CHARACTER-BASED COMMUNITY ANALYSIS: THE THEORY AND AN APPLICATION PROGRAM

Valério DePatta Pillar and László Orlóci

Electronic Edition
Available at http://ecoqua.ecologia.ufrgs.br/

CONTENTS

FOREWORD TO THE ELECTRONIC EDITION

The first edition was published in 1993 by SPB Academic Publishing in The Hague, The Netherlands. Since it is out of print we offer this electronic version free of charge on the Internet. The contents remain essentially unchanged. The field has developed further since, particularly in the interpretation of plant functional types, and the methodology has been refined by new developments (Pillar 1999, Pillar and Sosinski 2003, and references therein). The application program SYNCSA has been updated accordingly. The new version for Macintosh and Windows systems is available at http://ecoqua.ecologia.ufrgs.br, which includes all features described in Chapter 6 and new options that are explained in the auxiliary files that come with the program. The developments notwithstanding, we feel that the fundamental problem in the analysis of plant attributes in sets by way of a nested statistical model has not bypassed the original text.

How to cite this edition:

Pillar, V.D. & L. Orlóci. 2004. *Character-Based Community Analysis: The Theory and an Application Program*. Electronic Edition available at http://ecoqua.ecologia.ufrgs.br. 213 p.

Porto Alegre
October 2004

PREFACE

The taxonomies applied to recognize populations may lead to different perceptions of the vegetation. In other words, vegetation analysis is definitely taxonomy dependent. This fact has been notoriously unobserved in the dominant ecological traditions that used species composition as a basis for community comparisons. But we know that species based comparisons have only limited applicability which raises the taxonomy dilemma in vegetation study.

The problem is amplified by the expected changes in the global environment, calling for models to predict the broad-scale effects on vegetation. Species could be the basis of modeling, if it were not for the requirement of broad geographic applicability, and because of this vegetation science needs alternative, character-based comparative schemes. There are many character-based, species-free schemes that may serve the purpose, but which could be best suited for the task?

The approach offered in this book may help to answer this question.

Character-based community analysis is not new, but in its recent state it lacks a coherent theory and appropriate analytical tools. This book is an attempt to develop both theory and tools adopting Orlóci's character-based hierarchical analysis as the model. The result is a new comprehensive approach and a complex algorithm for its implementation. Specifically, Chapter one explains basic concepts of a unifying character-based approach and reviews past work on alternative plant taxonomies. Chapters 2 to 5 present original propositions regarding: community resemblance measures, structural evaluation functions, character ranking, randomization-based probabilistic assessments, fuzziness of population taxa in community studies, and the application of eigenordination. Chapter 6 describes how to use Pillar’s application program SYNCSA to perform the analyses, and Chapter 7 gives examples suggesting areas of utility for the approach.

SYNCSA is written in C and offered as a stand-alone application to users of Macintosh computers. It implements character-based analysis as described in the book and also other methods not specific to the approach, composing an integrated tool for the numerical exploration of community, population and environmental data structures.

Typical areas of application are the interpretation of compositional structures in terms of environmental conditions and plant character patterns, the identification of plant attributes that are ecologically more relevant and useful to predict vegetation pattern, and the quantitative assessment of structural convergence in geographically distant communities that do not share a common flora. The methods are also applicable to community descriptions that inform only the species composition, but, as in these the defining character set contains only one character, the species affiliation of the organisms, character ranking and the fuzzy population taxa approach do not apply.

Those to whom we are much indebted for invaluable assistance in data collection include Drs. J. P. Lewis, J. Ambrosio de Araújo, Mrs. Ilsi Boldrini, Mrs. Márta Orlóci, Mr. Q. Wu, Drs. D. Prado, F. Vervoorst, Mrs. Sandra S. Silva, and the anonymous rural workers in the Caatinga. We also extend our thanks to Drs. R. C. Jancey, A. Lachance, J. B. Phipps, S. X. Yu, A. Palmer and A. Szilárd for their comments and suggestions, to Ms. Xiaoshuang He for editorial assistance, and to the Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (CNPq), the National Science and Engineering Research Council of Canada (NSERC), and University of Western Ontario for financial support and facilities.

