Numerical Ecology

L. Legendre
and
P. Legendre

Developments in
Environmental
Modelling 3

Developments in Environmental Modelling, 3

Numerical Ecology

Louis LEGENDRE, Professor

Department of Biology, Université Laval, Québec, Québec G1K 7P4, Canada

and

Pierre LEGENDRE, Associate Professor

Department of Biological Sciences, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Preface by Prof. F. James Rohlf

Expanded translation from the French.
Translated by Robert Bélanger (chap. 11), Wendy E. Biro (chap. 8), Robert Catalfamo (chaps. 3 and 4), Luc Alain Giraldeau (chap. 7), Marc Labelle (chaps. 6 and 9), Stefan Ochman (chaps. 5 and 10), Gilles Roussin (Preface to the French Edition and Foreword), Vahé Sarafian (chap. 2), and Peter P. Zwack (chap. 1). Translation revised by Elizabeth A. Diamond and Marcia S. Downing.
Translation under the supervision of the authors, and subsequently expanded by them.
The translation and composition of this book were supported by grants from the Université du Québec à Montréal (former affiliation of the second author) and Université Laval.

ELSEVIER SCIENTIFIC PUBLISHING COMPANY
Amsterdam – Oxford – New York 1983

ELSEVIER SCIENTIFIC PUBLISHING COMPANY
Molenwerf 1
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY INC.
52, Vanderbilt Avenue
New York, N.Y. 10017

Library of Congress Cataloging in Publication Data

Legendre, Louis.
Numerical ecology.

(Developments in environmental modelling ; 3)
Rev. and expanded translation of: Ecologie numérique.
Bibliography: p.
Includes index.
1. Ecology--Mathematics. 2. Ecology--Statistical methods. I. Legendre, Pierre, 1946- . II. Title. III. Series.
QH541.15.M34L4313 1983 574.5'01'51 82-21115
ISBN 0-444-42157-2 (U.S.)

First published in French under the title *Ecologie numérique.* T. 1. *Le traitement multiple des données écologiques.* T. 2. *La structure des données écologiques.* Masson, Paris, and Les Presses de l'Université du Québec, 1979.

ISBN 0-444-42157-2 (Vol. 3)
ISBN 0-444-41948-9 (Series)

Printed in The Netherlands

PREFACE

I felt quite pleased and honored to be asked to write a preface to this new edition since it is likely to become a standard introduction and reference to a variety of important topics in mathematical and statistical ecology. I also appreciated the opportunity to preview the manuscript in advance of publication (and perhaps even to influence their discussions of topics which are of particular interest to me).

The book is intended to be a practical handbook. While it includes some very technical material, the authors begin each topic at a very elementary level (facility with advanced mathematical techniques is not a prerequisite for fruitful use of the book). They emphasize the biological interpretation of the various mathematical and statistical topics they cover, rather than worrying about the algebra and the mathematical proofs (they give, however, references to guide the more mathematical reader to the more advanced treatments of the subjects).

The authors start by giving considerable attention to the variables themselves. They discuss the advantages and disadvantages of different kinds of variables that an ecologist might consider using. I found their presentation on the application of dimensional analysis to ecology to be unique and quite interesting. This approach should prove to be a powerful new tool for understanding and checking the internal consistency of complex mathematical models and it will also be a valuable aid for the development of new ecological models. It should be especially useful for teachers and others who need to introduce and explain mathematical models to more general audiences.

I was especially pleased to find that they had devoted a considerable fraction of their manuscript to multivariate analysis. They recognize that nature is multidimensional and that one must, therefore, measure and then analyze numerous variables simultaneously if one is to begin to understand even the simplest ecological system. This is because environmental variables and organisms can interact in complex ways. It is difficult to interpret the more conventional studies which deal with only a single variable or at most a few variables at a time, since the investigators are (in effect) making the naive assumption that the neglected variables are either unimportant or that they have been held constant experimentally. Good introductions and examples are given for such topics as ordination analysis, factor analysis, cluster analysis, multiple regression analysis, and discriminant analysis.

It will be quite exciting to see what new understanding of ecological systems and processes will be gained in the next few years as more ecologists make use of these more powerful analytical tools.

