and management: it is the forecast and ability to predict confidently that count, not so much the underlying concept.

It need not be taken as a sign of "physics envy" to assert that as a science matures, these two processes must converge — more quantification must be used in concept validation and more concepts must be incorporated into the methodology of quantification.

The purpose of a worthwhile statistical ecology activity is to encourage that convergence within the science of ecology an promote its application in the study of the environment and environmental stress. To be meaningful and defensible, monitoring and assessment activities need: (i) a conceptual and philosophical basis, (ii) a theoretical framework, (iii) methodological support, (iv) a technological toolbox, and (v) administrative management. The ultimate purpose is to help identify and integrate the specifics to these important factors responsible for protecting the environment.

We need to take as our theme the better melding of fundamental ecological concepts with rigorous empirical quantification. The overall result should be progress toward a stronger body of general ecological theory and practice.

5. Research training perspectives

While effort within a single discipline is not necessari-

---

* I am indebted to Dr. Frederick Williams for working with me closely on this section.

---

(ii) How many of them are out there? This scenario takes place in a court of law. The issue is about the abundance of a species seemingly endangered, threatened, or rare. The judge orders an investigation. A seasoned investigator conducts the survey. He reports having seen 375 individual members of the species under consideration. The judge invites comments.

**Industrial Lobby**: The reported record of 375 members makes sense. The visibility factor is low in such surveys. The investigator has surely missed most of them that are out there. The exploitation should not cause alarm.

**Environmental Lobby**: The reported record of 375 makes sense. The investigator is an expert in such surveys. He has observed and recorded most of them that are there. And, therefore, only a few are out there. The species population needs to be protected.

The scenario is a typical one. It brings home the is-
issues characteristic of field observations often lacking a sampling frame necessary for the classical sampling theory to apply. One needs to work with visibility analysis instead. Satisfactory estimation of biological population abundance depends largely, in such cases, on accurate measurement of visibility, variously termed catchability, audibility, etc. And, this is not a trivial problem!

(iii) The blind men and the elephant: We are all familiar with the dialogue and the debate around the issues involving sampling, modeling, and analysis whether in theory, practice, or in the interaction of the two. The following extract of a poem on the blind men and the elephant brings out the dilemmas and the difficulties with great eloquence and clarity. It is a story to recall and to ponder whenever one is engaged in any extrapolation in the face of uncertainty, and in its communication.

The Blind Men and the Elephant
By J. G. Saxe (1816-1887)

1. It was six men of Indostan
   To learning much inclined,
   Who went to see the Elephant
   (Though all of them were blind),
   That each by observation
   Might satisfy his mind.

2. The First approached the Elephant
   And happening to fall
   Against his broad and sturdy side
   At once began to bawl:
   «God bless! but the Elephant
   Is very like a wall!»
   ...
   ...

3. The Third approached the animal,
   And happening to take
   The squirming trunk within his hands,
   Thus boldly up and spake:
   «I see, quot he, «the Elephant
   Is very like a Snake!»

4. The Fourth reached out an eager hand
   And felt about the knee.
   «What most this wondrous beast is like
   Is mighty plain», quot he;
   «Tis clear enough the Elephant
   Is very like a tree!»
   ...
   ...

5. The Sixth no sooner had begun
   About the beast to grope,
   Than, seizing on the swinging tail
   That fell within his scope,
   «I see», quot he, «the Elephant
   Is very like a rope!»

6. And so these men of Indostan
   Disputed loud and long,
   Each in his own opinion
   Exceeding stiff and strong,
   Though each was partly in the right
   And all were in the wrong!

(iv) Improbable and impossible: An important key for satisfactory stochastic modeling lies in effecting a judicious equivalence between improbable and impossible events. In the classical example of a coin tossing problem, the binomial model turns the improbable event of standing on the edge into an impossible event by virtue of admitting only head and tail as the potential outcomes. However, in modeling the life-length of a newborn, the popular models for life-length have infinite range, turning impossible events of excessive life-length into improbable ones. While the nature and extent of matching or fine-tuning between the model and the reality depend on the problem at hand, the judicious equivalence between the improbable and the impossible provides an interesting way of looking at a stochastic model. The following musical note should strike well in this connection.

   A musical note from an opera: H. M. S. Pinafore by
   W. S. Gilbert*
   All: What, never?
   Captain: No, never!
   All: What, never?
   Captain: Hardly ever!

   * I am indebted to Dr. Charles Taillie for bringing this reference to my attention.

(v) Problem formulation: Every interdisciplinary effort needs to involve continued problem identification and formulation — perhaps much more than the one involved in a single discipline effort. Edward David, former Presidential Science Advisor for USA quotes General David Sarnoff from his speech to the listening RCA researchers. This important message has great relevance and significance for statistical ecology also. The quote is: «Fifty percent of invention is knowing what to invent». See Science, February 1980.