Most of all, we thank Mariza and Márta for their enduring support and patience.

V. D. Pillar
L. Orlóci

Porto Alegre and London,
March 1993

Chapter 1

ALTERNATIVE TAXONOMIES

1.1. The problem of perception

It is quite evident that the taxonomy we use and the scale of our investigation affect the perception of natural objects and phenomena. Taxonomies categorize objects and are needed to simplify the display of diversity at the chosen scale. Different taxonomies reflect differences in objectives. Furthermore, the categories recognized may be crisp or ill-defined, fuzzy. When the objects are plant organisms and communities are to be described, we have to define the unit plant community and the community components. This can be a complicated task since the unit plant community is an areal dissection of the landscape, and the community components are dissections of the organismal multitude. Therefore, it is inexorable we have to choose the scale and the taxonomy (or accept others' choices), and these choices will affect the results.

The study of vegetation on both the population and community levels is mainly comparative. It involves the description of *structures* and the analytical search of biotic and abiotic *structural connections*. In the real space of a landscape, structures are manifested through variation in intrinsic and extrinsic attributes. The intrinsic attributes pertain to the plants themselves (termed the *community components*), on any level of organization, from organism to population, or community. Extrinsic attributes are environmental, biotic and abiotic. Structures also exist in abstract, analytical space (Goodall 1963, Feoli 1984), which is usually multidimensional. In one case, the reference axes of the space are the attributes and the points are the *unit plant communities*. In the other case, the role of attributes and units as reference axes is reversed. The *structural connections* are physical properties, such as the causal links in real space and the density and correlation patterns in analytical space.

The definition of the *unit plant community* is a question of one's purpose and one's adherence to a plant community concept (Mueller-Dombois and Ellenberg 1974:33). To further complicate things, arbitrary decisions are involved, linked with sampling considerations of the unit (plot) size and shape (see, *e.g*, Kershaw 1964:26, Greig-Smith 1983:19, Juhász-Nagy and Podani 1983, Podani 1982, 1984, Kenkel 1984:54, Kenkel, Juhász-Nagy and Podani 1989, Orlóci and Pillar 1989). Here the term unit plant community is used interchangeably with *quadrat*, *plot*, *stand*, or in a looser sense, *plant community* whose description is the *relevé*. In either case, the size and shape of the unit plant community determine the structures and connections that we can see. This implies that a scale dependence is inherent to community studies.

The *community components* are the entities on the basis of which intrinsic vegetation structures are defined. The individual plant organism is the basic community component that survives environmental adversity and reproduces in the community's site. But, for practical reasons we represent the organism by a population of similar organisms as the operational community component, and this requires a taxonomy of plant organisms. The dominant tradition in vegetation science has given emphasis to the taxonomic scheme that uses common inheritance as its defining criterion for taxa. When inheritance is used, the community structures are defined in terms of species. However, other taxonomic schemes also exist, such as the ones recognizing life-forms (*e.g.*, Raunkiaer 1904, Mueller-Dombois and Ellenberg 1974), growth-forms (*e.g.*, Warming 1909, Dansereau 1957:148, Barkman 1988a, Halloy 1990), phenological types (Barkman 1988a), or character set types (Orlóci and Orlóci 1985). It should be expected that these different taxonomies will dissect the multitude of community components differently, and, similarly to the case of scale dependence, the perception of structures and structural connections will be affected (see Feoli, Orlóci and Scimone 1984). Therefore, there is indeed a *taxonomic dependence* of community studies, a fact which has been neglected in the literature.