F. JAMES ROHLF

Professor of Biological Sciences
State University of New York
Stony Brook, NY 11794 U.S.A.

PREFACE TO THE FRENCH EDITION

Influenced by a good dose of friendship and probably by their recollection of some of our rather critical discussions on the relationship between ecology and statistics as well, the authors of this book invited me to write its preface. I regarded this as an honour since it entailed both the privilege of reading their manuscript before its publication, and of expressing personal opinions on topics of my choice. It also represented an embarrassing assignment since the mathematical education that ecologists of my generation received was comparable to their sexual education: they acquired it haphazardly and on the street. For this reason, some facets of the subjects dealt with in this book may have escaped me. I do not, however, suffer from an excess of modesty; hence I feel that I have the ability to judge whether complex mathematical expressions are projections of natural situations or merely futile mental exercises.

My relative ignorance of mathematics leads me to affirm that this book is intelligible and didactic. Directing their efforts to the vigilant reader, the authors develop their subject with care and, in my view, quite logically. This book is written by and for ecologists, differing in this respect from the numerous books written by those statisticians who are merely trying to sell their goods, and who are satisfied with using and re-using imaginary examples, limiting their choice of problems to those with available answers. This book is both very scientific, and faithful to nature's complexity. It provides the reader with an overall and critical insight, allowing for a choice of appropriate solutions to each problem. Further, each solution is logically justified and does not involve any simplification or deformation of nature. Everything is done quite naturally and most of the examples given are authentic.

The book has limitations of course. It is mostly of a statistical nature, and there are areas where statistics and analysis become entangled. Moreover, the study of diversity, as well as matrix expressions of a population, readily come up against the constraints imposed by the fact that real-life ecosystems are structured in a physical space. The interesting problems that follow are not discussed in the book, but since this is normal in quantitative ecology, one cannot really criticize the authors in this respect. I had planned to adopt a judiciously critical approach and as I was reading, I kept taking notes and asking myself questions. I am happy to say that all my questions were answered in a context that offered a larger scope than that prevailing in my mind when I had conceived of the questions. The fact that the authors are conscious of the necessity of the models having a physical content pleases me immensely. Einstein is said to have wittily remarked that one of the most remarkable properties of mathematics is that they may be applied to physical sciences. The remark is even more applicable to biology. Living systems are primarily physical systems, not logical beings, and one makes substantial progress if one applies mathematics to biology through the filter of *physics. The authors are well aware of this and reflect this preoccupation in the chapter dealing with dimensional analysis, as well as in several other instances throughout the book.*

An ecologist must use mathematical models for self-analysis and therapy, in order to express his mental processes and make them understandable to others through a common logical base. This is particularly evident in the typology, ordination and classification of communities (a complex subject if there ever was one), which the authors have treated in a concise and well-balanced manner. I believe that they have been particularly successful here, considering the multitude of details which were involved and which may well startle the newcomer to the field. The latter will nevertheless be able to make a reasoned and responsible choice, given a proper amount of attention, since he is made aware of the traps and artifacts. It is rewarding to read a text written in the French language which does not place undue emphasis on only one of the analytical methods available.

This book is a very complete treatise, the contents of which may be advantageously compared to those of all other scientific works that I know of. It is a textbook that fulfills the ecologist's needs. The authors have judiciously and directly made their choices, ignoring in the process, without acrimony, obsolete material, and accepting new ideas, not so much because of their novelty but because they are essential to the logical nature of their work. They have done so without any fracas, and with courtesy to the quoted authors and readers, all of which we perceive and gratefully appreciate.

RAMÓN MARGALEF,

Professor of Ecology
University of Barcelona

FOREWORD

The delver into nature's aims
Seeks freedom and perfection;
Let calculation sift his claims
With faith and circumspection.

GOETHE

As a premise to this textbook on *Numerical Ecology*, the authors would like to state their opinion as to the function of numerical methods in ecology. In the above quotation, Goethe condemns the application of mathematics to natural sciences: in his opinion, mathematics only render more obscure, using an often esoteric language, the natural phenomena that the scientist is trying to elucidate. Unfortunately, many of the instances where mathematics have been used in ecology only serve to verify Goethe's thesis, particularly since the advent of computers facilitated access to even the most complex numerical treatments. Nevertheless, as is well illustrated throughout this book, when the ecologist masters the theoretical bases of numerical methods and knows how to use them, he can get a more thorough understanding of reality.

