Chapter 1. General introduction

by Guy BOUXIN°

This text dealing with the analysis of vegetation data is mainly concerned with the statistical treatment of the tables of floristic and mesological relevés. It is not exhaustive but emphasizes the understanding of the successive steps of the analytical approach, the only way for us to arrive at a good use of statistical techniques.

The study of vegetation has undergone numerous developments over the last few decades, notably owing to the widespread access to computer techniques and multivariate analysis softwares. This has favored the use of statistical techniques which were formerly decried or at least viewed with great suspicion. But the results did not always follow because the automatic use of available tools has too often eliminated the preliminary reflection and indispensable to carry out all the steps of a study of vegetation.

One of the objectives of this text is to show that a study of vegetation is a whole, that the successive steps of the scientific approach are intimately linked, from field collection to final synthesis. Botanists generally do their fieldwork by collecting data in relatively small and few relevés, in relation to the area under consideration, separated from one another by large distances. Generally, it is only after, that is too late, that they care about the analysis of the data. This process brings only limited results with the false security of a mathematical treatment. Often, the user is satisfied only if the results of the analysis confirm the field impressions.

Let us resume the usual steps of the analysis of vegetation data:

- the field data collection; these are floristic or mesological data;
- the presentation of the data in a double entry table, the lines corresponding to the species or factors of the environment and the columns to the sites;
- the analysis and synthesis of data;
- making relationships vegetation data with environmental data.

However, to be successful, the observer must be fully aware of the type of work he is undertaking and, above all, of defining his objectives (JONGMAN *et al*., 1987). How does his work fit in? All the steps should logically be adapted according to the objectives.

° rue des Sorbiers, 33 à B.5101 Erpent mail : guy.bouxin@proximus.be

Yet, as GOUNOT (1969) pointed out, the use of the statistical tool should lead to the identification of principles that can be used to synthesize and integrate as far as possible existing theories and methods of vegetation study. The quantitative approach of this study was made possible by the efforts of Anglo-Saxon researchers of previous decades. But we must endeavor to incorporate in it all the contribution of continental European schools, empirical but certainly nourished by an unparalleled concrete experience. GOUNOT (1969) adds: "It seems essential to us to identify the specific nature of vegetation problems from the outset, since methods are never neutral despite appearances. The theoretical model assumed (implicitly or explicitly) by any statistical technique is well compatible with the reality that we seek to apprehend. It is worth recalling briefly the problem of the relations between phytosociology and ecology (GUINOCHET, 1973). For the latter, phytosociology is the study of plant associations, the very existence of which, or rather the possibility of bringing them to light by carrying out a certain number of well-defined operations, constitutes a phenomenon in itself of which it is legitimate to take care of himself. Phytosociology appears to be a discipline in its own right, the objective of which is not only the floristic diagnosis and the classification of plant associations, but also the study of their dynamics, their relations with environmental variables, their history, that is to say their evolution and their genesis. For GUINOCHET, phytosociology should not be subsumed under ecology. This shows the opposition with GREIG-SMITH (1964) who was one of the pioneers in the quantitative study of vegetation and who writes at the beginning of chapter 1 of his book "Many ecological data take the form of description of vegetation with or without concurrent recording of factors of the environment. Such data have formed the main basis of most ecological theory and concepts and are likely to continue to do so". In the same vein, GAUCH (1982) speaks of "community ecology" and ORLÓCI (1978) of "vegetation ecology". This different conception of the place of the study of vegetation has had a major influence on the way of organizing scientific work. Phytosociology, *sensu stricto*, has developed a very elaborate formalism, making it difficult to access the "uninitiated" and isolating it, while limiting its development.

It is impossible to take an interest in the study of vegetation without referring to another passionate opposition between those who see vegetation as a *discontinuum* (Sigmatist school, see GUINOCHET, 1973, p. 159 *et seq*.) and those who have an individualistic interpretation of species behavior and base their studies on the concept of *continuum* (WHITTAKER 1973). A thorough analysis of vegetation shows how futile this opposition is. Some authors thus confuse model and reality, that is, they look at vegetation according to the model of description they use. Their approach to description is necessarily influenced by this model, which becomes a paradigm, even a dogma (see LEGAY, 1996).

The problem of describing vegetation data in plots or relevés and the elaboration of a sampling frame will not be developed in a particular chapter since detailed informations were presented by GOUNOT (1969) GREIG-SMITH (1964), GUINOCHET (1973), KERSHAW (1964), KNAPP (1984), ORLÓCI & KENKEL

(1985), PODANI (2000) and VANDEN BERGHEN (1973). Sampling will be evoked in the selected examples.

It is thus possible to construct many different tables, depending on the mode of data collection (*e.g.* in a series of relevés located in a single site or in sites of one or more types of communities). Each column in the table corresponds to a plot, site or relevé and each line to species or environmental parameters. The phytosociological table thus forms a primary matrix from which the ecologist hopes to arrive at a synthesis of the description of the pattern of variation in nature and consequently to interpret it.