Perception also depends on fuzziness. Plant communities and populations, and their classes, are ill-defined entities, in general with no sharp limits in real or abstract space. Vegetation structure is inherently fuzzy (see Orlóci 1988a, Roberts 1989, Feoli and Orlóci 1991:ix). The way this fuzziness is treated in data analysis may determine the nature of biotic or environmental connections that are perceived. For instance, a fuzzy cluster analysis interprets each community as a member, though with different degrees, in all groups (see Feoli and Zuccarello 1986, 1988), emphasizing the perception of spatial continuity instead of discreteness. Furthermore, analysis considering community components as fuzzy entities interprets the absences of taxa in one community as not absolute, avoiding the problem of analytical indeterminacy, and leading to the perception of vegetation structures that otherwise would not be revealed (see Chapter 3).

The fact that there are alternative taxonomies of plant organisms, potentially leading to different perceptions of the vegetation, brings up the question of which taxonomy to use. Under a utilitarian viewpoint, the answer has to depend on the objectives of the study. The approach adopted in the next chapters treats the taxonomy as a variable, and for each aspect of the community structure and connection to be described with clarity there should be a taxonomy that best suits the task.

1.2. The relevé

The description of a unit plant community involves the *delimitation* and *evaluation* of populations The delimitation of populations requires a taxonomy, which we call the *primary taxonomy*. The evaluation of populations uses (1) one or more intrinsic, organismic level characters (*e.g.*, species, life-form, leaf deciduousness, etc.), which we call the *defining characters*, and (2) population level characters (*e.g.*, cover-abundance, local frequency, or other performance measure). The defining characters describe another taxonomy of organisms, to which we refer as the *analytical taxonomy*. On the basis of the analytical taxonomy community components are recognized and comparisons between communities are made.

The primary taxonomy is informal, intuitive, necessary to distinguish discrete populations in a unit plant community during the field survey. In practice, plant populations so delimited may coincide with species populations, but not necessarily. It is essential, though, to take into account the relative homogeneity of the defining characters within a population delimited in the primary taxonomy. As a consequence, the populations delimited in the primary taxonomy the first time may have to be subdivided the next time in an iterative process in the field. The iterative redefinition is the sole mechanism to refine primary taxa. In the analytical taxonomy this cannot be done; primary taxa can only be lumped. Ideally the ratio of the within and between population variability should be minimal. Its actual magnitude, however, depends on the defining character set and on the heterogeneity of the unit plant community, such as its areal size and shape, seasonal state, ground pattern, etc.. For instance, if the primary taxonomy delimits species and no evaluation is pursued of the defining characters other than the population's species identity, there will be no within population heterogeneity. In this case, the analytical taxonomy is defined by one character, which is the "species affiliation". In another cases, the heterogeneity within populations delimited by the primary taxonomy may become a constraint. For instance, the primary taxonomy recognizes species and the defining characters of the analytical taxonomy are life-form and plant height; if one of these characters varies more within species than between species, the primary taxonomy will be proportionately unreliable and will have to be redefined, or the character will not provide valuable information.

When the primary taxonomy recognizes species, and their names are recorded, the information on the defining characters can be drawn from external sources, instead of an in-site evaluation. This is a common procedure in community studies for understandable reasons, considering the technical problems involved. But, accuracy is critical and it depends on the species' phenotypic variability in the defining character set. The danger exists that the external data may not match the actual states of the defining characters in the site.

1.3. Taxonomy

Taxonomy as a process has been defined by Simpson (1961:11) as "the theoretical study of classification, including its bases, principles, procedures and rules." Here, though, *taxonomy* is mainly used to designate the result of the taxonomic process, as a synonym of *classification*. A taxonomy is based on one or more characters. In the context of organismal taxonomy, a *character* refers to "a feature which varies from one kind of organism to another" (Michener and Sokal 1957). *Character states* are the different qualitative or quantitative aspects of a character within a set of organisms. Character types range from very simple to very complex. In one extreme, a character can be a *unit character* defined by Sneath and Sokal (1973:72) as "a taxonomic character of two or more states, which within the study at hand cannot be subdivided logically, except for subdivision brought about by the method of coding." For instance, presence of hairs on the leaf dorsal surface can be taken as a unit character. Most characters are *complex*, in which case the states are the taxa, or forms, defined by an implied *subtaxonomy*. For instance, leaf shape is a complex character that requires a classification of leaf shapes to decide the state to which a given organism belongs. The same can be said for growth-form, life-form, phenological types, or even species; they can all be interpreted as complex characters. By extension, the analytical taxonomy that recognizes species populations has one defining character, the species, and the species, to be identified, needs in turn a subtaxonomy, with its own defining characters.