(vi) Martian philosophy: A student wishes to study 'Martian Philosophy' but finds that there is no instructional program available in Martian philosophy. He is advised to take courses in astronomy, which may have some bearing upon Mars; he is also asked to take courses in philosophy that may have some context of the universe; and in due course, he is declared to have completed a program in Martian Philosophy! The inadequacy of this approach is clear. It would be important to make sure that neither the student nor the supervisor falls into this trap. Integrated and interactive research training programs should be made available to those interested and concerned.
(vii) Information paradox: Generally speaking, we have infinite information around us; but often we do not quite know what to make of it. At the same time, however, for a specific problem, we may have a mere infinitesimal amount of information. Often, we do not quite know what to make of it either. We find ourselves in need of additional information, even when we know of the infinite information around us.

Is there a satisfactory resolution of the paradox? (viii) Comprehensive versus comprehensible: Consider the following. We wish to comprehend a given situation.

— For lack of information, we do not (quite) comprehend the situation.
— We collect information, and we tend to collect comprehensive information.
— Because the information is comprehensive, we do not (quite) comprehend it.
— So we summarize the information through a set of indices (statistics) so that it would be comprehensible.
— But now, we do not comprehend quite what the indices exactly mean.
— And, therefore, we do not (quite) comprehend the situation.
— 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 state clearly the purpose, formulation, and solution for the study under consideration.

(ix) Scientific content and statistical extraction: The following quotation of Fisher should be quite perceptive and enlightening.

«The statistician is no longer an alchemist expected to produce gold from any worthless material offered him. He is more like a chemist capable of assaying exactly how much of value it contains, and capable also of extracting this amount, and no more. In these circumstances, it would be foolish to commend a statistician because his results are precise or to reproach because they are not. If he is competent in his craft, the value of the result follows solely from the value of the material given him. It contains so much information and no more. His job is only to produce what it contains».


(x) Statistical research training: In his Presidential address to the International Statistical Institute held in Manila, Philippines in December 1979, Rao made following comments that should be very instructive and useful.

«...It is generally agreed and pious resolutions have been passed by several committees to the effect that all statistical courses must have practical orientation. But very few attempts have been made to evolve integrated courses with the desirable blend of theory and applications. While the discussion of practical problems in teaching statistics is essential, it must not be done at the expense of weakening the theoretical and mathematical contents of the courses. No doubt, a statistician should be equipped with the practical skills to enable him to choose the appropriate methodology for analysis of given data; at the same time, he should have the theoretical knowledge to formulate a new statistical problem for study if the existing methodology is inadequate in any given practical situation... The gap between theoretical statisticians and practitioners may be partly bridged if we are able to produce statisticians with the proper knowledge of theory and flair for applications and give them opportunities to improve their skills during the course so their work...»

6. A perspective of a research problem

To illustrate a perspective on research in statistical ecology, I would like to share with you some of the recent advances in the theory and application of ecological diversity. What appears to be true of the present and the future of this facet of community structure is by and large true of most any other quantifiable concept in ecology. A four hundred page volume has just appeared on ecological diversity in theory and practice, edited by Grassle et al. (1979). This volume has in it also a bibliography of over one thousand publications devoted to the subject of ecological diversity in one context or the other. Understandably, I will try to be brief. First, two quotes:

«...Diversity so pervades every aspect of biology that each author may safely interpret the word as he wishes and there is consequently no central theme to the subject. We cannot be sure if this flexibility is healthy or due to lack of discipline, but it can be traced back to the beginnings of interest in biological diversity...»

L. R. Taylor (1978)
Inaugural Address to the Symposium on the Diversity of Insect Fauna.

«...The richness and variety — in a word, the diversi- ty — of natural ecological communities have never been more highly valued than they are now, as they become increasingly threatened by the environmental crisis. Students of what has come to be known as "ecological diversity" realize that their work now has practical importance (indeed, urgency) in addition to the academic interest it has always had...»

E. C. Pielou (1975)
Preface to Ecological Diversity

To set the stage, consider the following question: Am
I a specialist or generalist?... My college dean may say that I am a specialist... because I do statistics; not physics, not chemistry, not astronomy, etc.; whereas, my department head may say that I am a generalist... because I do statistical ecology, statistical distributions, statistical characterizations, statistical programming (instituted!), etc. Clearly, the degree of specialization/diversification has to be related to the categories identified.

Let us now assume the categories of interest given. And, we consider the apportionment of a resource, such as abundance, biomass, energy or time. Suppose John and Jane apportion their time effort on mathematics and music as follows:

<table>
<thead>
<tr>
<th>Person</th>
<th>Category</th>
<th>Mathematics</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Jane</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

If we ask, «Does John have a different kind of specialization/diversification than Jane?», the answer is yes; the subject identity matters. Instead, if we ask, «Does John have a different degree of specialization/diversification than Jane?», the answer is no; the subject identity does not matter. The degree of specialization/diversification is permutation invariant. You may have observed that we are already in the realm of ecological diversity.

For conceptualization purposes, consider a community of s species with relative abundance vector \( \pi = (\pi_1, \pi_2, ..., \pi_s) \) with labels arranged so that \( \pi_1 \geq \pi_2 \geq ... \geq \pi_s \).