With the built table, the analysis of data can begin, but first of all what is a data analysis? Let us return to the terms of BOUROCHE & SAPORTA (1980): "Classical statistics have focused on the study of a single character (or variable) measured on a small set of individuals. However, in practice, the individuals observed are frequently characterized by many characters. Data analysis methods allow an overall study of these variables, highlighting relationships, resemblances or differences by dipping individuals and variables into geometric spaces while making the largest saving of hypotheses and transforming the data to visualize them in a plane or classifying them into homogeneous groups and this while losing the minimum information ". It is therefore a "multidimensional descriptive approach". "We usually distinguish two sets: the individuals and the characters relative to these individuals". Always quoting the authors: “The whole of the observed individuals comes either from a sampling in a population (in the case of a relevé) or from the entire population. The objective is not to look for characteristics of a population but to describe the structure of the set of observed individuals, abandoning any pseudo-inductive will".

Correspondence analysis was born in this spirit. HUSSON *et al*. (2009) explain that correspondence analysis was invented to deal with large linguistic data and that it was the first link in a large set of statistical techniques which proposed to approach the statistical tables independently of any mathematical hypothesis or probabilistic data structure. Still quoting the same authors, "A multidimensional approach implies taking into account simultaneously all the relationships between the variables taken two by two. This is what is done, for example, in the highlighting of synthetic variables: such a variable represents several others, which implies that it is linked to each of them, which is only possible if the latter are themselves linked together in pairs. The notion of synthetic variable is therefore intrinsically multidimensional; it is a powerful tool for describing an array individuals x variables ".

Why give so much importance to the statistical tool? The definition of SCHWARTZ (1994) is already a response: "Statistics is a way of thinking to collect, process and interpret the data we meet in various fields, particularly in life sciences, because these data have an essential characteristic: the variability ". Or again: "The individual differs from other individuals, he differs from himself from moment to moment. Thus, the domain of life is made up of particular cases. But there is no science but the general. How then can there be a science of life? This poses the problem: we must adapt science to the domain of the particular. Because of the variability, we are in the field of uncertainty. To adapt a science to the particular is to invent a science of the uncertain ".

Our aim is not to write theoretical notes on all aspects of the analysis of vegetation data, but to understand the logic of a complete statistical approach in a vegetation study, knowing full well that it will be difficult to persuade those who continue to use the simple technique of row and table swapping or who use techniques that I call mixed, that is, data collection using traditional techniques followed by a summary statistical analysis .

The foundations for a complete analysis of the data were given by GOUNOT (1969) and GREIG-SMITH (1979). GOUNOT was very interested in the structure of the vegetal cover and said in particular: "First of all, the vegetal cover, often, if not always, has a mosaic appearance, consisting of the repetition of structural patterns, usually not many, related to concomitant variations of quantitative and qualitative floristic composition. The homogeneity of the vegetal layer can only exist if the mosaic is repetitive, that is to say, results from a more or less regular arrangement of its different parts. Homogeneity is generally a matter of scale, and the parts of the mosaic which appear to be homogeneous on a certain scale usually show a complex structure if we move to a finer scale. Nested structures imply that there are more or less distinct discontinuities in the vegetal cover, and therefore preferential levels for the analysis of vegetation". Gounot had observed how complex the vegetation layer may be in certain cases, and how difficult it is to give criteria for defining the sampling unit in the general case. A point on which we insist is the opinion of the author on the area-species curve: in the end, it seems that the use of the area-species curve is not a valid method of definition of the minimal area. The methods of practical determination of the curve are either questionable or difficult to attain. On the other hand, it is not even certain that the area-species curve presents a horizontal asymptote. Rather, it appears that the number of species grows more or less linearly as a function of the logarithm of the inventoried surface, in the range of surfaces where the minimum area is generally located. Finally, the shape of the curve is certainly very influenced by the density and the mode of distribution of the species. The result therefore depends on causes so diverse that its theoretical interpretation is difficult. Any reader interested in the species-species relation will consult WILLIAMSON (2003) and WILLIAMSON *et al*. (2001 and 2002). Although GOUNOT (1969) abandoned the technique of the area-species curve, he still persists in seeking true homogeneous communities using ecological, structural and dynamic criteria in order to discriminate mosaics. And he adds: "In any case, there is a lower limit to the resolving power of the vegetation which is of the same order of magnitude as the minimal area (insofar as an area smaller than the minimal area is not representative of the community) ". It is difficult for us to subscribe to this last statement. In 1979, GREIG-SMITH declared: "Though it is convenient to distinguish variation between communities and variation within communities, because techniques of investigation are generally different, there is clearly a continuum both of scale and of intensity of pattern. Ecologists are increasingly aware of the need to "build a bridge" between community ecology and population ecology if we are to further our understanding of the functioning of ecosystems and ultimately to predict or control changes in them". We will see how it is now possible to "throw this bridge" and the impact of this finding in the definition of plant communities.