A fundamental part of a taxonomy is the defining character set. Another is the set of rules that are followed to classify the organisms. The defining characters are selected according to the objectives of the taxonomy, no matter whether they are defining the analytical taxonomy or a subtaxonomy. When, for instance, species affiliation is the defining character in the analytical taxonomy, the criteria for character selection in the subtaxonomy that recognizes species are a priori defined in a context and with objectives not necessarily consistent with a vegetation study. The objectives of the plant taxonomy that recognizes species as a taxonomic rank have been debated and appear to be mainly to maximize predictivity via a general purpose, phenetic classification, or to reflect common ancestry in a phylogenetic sense (see Michener 1970, Stace 1989:11, Stuessy 1990:51). Linked with this is the controversial nature of the species concept, whether phenetic or biological (see Mayr 1942:119, Davis and Heywood 1963:89, Sokal and Crovello 1970, Michener 1970, Stuessy 1990:161). Despite the controversy about objectives or the species concept, little change is evident in these respects (see Walters 1963) that would have affected the results of the use of "species" in many community descriptions. Concerning species recognition, the plant taxonomist chooses from characters that are constant within taxa and relatively insensitive to environmental variation (Davis and Heywood 1963:119, Heywood 1976:44, Harper 1982, Stace 1989:184, Stuessy 1990:38). As such, these characters have more to do with the transfer of genetic information than of matter and energy (Lausi and Nimis 1985). Alternative taxonomic schemes have been specifically proposed for vegetation description (see later in this chapter). These schemes delimit plant taxa by morphological and functional characters, often in connection with a view at plant survival under specified environmental conditions.

1.4. Adaptive significance of characters

The ultimate aim of vegetation studies is to find causal explanations for structures or phenomena, in order to allow postdictions or predictions. Linked with this is the search for connections between vegetation and environment in space and time, and their explanation on the basis of plant adaptations to biotic and abiotic conditions. The meaning of *adaptation* has to be clarified. Harper (1982) defines adaptation as "any aspect of form or behavior that *at a reasonable guess* is the result of natural selection". There is a controversial aspect implied, since "adaptation" appears to reinforce a teleological view of nature. In this, an organism has goals (functions), which are better achieved when it "acquires" adaptations (see Thompson 1917:672, Rosenberg 1985:37). In other words, teleology presents "goals" as the causes of adaptation. With recourse to Darwin's theory of natural selection, the causal explanation focuses on variation and selection in the past (Rosenberg 1985:46). In this respect, Mayr (1961, 1982:67-9) points out that biological processes have "proximate" causes, shared with other physical phenomena, and "ultimate" ones, typically evolutionary, specific to biological phenomena, that acted in the past. The controversy is not settled by tautological statements (see Peters 1976) that adaptation is "any structure, physiological process, or behavioral pattern that makes an organism more fit to survive and to reproduce" (Wilson 1975:577), for "to explain the survival of one group of organisms by appealing to their fitness is to explain survival by appeal [ing] to survival" (Rosenberg 1985:127). Fitness, however, can be explained by other means, such as mechanical efficiency (Thompson 1917:673), energy capturing efficiency, and water use efficiency (see among others Parkhurst and Loucks 1972, Orians and Solbrig 1977b, Givnish 1979, 1986), but one has to assume that plants are efficient.