We do not consider the numerical approach a substitute for ecological reflection on natural observations, but rather as an objective and non-exclusive basis for an in-depth analysis of the scientist's data. Consequently, throughout the book, we insist on analyzing ecological applications which illustrate the pathways from the numerical to the ecological approach.

Our textbook is directed to practical ecologists — both graduate students and professional scientists — and we regard it as a practical handbook as well as a reference textbook. Our goal is to present the ecologist, in a clear and synthetic manner, with the mathematical techniques that have recently been developed for the analysis of ecological data complexes. A fair number of these numerical methods —derived from mathematical physics, communication theory, parametric and non-parametric statistics, taxometrics, sociometry, etc. — are available to only those ecologists who are specifically involved in numerical treatments, since the use of these methods entails numerous problems for the practical ecologist. Due to these difficulties, the analysis of ecological data is too often limited to elementary statistics, which are poorly adapted to the ecologist's multidimensional problems. This results in sub-optimal extraction of the available information.

With the advent of computer packages, scientists may now have access to the most sophisticated treatments, without having to write their own programs. Unless he has a strong background, the use of high-level packages may often place the ecologist in a dead-lock. It is thus essential that the principles underlying the various numerical methods and the extent to which the latter can be used be clearly established and that guidelines be set for the interpretation of the results generated by the computer. We have therefore organized this text, not only as a synthetic outline of the modern analytical methods for ecological data complexes, but also as a practical guidebook based on the most accessible packages.

Following our experience in graduate teaching and in consulting for major land management agencies, we are fully aware of the problems created for most ecologists by the use of advanced mathematical methods. Any earnest approach must be based on a deep understanding of the general principles and theoretical bases of the methods. We have therefore dealt with our subject in a manner adapted to ecologists, founded on unified symbolism, and making use of illustrations, and to some extent appealing to one's intuition. With a reasonable effort, the reader will get to the heart of numerical ecology. We have verified this by using our manuscript for our graduate teaching. On the other hand, one needs a knowledge of the finality and limits of numerical methods, and of the conditions under which they can be applied, in order to make efficient use of them. Most of these methods are already available in computer packages or can be found in the scientific literature; we therefore insist on the ecological interpretation of the results rather than on the computation algorithms, which can be found in reference materials. Finally, all the methods described in the book are illustrated by examples from the ecological literature, mostly in the English language (references in other languages than English or French are generally of historical nature).

Since the very expression *numerical ecology* may induce some confusion, we propose to give it the following meaning: *mathematical ecology*, which embraces the whole area of mathematical applications to ecology, may be divided into *theoretical ecology* and *quantitative ecology;* and the latter in turn embraces a plurality of areas among which may be mentioned *models* — descriptive, simulation or management models —*ecological statistics*, and *numerical ecology*. *Numerical ecology* is the field of quantitative ecology devoted to the numerical analysis of ecological data sets (related mainly to synecological problems), in order to determine and interpret their multidimensional and/or process structure. Numerical ecology is distinct from descriptive or inferential biological *statistics* in that it makes extensive use of nonstatistical procedures, and systematically combines ecologically relevant multidimensional statistics with nonstatistical numerical techniques (cluster analysis, etc.), frequently without any reference to theoretical distributions (i.e., statistical tests). Numerical ecology is also different from ecological *modelling*, even though extrapolation of ecological structures (through multiple regression or other procedures) may be used to *forecast* the extension of systems in space and/or time (sometimes called correlation modelling). When a *prediction* of the critical consequences of alternatives is sought, the ecologist must resort to ecological modelling. The development of models that predict the effects on some variables caused by changes in the others (see, for instance, de Neufville & Stafford, 1971) requires a deliberate causal structuring — based on ecological theories — and also a validation procedure. Such models are difficult to devise and are quite costly, so that good causal models are fairly rare. The ecological hypotheses underlying causal models are often developed in the context of numerical ecology studies, however, so that the two fields are closely related. Numerical ecology therefore provides a natural basis for ecological modelling (Jørgensen, 1980).

