Preface

In May 1975, a dozen or so ecologists, mostly marine, sat during three days in a (then) dusty conference room on the first floor of a historical building of the *Station marine de Villefranche-sur-Mer* (Université Paris 6, France), metres away from the Mediterranean shore, to discuss developments concerning a new trend in the ecological literature: the statistical analysis of multivariate ecological data. We, the authors of this book, had been independently invited to participate in the seminar. On the evening of the closing day of the meeting, sitting at the terrace of a restaurant, we wrote on a paper place mat a list of subjects that was to become the table of contents of the book that we published a few years later under the title *Écologie numérique* (first edition, in French; Legendre & Legendre, 1979a and b).

During the 1970's, community ecology, which had traditionally been a descriptive science until then, slowly adopted the hypothesis testing approach. Testing hypotheses required the analysis of numerical data. The theoretical foundations of community ecology had been developed during the 1950's and 1960's (niche theory, succession, biodiversity concepts, food webs, etc.) and statistically inclined researchers had already suggested to analyse ecological data using multivariate methods (e.g. Odum, 1950; Goodall, 1954; Bray & Curtis, 1957; Margalef, 1958; Williams & Lambert, 1959; Dagnelie, 1960, 1965; Gower, 1966; Pielou, 1966, 1969). We felt, in 1975, that the time was ripe to inventory the available numerical methods, compare them to the array of ecological questions found in the literature, describe the correspondences between questions and methods, provide a structure to interlink the various methods, and identify methodological gaps in the edifice. This is what we did in the first editions of this book, published in French in 1979, then in English under the title *Numerical ecology* (Legendre & Legendre, 1983a), quickly followed by a second French edition (Legendre & Legendre, 1984a and b).

Following the inventory and educational work described above, and with the help of graduate students and research assistants, we got to work to develop new numerical methods to answer emerging ecological questions and help fill gaps in existing numerical methodologies. Similar movements towards development of numerical methods took place in several laboratories throughout the world. In the late 1990's, the

time was ripe for a new synthesis of the field, and we worked on the second English edition of the ***Numerical ecology*** textbook (Legendre & Legendre, 1998). A decade later, the field of multivariate community ecology had developed so much that a new synthesis had become necessary. We spent most of the past three years preparing the 2012 edition of ***Numerical ecology***. This edition includes numerous developments in statistical computing made available in the R statistical language, and refers to many R packages written for ecologists by researchers in several laboratories around the world.

During our teaching in universities at home and abroad, we have been repeating a key message to graduate students: *While it is important to learn about the methods developed by previous generations of scientists, do not let yourself be silenced by their aura. If you think you have a good idea, work on it, develop it, listen to criticisms, and publish it, thus contributing to the advancement of the field. Do not let people tell you that everything is known, or that you are too young or not good enough to contribute to this field — or any other field of science.*

The ***Numerical ecology*** book is written for practising scientists — graduate students and professional researchers, in classical and molecular ecology, oceanography and limnology, environmental sciences, soil science, agriculture, environmental engineering, and related fields. For that reason, it is organized both as a practical handbook and a reference textbook. Our goal is to describe and discuss the numerical methods that are successfully used for analysing ecological data, using a clear and comprehensive approach. These methods are derived from the fields of mathematical physics, parametric and nonparametric statistics, information theory, numerical taxonomy, archaeology, geography, psychometrics, sociometry, econometrics, and others. Meaningful use of most of these methods requires that their theoretical bases be mastered by users. For that reason, analyses reported in the literature have at times been carried out with methods that were not fully adapted to the question or the data under study, leading to conclusions that were sub-optimal with respect to the quality of the field observations. When we were writing the first English edition of ***Numerical ecology***, this warning mostly concerned multivariate versus elementary statistics. Nowadays, most ecologists are capable of using multivariate methods; the above remark now especially applies to the analysis of spatially or temporally correlated data (see Section 1.1; Chapters 12 to 14) and the joint analysis of several data tables (Chapter 11).