Related to the conceptual controversy surrounding "adaptation", a relevant question in vegetation studies is whether the state of a form, or phenotype in a broader sense, is indicative of particular extrinsic conditions. On the one hand, this is intuitively true for (1) traits developed under selective pressures still prevailing. These traits are here understood as *adaptive*, or *functional* in ecological usage. An analytical taxonomy which draws upon these traits, or upon other traits linked to these, is expected to reveal community structures closely related to environmental variation. On the other hand, there are (2) traits selected by past conditions, no longer prevailing, which are selectively neutral under present day conditions, and (3) traits arisen by genetic drift which have been neutral to the environmental conditions so far prevailing (see Wilkins 1968). Traits among (2) and (3) are probably not adequate to define an analytical taxonomy aiming at revealing ecologically relevant community structures.

Comparative studies have shown environmental trends in the variation of several plant characters (see Givnish 1987). These trends indicate that probably the characters in focus are adaptive in the foregoing sense. In this respect, it is observed, for instance, that leaf size tends to decrease with decreasing precipitation, humidity and/or soil fertility (Schimper 1898:8, Raunkiaer 1916, Sarmiento 1972, Lausi and Nimis 1986). Also, leaf size decreases with increasing light intensity from the ground up through the canopy (Cain *et al.* 1956, Parkhurst and Loucks 1972, Bongers and Popma 1988), and with increasing altitude (Parkhurst and Loucks 1972). Leaf thickness also increases with aridity and light intensity through the canopy (Schimper 1898:8, Lausi and Nimis 1986). Furthermore, compound leaves are most common in plants growing in drier habitats or in those invading light gaps (Givnish 1978, Stowe and Brown 1981). Leaf pubescence and the presence of glands or salt bladders tend to increase with increasing aridity (Ehleringer *et al.* 1981, Ehleringer and Werk 1986, Lausi, Nimis and Tretiach 1989). Steeply inclined leaves are observed in more arid sites (Ehleringer and Werk 1986), as the N-S oriented leaves of *Larrea divaricata* in the Argentinian Monte, or in taller plants in other types of desert vegetation (Lausi and Nimis 1986). Despite what intuition may suggest, stomata density increases in helio and xerophytic plants (Lausi, Nimis and Tretiach 1989). Leaves with stomata on both leaf faces are more frequent in plants growing in sunnier and/or drier sites (Wood 1934, Lausi, Nimis and Tretiach 1989). Leaves tend to have smaller and isodiametric cells in more arid sites (Lausi, Nimis and Tretiach 1989), or in taller plants in desert vegetation (Lausi and Nimis 1986). In tropical forests, deciduousness is more common in the upper tree story and increases with the lengthening of the dry season (Walter 1973:61). Taller species tend to dominate in soils rich in water and nutrients (see Tilman 1988:146). Barkman (1988b) attributes differences in the drought resistance of *Quercus robur* and *Q. rubra* to leaf inclination, leaf size and leaf thickness.

It is important to note that different characters may have parallel ecological trends. As Warming (1909:3) points out, the plant carries hereditary constraints ("phylogenetic constraints" in Givnish 1987) which

> "render it possible for different species, in their evolution under the influence of identical factors, *to achieve the same object by the most diverse methods*. While one species may adapt itself to a dry habitat by

> means of a dense coating of hairs, another may in the same circumstances produce not a single hair, but may elect to clothe itself with a sheet of wax, or to reduce its foliage and assume a succulent stem, or it may become ephemeral in its life-history."

Despite the rather teleological phrasing, this fact, on the one hand, explains the relatively high diversity of plant forms that is found within similar environmental conditions. On the other hand, it makes even more complex the task of revealing vegetation patterns using plant characters other than the species affiliation.