The scientific approach that we follow in this book is placed in the logic of creation and rejection of models (FOUREZ, 1992). A model does not bring the truth, but it is its effectiveness that interests us. The model of description of the vegetation to which we arrive has already been presented in BOUXIN (1995, 1999) but it evolves constantly (BOUXIN 2011 and 2013). We wanted to develop it completely and especially to explain the stages that led to its elaboration.

The current site includes five chapters:
- A general introduction
- Presentation of the Akagera data file
- Description of the main multivariate analyzes found in the literature
- The approach showing the benefit of data analysis and the steps to get them rather than an ordination
- Personal software made available to researchers.

An appendix includes the Akagera file and the complete results of multivariate analyses.

This book is also an occasion to pay tribute to some pioneers in the statistical analysis of vegetation and who directly helped or influenced me in one way or another: P. BERTHET, D. CHESSEL, P. DAGNELIE, GOUNOT, L. ORLỎCI and E. van der MAAREL. Without them, this work would never have come about.

January 2024

References

BOUROCHE, J.-M. & SAPORTA, G.. (1980). *L'analyse des données*. Presses universitaires de France. Collection "Que sais-je ?". 127 pp.

BOUXIN, G. (1995). Démarche conduisant à la mise au point d'une technique de description de la végétation. *Acta Botanica Gallica* **142**: 533-540.

BOUXIN, G. (1999). Description de la végétation aquatique et du bord de l'eau dans le bassin hydrographique de la Molignée (Condroz, Belgique) par l'analyse multiscalaire des motifs. *Ecologie* **30**: 139-163.

BOUXIN, G. (2011). Evolution de la végétation macrophytique et trophie dans deux ruisseaux du bassin hydrographique de la Molignée (Condroz, Belgique). *Revue des Sciences de l'Eau* **24**: 256-266.

BOUXIN, G. (2013). Végétation macrophytique, environnement et qualité d'eau dans le bassin versant du Bocq (Belgique, Wallonie). *Revue des Sciences de l'Eau* **26**: 1-19

FOUREZ, G. (1992). *La construction des sciences*. Deuxième édition revue. De Boeck Université. 288 pp.

GAUCH, H.G., Jr. (1982). *Multivariate analysis in community ecology.* Cambridge University Press. 298 pp.

GOUNOT, M. (1969). *Méthodes d'étude quantitative de la végétation*. Masson, Paris. 314 pp.

GREIG-SMITH, P. (1964). *Quantitative plant ecology (2nd ed.)*. Butterworths, London. 256 pp.

GREIG-SMITH, P. (1979). Pattern in vegetation. *Journal of Ecology* **67**: 755-779.

GUINOCHET, M. (1973). *Phytosociologie*. Masson et Cie. 227 pp.

HUSSON, R., LÊ, S. & PAGÈS, J. (2009). *Analyse des données avec R*. Presses Universitaires de France. 224 pp.

JONGMAN, R.H.G., ter BRAAK, C.J.F. & van TONGEREN, O.F.R. (1987). *Data analysis in community and landscape ecology*. Pudoc Wageningen. 299 pp.

KERSHAW, K.A. (1964). *Quantitative and dynamic ecology*. Edward Arnold, London. 183 pp.

KNAPP, R. (1984). *Sample (relevé) areas (distribution, homogeneity, size, shape) and plot-less sampling*. In Knaap, R. (ed.), Sampling methods and taxon analysis in vegetation science. Dr W. Junk Publishers, The Hague. 105-119.

LEGAY, J.-M. (1996). *L'expérience et le modèle. Un discours sur la méthode*. INRA editions. 112 pp.

ORLÓCI, L. (1978). *Multivariate analysis in vegetation research* (2nd ed.). Dr. W. Junk, Publishers, The Hague. 451 pp.

ORLÓCI, L. & KENKEL, N.C. (1985). *Statistical Ecology Monographs. Vol. I. Introduction to data analysis*. International Co-operative Publishing House. 339 pp.

PODANI, J. (2000). *Introduction to the exploration of multivariate biological data*. Backhuys Publishers, Leiden. 407 pp.

SCHWARTZ, D. (1994). *Le jeu de la science et du hasard. La statistique et le vivant*. Flammarion. 111 pp.

VANDEN BERGHEN, C. (1973). *Initiation à l'étude de la végétation*. Les Naturalistes Belges. 236 pp.

WHITTAKER, R.H. (1973). *Ordination and classification of communities.* Handbook of vegetation science. Part V. Dr. W. Junk b.v. - Publishers - The Hague. 737 pp.

WILLIAMSON, M. (2003). Species-area relationships at small scales in continuum vegetation. *Journal of Ecology,* **91**: 904-907.

WILLIAMSON, M., GASTON, K.J. & LONSDALE, W.M. (2001). The species area relationship does not have an asymptote ! *Journal of Biogeography,* **28**: 827-830.

WILLIAMSON, M., GASTON, K.J. & LONSDALE, W.M. (2002). An asymptote is an asymptote and not found in species-area relationships. *Journal of Biogeography,* **29**: 1713.