Ecologists have taken advantage of quantitative analysis since the publication, by Jaccard in 1901, of the first association coefficient. Floristics developed from this seed and the method was eventually applied to all fields of ecology, often achieving a great degree of development. Social scientists, on their own and following Hotelling (1933), have developed mathematical tools which were first derived from multidimensional statistics and later from non-parametric statistics. The advent of computers made it possible to tackle large data sets, then combining methods evolved from various approaches supplemented with new mathematical developments; the first synthesis was published by Sokal & Sneath (1963), establishing *numerical taxonomy* or *taxometrics* as a new discipline.

Numerical ecology is a step towards integrating the experience of numerous disciplines, into a general methodology for the analysis of data complexes. The chief charateristic of this method is the *convergent* and *unified* use of diversified treatments, evolved from distinct mathematical sectors. Thus, numerical ecology is founded on the recognition of the *complementary* character of the numerous existing numerical methods, each of which makes it possible to study one aspect of the information underlying the data, and it institutes principles with which to interpret results in an integrated fashion. This textbook is obviously far from being perfect, either from the standpoint of an inventory of the methods that may eventually be applied to ecological data, or from that of their integration. However, we have attempted to synthesize the information available in the literature, in a structured and prospective framework, so as to enable practical ecologists to take immediate and maximum advantage of the existing methods and to provide methodologists with an incentive for promoting advancement of the discipline.

This book is organized in such a way as to induce the scientist who is interested in a particular method to simultaneously examine alternative ones: numerous cross-references from one chapter to another as well as partial syntheses, sometimes making up the larger part of a chapter, favor an integrated approach to data analysis.

We could not have brought this work to completion without the assistance of many colleagues to whom we here wish to express our most sincere thanks. We first wish to acknowledge the outstanding collaboration of Profs. Serge Frontier (Université des Sciences et Techniques de Lille) and F. James Rohlf (State University of New York at Stony Brook), who critically reviewed the whole of our manuscript, and whose suggestions were very often incorporated into the final text. Many of the concepts underlying the chapters on dimensional analysis and spectral analysis had already been discussed by Dr. Trevor Platt, in 1974, in a series of lectures given to graduate students of Université Laval on the *thermodynamics of the ecosystem*. As a result of useful discussions with mathematicians Yves Escoufier, of Université des Sciences et Techniques du Languedoc (Montpellier), Michel Fortin, of Université Laval, Leonard P. Lefkovitch, of Agriculture Canada, and André Plante, of Université du Québec à Montréal, we have improved our original conception of ordination techniques. A fair number of the ideas expressed in the following pages were developed through regular and fruitful meetings with the scientists from the Station zoologique de Villefranche-sur-Mer, France. Finally, the preliminary drafts of our text were efficiently criticized

by graduate students of Université Laval and Université du Québec à Montréal.

Numerical ecology is a translation from the French *Ecologie numérique*. However, in preparing this translation many errors and inconsistencies found in the first edition were corrected. Additions have also been included in most chapters, in order to account for recent advances. In some cases (e.g., sections 4.2 and 4.3), sections have been completely rewritten, and tables and figures have been redesigned. *Numerical ecology* is therefore a revised and expanded translation of *Ecologie numérique*.

While writing this book, we benefited from competent and unselfish advice... which we did not always follow; we thus assume full responsibility for any gaps in this work as well as for all the opinions expressed therein. We shall therefore welcome with great interest any suggestions or criticisms that our readers may wish to address to us.