Optimality models have been proposed to explain the causes of plant patterns observed in nature (*e.g.*, see Parkhurst and Loucks 1972, Givnish and Vermeij 1976, Orians and Solbrig 1977b, Givnish 1979, 1986, 1987, Ehleringer and Werk 1986, Tilman 1988). The principle of "optimal design" underlies the modelling (Mooney 1974, Cody 1974, Givnish 1986:3), *i.e.*, "natural selection leads to organisms having a combination of form and function optimal for growth and reproduction in the environments in which they live" (Parkhurst and Loucks 1972). The models are primarily based on the economics of gas exchange, plant support, soil resources supply, and biotic interactions (Givnish 1986). In these models the cost and income of different strategies is assessed under different constrains, in particular the availability of light and soil resources (Orians and Solbrig 1977b, Givnish 1979, Tilman 1988), and rate of loss and mortality (Tilman 1988). The models are built to take into account such things as the trade-offs between carbon gain and water loss, and between leaf photosynthetic capacity and the energetic costs of nutrient capture to construct and maintain a highly photosynthetic tissue (Givnish 1986:11). Model validation is a function of the agreement of optimal strategy dictated by the model and the observed patterns in nature. For instance, Orians and Solbrig (1977b), using a cost-income model, explain the trends in plant strategies found in a gradient of water availability. They conclude that "mesophytic leaves have lower costs and higher potential incomes than xerophytic leaves" and payoff "in a much shorter time provided there is water", and that there is an inevitable association "between ability to photosynthesize rapidly when soil moisture is readily available and inability to extract moisture when soils are drier, and vice versa." Furthermore, the occurrence of plants with mesophytic leaves in very arid conditions is explained by "their shorter amortization time and higher income rates [that] may yield higher yearly profits than xerophytic leaves even when the period of high moisture availability is brief". These plants occupy a niche space available for a very short time.

Givnish (1979) explains leaf size by taking account of its effect on leaf temperature and transpiration rates. The thinner air boundary layer associated with smaller leaves allows them to exchange heat faster than larger leaves, thereby avoiding excessive temperature rise and the costs in unproductive roots and xylem associated with increasing transpiration. Thus, smaller leaves are more efficient when soil moisture is a limiting resource (see also Horn 1971:55). Along the same lines Ehleringer and Werk (1986) explain the functional advantage in arid conditions of surface modifications (waxes, hairs and salt bladders), for they increase leaf reflectance. Also some patterns of leaf inclination are advantageous, for they allow maximal rates of photosynthesis when the evapotranspiration potential is the lowest, and reduce heat at middle day hours. It is a fact that pubescent leaves, such as in *Encelia farinosa*, transpire less than non-pubescent ones, and can remain photosynthetically active for a longer period of time, which compensates for the extra cost of growing the hairs. Horn (1971:53) explains the adaptive advantage of a multilayer over a monolayer arrangement of leaves in drier soil conditions on the basis of leaf temperature too. The multilayer arrangement has a lower heat load than the monolayer. Also, for optimal light interception the optimal leaf size for the multilayer arrangement is smaller than that for the monolayer.

Tilman's (1988) model invokes the rule that trade-offs originate in the physical separation of essential plant resources: terrestrial vascular plants need light, which is available above the soil surface, and water and mineral nutrients, which are available in the soil. The allocation of growth to increase the capture of one of these resources necessarily reduces the growth that can be allocated for the capture of the other, and thus, depending on which resource has limiting availability, different strategies will be optimal. For instance, if more light is to be obtained under light competition conditions, plants must have been selected to allocate more of their growth to stems and leaves instead of allocating to roots; analogously, if more water or soil nutrients is to be obtained in drier or poorer soil conditions, plants must have been selected to allocate more growth to roots than to stem and leaves, which is accompanied by a reduction in light competition. Thus, the model predicts that each habitat along a gradient of limiting soil resources will have a particular physiognomic pattern dictated by the differential allocation between stem/leaves and roots. The model is validated with observed patterns in nature (Tilman 1988:146).