Computer packages provide easy access to the most sophisticated numerical methods. Ecologists with inadequate background often find, however, that using high-level packages leads them to dead ends. In order to efficiently use the available numerical tools, it is essential to clearly understand the principles that underlay numerical methods, and their limits. It is also important for ecologists to have guidelines for interpreting the heaps of computer-generated results. We therefore organized the present text as a comprehensive outline of methods for analysing ecological data, and also as a practical handbook pointing to the most commonly-used packages.

Our experience with graduate teaching and consulting has made us aware of the problems that ecologists may encounter when they first use advanced numerical methods. Any earnest approach to such problems requires in-depth understanding of the general principles and theoretical bases of the methods to be used. The approach followed in this book uses standardized mathematical symbols, abundant illustration, and appeal to intuition in some cases. Because the text has been used for many years for graduate teaching and greatly improved along the process, we know that, with reasonable effort, readers can get to the core of numerical ecology. In order to efficiently use numerical methods, their aims and limits must be clearly understood, as well as the conditions under which they should be employed. In addition, since most methods are well described in the scientific literature and are available in computer packages, we generally devote most of the text to the ecological interpretation of the results; computation algorithms are described only when they may help readers to understand methods. Methods described in the book are systematically illustrated with numerical examples and/or applications drawn from the ecological literature, mostly in English; references in languages other than English or French are generally of historical nature.

The expression *numerical ecology* refers to the following approach. *Mathematical ecology* covers the domain of mathematical applications to ecology. It may be divided into *theoretical ecology* and *quantitative ecology*. The latter, in turn, includes a number of disciplines, among which *modelling*, *ecological statistics*, and *numerical ecology*. *Numerical ecology* is the field of quantitative ecology devoted to the numerical analysis of ecological data sets. Community ecologists, who generally use multivariate data, are the primary users of these methods. The purpose of numerical ecology is to describe and interpret the structure of data sets by combining a variety of numerical approaches. Numerical ecology differs from descriptive or inferential *ecological statistic* in that it combines relevant multidimensional statistical methods with heuristic techniques (e.g. cluster analysis) that do not have a firm statistical foundation. In addition, it often incorporates into the analysis of multivariate data constraints that represent ecological hypotheses, e.g. spatial or temporal contiguity, or relationships between community structure and environmental variables. Numerical ecology also differs from *ecological modelling*, even though the extrapolation of ecological structures is often used to *forecast* values in space or/and time (through multiple regression or other similar approaches, which are collectively referred to as *correlative models*). When the purpose of a study is to *predict* the critical consequences of alternative solutions, ecologists must use *predictive ecological models*. The development of such models, which predict effects on some variables caused by changes in others, requires a deliberate causal structuring. This approach must be based on ecological theory and include a validation procedure. Because the ecological hypotheses that underlay causal models are often developed within the context of studies that use numerical ecology, the two fields are often in close contact.

Ecologists have used quantitative approaches since the publication by Jaccard (1900) of the first association coefficient. Floristics developed from that seed, and the method was eventually applied to all fields of ecology, often achieving high levels of

complexity. Following Spearman (1904) and Hotelling (1933), psychometricians and social scientists developed non-parametric statistical methods and factor analysis and, later, nonmetric multidimensional scaling (nMDS). During the same period, anthropologists (e.g. Czekanowski, 1909, 1913) were interested in numerical classification, and economists started to developed numerical indices (e.g. Gini, 1912). The advent of computers made it possible to analyse large data sets, using combinations of methods derived from various fields, supplemented with new mathematical developments. The first synthesis was published by Sokal & Sneath (1963), who established *numerical taxonomy* as a new discipline.

Numerical ecology combines a large number of approaches, derived from many disciplines, in a general methodology for analysing ecological data sets. Its chief characteristic is the *combined* use of treatments drawn from different areas of mathematics and statistics. Numerical ecology acknowledges the fact that many of the existing numerical methods are *complementary* of one another, each one allowing the exploration of a different aspect of the information underlying the data; it sets principles for interpreting the results in an integrated way.