LOUIS LEGENDRE
PIERRE LEGENDRE

CONTENTS

CHAPTER 1

COMPLEX ECOLOGICAL DATA SETS

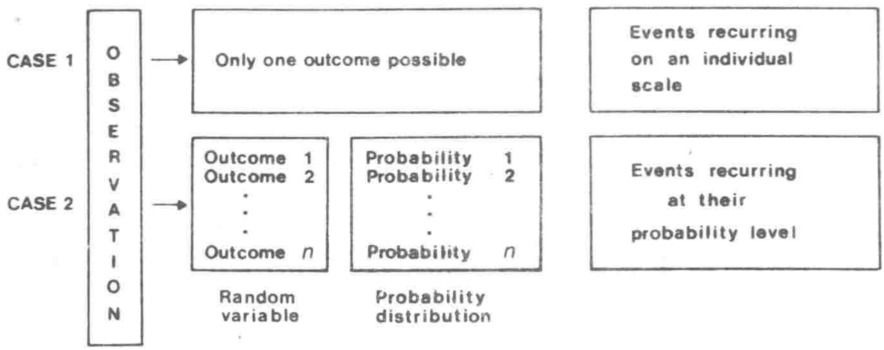
1.1 — NUMERICAL ANALYSIS OF ECOLOGICAL DATA

The basis of a general methodology for the analysis of ecological data can be derived by examining the relationships which exist between the *conditions* surrounding an ecological observation and its *outcome*. *In the physical sciences for example, a cause and effect relationship can often be established between the natural or experimental conditions* and the *outcome* of an observation or experiment. That is to say, given a certain set of conditions, it is possible to predict exactly the outcome that will follow. Such totally deterministic relationships are only characteristic of extremely simple ecological conditions.

Generally in ecology, a number of different outcomes can arise from a given set of conditions because of the large number of determining variables, many of which are not available to the scientist. However, if the same observation is repeated many times under similar conditions, the relative frequency of each possible outcome tends to stabilize at a given value called the *probability* of this outcome. Following Cramer (1946: 148) it is possible to state that "whenever we say that the probability of an event with respect to an experiment [or an observation] is equal to *P*, the concrete meaning of this assertion will thus simply be the following: in a long series of repetitions of the experiment [or observation], it is practically certain that the [relative] frequency of the event will be approximately equal to *P*." This corresponds to the frequency theory of probability and it excludes the Bayesian or likelihood approach.

In the previous case, the outcomes repeated themselves on an individual scale, while in this case they are repetitive only according to their probability. When each of the possible outcomes is characterized by a given probability, the set of all the possible events is called a *random variable* while the set of their respective probabilities is called their *probability distribution*. Table 1.1 summarizes these basic ideas. It is naturally possible to imagine other results from a series of observations, for example, that no relationship exists between the initial conditions and the resulting events, i.e., that the

Table 1.1 — Two types of recurrence of observations.



results are absolutely *unpredictable.* On the other hand, *strategic* relationships may be found between the surrounding conditions and the resulting events, when some action — or better, the expectation of this action —produces or modifies the reaction of others. It is possible that such strategic-type relationships can explain phenomena such as succession or evolution (Margalef, 1968), and that they might become of central interest to ecological research.

In table 1.2 these four types of relationships (between the conditions and the outcome of an observation) are regrouped along with methods for their processing. This text is limited to the numerical analysis of random data, which confront the ecologist most frequently.

Table 1.2 — Numerical analysis of data.

Relationships between natural conditions and the outcome of an observation	Methods for the use and analysis of the data
deterministic: only one possible result	deterministic models
random: many possible results, each one with a recurrent frequency	methods described below (figure 1.1)
strategic: results depend on the respective strategies of organisms and their environment	game theory
uncertain: many results possible, of an unpredictable nature	?

Numerical analysis of ecological data uses mathematical tools developed for use in many different disciplines. However, a formal exposition should use a unified method of presentation. For the ecologist, the most suitable and natural language — as will become evident in the next chapter — is that of *matrix algebra* which is derived from mathematical algebra. This method of calculation is best adapted to computer processing of data and has the advantage of being simple and efficient while, at the same time, familiar to a large number of ecologists.

Three distinct sciences provide the ecologist with powerful tools well adapted to the complexities of ecological data. From mathematical physics comes the powerful *dimensional analysis* that provides simple and elegant solutions to complex ecological problems. Communication theory contributes *information theory* which allows very general processing of any type of data. Finally, parametric statistics makes *multidimensional statistics* available to the ecologist.

Together, these four fundamental contributions can be used in the analysis of complex ecological data sets. Because such data usually come in the form of highly complicated sets of variables, the capabilities of elementary statistical processing are usually exceeded. Such elementary methods are the subject of a number of excellent texts, while this textbook expounds the more advanced methods upon which the ecologist must rely.