1.5. The species-based taxonomy

The emphasis on a species-based taxonomy for vegetation description has been the dominant tradition since the beginning of this century. We call it species-based because the community components are species populations. That is, the defining character in the analytical taxonomy is only species affiliation. This can be attributed to increased taxonomic knowledge of the worldwide flora and to concerns with economy and generality. Economy of the description comes from applying class (species) properties to describe the plants, rather than field observations of the character states. Generality is enhanced through the use of a unified taxonomy. As pointed out by Braun-Blanquet (1928:21),

> "The countless individuals may be grouped in two distinct ways: under the concept of the taxonomic species or under the concept of the growth-forms or life-forms [...]. The Brussels Congress (1910) [3rd International Botanical Congress] rightly decided in favor of the species as the fundamental unit of the plant community. The concept of 'life-form' is indefinite, has not been adequately defined, and cannot be considered as a sufficient basis for a science of vegetation. Species, however, are groups of individuals with uniform inheritance and have been for many years the objects of careful investigation."

Several authors have questioned the ecological relevance of species as the fundamental unit of plant communities (Salisbury 1940, Constance 1953, Ehrlich and Holm 1962, McMillan 1969, Snaydon 1973, Grime 1979, Harper 1982, Orlóci 1991). As indicated by Ghiselin (1987), species are important only because they are fundamental reproductive units – but nothing more. The main concern in these criticisms is that species in general occur in a too wide environmental range. This is understood in the sense that phenotypic variation within species ("locality", Jancey and Wells 1987) cannot be accounted for, either because "the individuals of the species have wide tolerances or plasticity", or because "the individuals have very narrow tolerances but the nature of the species is such that the taxonomist includes a wide range of locally specialized genotypes within one taxon" (Harper 1982). In any case, the too wide ecological nature of the species is linked to the defining characters which, in general, are resistant to polymorphism or phenotypic variation (see Section 1.4 above). Related to this, the use of narrower units, such as the ecotypes, has been suggested (Braun-Blanquet 1928:21, Salisbury 1940). Despite the dissatisfaction, the use of species as the criterion to delimit populations is almost standard (see Section 1.2), even when further evaluation by other characters is aimed. This should, of course, not be proof that the species concept is necessarily or universally useful in vegetation work.

Further criticism of the species usage may be based on the fact that the species is a geographically limited concept. This is quite evident to those who attempt to compare communities from widely separated areas with different floras, or from different sites within the same floristic region, in which case enhanced analytical indeterminacy (Orlóci and Stofella 1986, Orlóci 1988a) undermines quantitative comparisons.

In another context the use of higher level plant taxa (genus, family, order) as the unit for community characterization has been attempted. Maarel (1972) and Dale (1978) report meaningful results on this basis. Dale and Clifford (1976) find little loss of interpretability when using genus and subgenus categories instead of species to classify communities from a small area, but not as good results are shown when using family and subfamily. The approach has not been explored further, but we can speculate on its shortcomings. First, the ecological range of populations belonging to the same higher taxa, possibly with few exceptions, is wider than in the species; therefore, little improvement is expected in the ecological significance of the results. Second, above the level of species there is less comparability between taxa, because the criteria defining what is a genus, a family, etc., are even less consistent than those defining species (Stace 1989:188). However, we should not disregard the potential utility of this approach in broad, worldwide scale comparisons, since many plant families show climatic trends in their distribution, such as the Cruciferae, Umbelliferae, Rosaceae, Rubiaceae, Euphorbiaceae (Good 1974:55).

1.6. Other taxonomies

The rejection of the species taxon as the unit for vegetation description brings the question of which characters should then be used in an alternative taxonomy. Should the plant characters be functional (or adaptive) is a question that has nourished intense discussion. Life-form for instance, is claimed to be merely a hypothesis about how plants survive. Du Rietz(1931) argues that "life-forms should be founded simply upon the characters observed by us, and not upon what we believe about the probable origin of these characters" (*loc. cit.* p. 42). With different names attributed to the types (life-forms, growth-forms, etc.), several alternative schemes have been proposed to classify plants. In this respect, there are several reviews (*e.g.*, Du Rietz 1931, Adamson 1939, Lacza and Fekete

1969, Fekete and Lacza 1970, Shimwell 1971:63, Barkman 1988a, and Orlóci 1991). In the next paragraphs, we describe and discuss some of these schemes from the viewpoint of finding the underlying characters on the basis of which the types are recognized.