The present book is organized in such a way as to encourage researchers who are interested in a method to also consider other techniques. The integrated approach to data analysis is favoured by numerous cross-references among chapters and the presence of sections devoted to syntheses of subjects. The book synthesizes a large amount of information from the literature, within a structured and prospective framework, to help ecologists take maximum advantage of the existing methods.

This third English edition of **Numerical ecology** is deeply revised and largely expanded compared to the second English edition (Legendre & Legendre, 1998). It contains a new chapter dealing with multiscale analysis by spatial eigenfunctions (Chapter 14). In addition, new sections have been added in several chapters and others have been rewritten. These include the sections (numbers given in parentheses) on: autocorrelation (1.1), singular value decomposition (2.11), species diversity through space (6.5.3), the double-zero problem (7.2.2), transformations for community composition data (7.7), multivariate regression trees (8.11), and matrix comparison methods (10.5). Sections 11.1 on redundancy analysis and 11.4 on canonical correlation analysis have been entirely rewritten, and a new Section 11.5 on co-inertia and Procrustes analyses has been added. New sections, found at the end of most chapters, list available computer programs, with special emphasis on R packages.

The present work reflects the input of many colleagues, to whom we express here our most sincere thanks. We first acknowledge the outstanding inputs of the late Professor Serge Frontier (Université des Sciences et Techniques de Lille) and Professor F. James Rohlf (State University of New York at Stony Brook) who critically reviewed our manuscripts for the first French and English editions, respectively. Many of their suggestions were incorporated into the texts that are at the origin of the present edition. We are also grateful to the late Professor Ramón Margalef for his support, in the form of an influential Preface to the two French and the first English editions. Over

the years, we had fruitful discussions on various aspects of numerical methods with many colleagues, whose names have sometimes been cited in the Forewords of previous editions.

During the preparation of this new edition, we benefited from the help of several colleagues. First and foremost is Daniel Borcard; after 20 years of scientific collaboration with one of us, he undertook to write a book, ***Numerical ecology with R*** (Borcard *et al*., 2011), which is the companion to the present manual. That book shows readers how to use the R language to carry out calculations for the methods described in the present book. In addition, Daniel Borcard revised several chapters and sections of this new edition, including Sections 1.1 and 10.5, Chapter 14, and all the Software sections found at the end of the chapters. He also carried out the simulations for the Dagnelie test of multivariate normality reported at the end of Section 4.6, and he developed the method of selection of rare species to be used before correspondence analysis (Box 9.2). Jari Oksanen developed an algorithm combining PCA/RDA/partial RDA and gave us permission to reproduce it in Table 11.5. We are most grateful to these two researchers for their major contributions to our book.

Other long-time collaborators and friends helped us by revising sections of the book that were either new or had been rewritten and modernized. We are most thankful to Marie-Josée Fortin who revised Section 1.1, François-Joseph Lapointe for Section 8.13, Miquel De Cáceres for Subsection 8.9.3, Stéphane Dray and Pedro Peres-Neto for Section 11.5, Patrick M. A. James for Subsection 12.5.4, and Helene H. Wagner for Subsection 13.1.4. Cajo J. F. ter Braak and Jari Oksanen commented on portions of Section 11.1. The new Chapter 14 received special attention: it was entirely revised by Daniel Borcard and Pedro Peres-Neto, whereas other colleagues revised the sections describing methods that they had contributed in developing: Stéphane Dray for Section 14.1 and 14.2, F. Guillaume Blanchet for Section 14.3, Helene H. Wagner for Section 14.4, Miquel De Cáceres for Subsection 14.5.1, and Guillaume Guénard for Subsection 14.5.2.

Graduate students in our home universities and those who participated in short courses that we gave in several countries abroad have greatly contributed to the book by raising interesting questions and pointing out weaknesses in previous versions of the text.

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 the work and for all the opinions expressed therein. We shall therefore welcome with great interest all suggestions or criticisms from readers.

PIERRE LEGENDRE, Université de Montréal
LOUIS LEGENDRE, Université Pierre et Marie Curie Paris 6   April 2012