Ecological data sets are for the most part *multidimensional*: the ecologist samples along a number of axes which, depending on the case, are more or less independent, with the purpose of finding a structure and interpreting it. When the ecologist is studying a multiply faceted reality, which is termed *hyperspace* (space of many dimensions) in mathematics, the data are multidimensional. One now classic example of ecological hyperspace is the *fundamental niche* of Hutchinson (1957, 1965). According to Hutchinson, the environmental variables that determine the existence of a species can be thought of as axes, one for each factor, of a *multidimensional* space. On each axis there are limiting conditions within which the species can exist indefinitely. The set of these limiting conditions defines a *hypervolume* called the species' *fundamental niche*. The spatial axes, on the other hand, describe the geographical distribution of the species. The ecologist usually collects data for a large number of variables which are related to the problem studied and, using appropriate numerical analysis techniques, tries to find an organized and significant structure.

The success of the analysis and interpretation is related in particular to the compatibility between the data and the processing methods. It is, therefore, important to take into acount the requirements of the numerical techniques when planning the sampling, because it is obviously useless to collect quantitative data inappropriate to the numerical analysis. The analysis of multidimensional data is based on *association matrices* of which a number of types exist, each one leading to more or less different results. Two wide avenues of attack are then open to the ecologist: (1) ecological *clustering* by algorithms of agglomeration or division; and (2) *ordination* in a space of reduced dimensionality, using principal component or coordinate analysis, nonmetric multidimensional scaling, and correspondence analysis. The *analysis of structures,* derived from clustering and/or ordination, can be accomplished by a number of approaches depending on the nature of the problem and the information available.

In addition to multidimensional data, sets of ecological data are sometimes composed of *process data,* that is to say, data sampled along only one axis in the form of a *series*. Ecological series, both in time and space, require intensive sampling which can often be automated when equipment exists that permits recording of the ecological variables of interest. A *process* is a set of phenomena organized in time or space. Mathematically speaking, ecological data series represent one of the possible realizations of a random process which is called a *stochastic process.* Repeating the same sampling would not necessarily produce identical results, which leads to the idea of a process having a *random* component.

The form of the series must meet the requirements of the numerical method, because only certain special types of data series can actually be analyzed. Therefore, poor planning of the sampling can render the series of data useless for numerical treatment. There are a number of different methods for analyzing *ecological series;* some are designed for extracting general trends in the data while others are used to

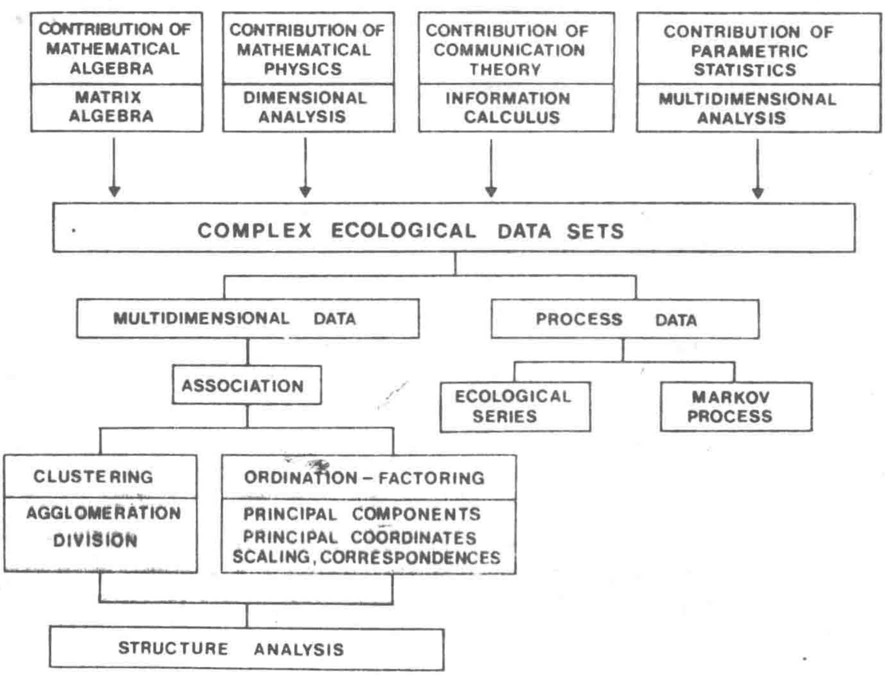


Figure 1.1 - Numerical analysis of complex ecological data sets.