Let us consider the diversity of the community to be an average property of the community. But average of what? Since diversity has to do with variety, and variety has to do with rarity, let us speak of the average of the species rarity. Thus, if \( R(i; \pi) \) stands for the rarity of the \( i \)-th species in a community of relative abundance vector \( \pi \), we may define the corresponding diversity of the community \( \pi \) to be

\[
\Delta(\pi) = \sum_{i=1}^{s} \pi_i R(i; \pi).
\]

This definition has an alternative interpretation in terms of an average over the individuals of the community also. For example, draw an individual from the community at random, and record the rarity \( R(i; \pi) \) of the species \( i \) to which the individual belongs. The expected rarity is nothing but \( \Delta(\pi) \).

How to choose the rarity function \( R(i; \pi) \)? We note that rarity is a relative concept. If we consider species \( j \) to be a standard, depending on our need to be sensitive to abundant or rare species, we may define

\[
R(i; \pi) = \begin{cases} 
1 & \text{if } i > j \\
0 & \text{if } i \leq j 
\end{cases}
\]

Corresponding to this zero-one rarity function, we obtain

\[
\Delta(\pi) = \sum_{i>j} \pi_i = T_{ij},
\]

the right tail of the discrete probability distribution \( (\pi_1, \pi_2, ..., \pi_s) \) representing the community \( \pi \).

But how to fix \( j \)? One need not. One may consider every \( j, 0 < j \leq s \), and look at the entire diversity profile

\[
\{(j, T_{ij}) : 0 < j \leq s\}
\]

instead of a single diversity number \( T_j \) for a fixed \( j \). In so doing, we are representing the community by the diversity numbers \( T_0, T_1, T_2, ..., T_s \), where \( T_0 = 1 \), and \( T_s = 0 \) rather than by the relative abundance numbers \( \pi_1, \pi_2, ..., \pi_s \), where \( \pi_1 + \pi_2 + ... + \pi_s = 1 \). Thus, there is no loss of information, unlike the case when one wants to speak of a single diversity measure for the community.

We may represent the \( T_j \) diversity profile by the curve \( (j, T_{ij}) \) as follows.

For illustrative purpose, let

\[
\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5) = (\frac{8}{20}, \frac{5}{20}, \frac{4}{20}, \frac{3}{20}, \frac{1}{20}),
\]

so that

\[
T_0 = 1, \quad T_1 = \frac{12}{20}, \quad T_2 = \frac{7}{20}, \quad T_3 = \frac{3}{20}, \quad T_4 = \frac{1}{20}, \quad \text{and } T_5 = 0.
\]

The simple looking \( T_j \) profile has much deeper significance. It effects an intrinsic diversity ordering between two communities \( \pi = (\pi_1, \pi_2, ..., \pi_s) \) and \( \nu = (\nu_1, \nu_2, ..., \nu_s) \). That is, \( \nu \) is intrinsically more diverse than \( \pi \) if and only if the \( T_j \) profile of the community \( \nu \) is never below the \( T_j \) profile of the community \( \pi \). Conceptually, \( \nu \) is defined to be intrinsically more diverse than \( \pi \) if \( \nu \) is «realized» from \( \pi \) through a series of steps that involve transfer of abundance from more abundant to less abundant spe-
cies or involve introducing a new species. It follows that the $T_i$ profiles of $\pi$ and $\psi$ will cross if neither community is intrinsically more diverse than the other, implying inconsistency in the comparisons based on individual diversity numbers. See Patil and Tallie (1979a) for an overview. You may also find Patil and Tallie (1979b) of further interest in this connection. It discusses diversity profiles and orderings for a bird community in the vicinity of Colstrip, Montana. Finally, you may be interested in knowing that these developments have helped discover in Preston's canonical hypothesis an implied tradeoff between richness and evenness leading to a natural composi-
tucceptualization of diversity as has been prevalent in ecological literature. See Patil and Tallie (1979c) for details.

In matters relating to the canonical hypothesis and its empirical basis, Preston (1962, 1979, and 1980) provide ecological formulation and insight. Considerable interest in visible in theories of island biogeography and in their application to the design of nature reserves (see May 1975a, b) needing reasonable species area relations. The canonical hypothesis and the related diversity considerations provide on such relation. See Usher (1979), and Higgs and Usher (1980) for a recent discussion.

7. Conclusion

In conclusion, I am reminded of a famous poet speaking on life pursuit! A budding enthusiast, in his twenties, went to Rabindranath Tagore for advice on what to do in life. The Nobel Laureate replied: «Anything my dear, but preferably not on an empty stomach». Lunch time is getting close. So, I will bring my talk to a close. Thank you very much for your attention and interest, and best wishes on your plants and programs in the times ahead.

Acknowledgments

My thanks are due to Brian Dennis, Keith Ord, James Rosenberger, Charles Tallie, and Frederik Williams for their comments and suggestions during the preparation of this paper.

References