Life-form and *growth-form* are in essence synonymous concepts, despite attempts (*e.g.*, Du Rietz 1931) to restrict the meaning of growth-form to designate types based primarily on shoot architecture. The term life-form was first coined by Warming ("lifsform", 1895) to emphasize the functional value of the defining characters in his plant forms classification, instead of *vegetative form* ("Vegetationformen", first used by Grisebach 1872). However, in the English version of Warming's book (Warming 1909:2) life-form is translated as growth-form. In plant taxonomy, a *form* (not *forma*, the taxonomic rank) is perceived by an intuitive, visual act, as a totality or a "Gestalt" (Heywood 1973, Lausi and Nimis 1985). This view has apparently been transposed into some of the classifications of life and growth-forms, such as Barkman's (1988a), to take the most recent case. That is, firstly types are perceived mentally, and then characters are found that support the types. Even in schemes in which there is no explicit a priori recognition of types, only combinations of character states (*e.g.*, Dansereau 1951, Lausi and Nimis 1985, Orlóci and Orlóci 1985), some of the characters may be complex and implicitly involve a priori types as character states.

The classifications created by Humboldt (1806), Kerner (1863), and Grisebach (1872) are purely physiognomic and do not make any consideration of the ecologically functional value of the defining characters. De Candolle's (1818) system cannot be considered a purely physiognomic one, for it is based on the height of the ligneous stem and on the duration of life, the latter which is not a morphological character. It is interesting to note that Humboldt (1806:220) opposed the use of species to describe vegetation:

> " On taking one general view of the different phanerogamic species which have already been collected into our herbariums [...], we find that this prodigious quantity presents some few forms to which most of the others may be referred. In determining those forms, on whose individual beauty, distribution, and grouping, the physiognomy of the country's vegetation depends, we must not ground our opinion (as from other causes is necessarily the case in botanical systems) on the smaller organs of propagation [...]; but must be guided solely by those elements of magnitude and mass from which the total impression of a district receives its character of individuality. [...] The systematizing

> botanist [...] separates into different groups many plants which the student of the physiognomy of nature is compelled to associate together."

Based on this kind of reasoning, Humboldt describes 19 forms of plants ("Hauptformen"), with some names taken from a characteristic species, genus or family, namely: palms, banana, Malvaceae and Bombacaceae, Mimosa, heath, Cactus, Orchideae, Casuarineae, conifer, Pothos, lianes, Aloe, grasses, ferns, Liliaceae, willow, myrtle, Melastomataceae, and laurel forms. The types in Humboldt's classification are merely physiognomic. As he notes, "The painter [...] distinguishes between pines or palms and beeches in the background of a landscape, but not between forests of beech and other thickly foliated trees". No functional aspects are considered, for the objective is to reveal a general picture of the vegetation, as an artist would paint it. Warming (1909:4) points out the lack of physiological and ecological foundations in Humboldt's physiognomical system.

Kerner (1863:7) distinguishes 11 morphological, purely physiognomic "Grundformen" (basic forms), namely: trees, shrubs, tall herbs, mat plants, herbs, leaf plants (unconspicuous stems), lianas, string (filamentous) plants, reeds, grass form, fungi, and crustose plants. In Kerner (1891:593-777) there is no formal plant growth-form classification, but a classification for each different morphological aspect, such as leaf, venation, stem, and root.

Grisebach's (1872:11) physiognomic system divides plants in "Vegetationsformen" (vegetative forms) named after a characteristic genus and arranged in 7 groups: woody plants (divided in 30 types), succulents (3 types), climbers (3 types), epiphytes (2 types), herbs (8 types), grasses (6 types) and cryptogams (divided in bryophites and terrestrial lichens). The woody plants are subdivided in intermediate groups with many forms. For instance, the "Sträucher" (shrubs) group has 7 vegetative forms: "Erikenform", "Myrtenform", "Oleanderform", "Proteaceenform", "Sodadaform", "Rhamnusform", and "Dornstraücher" (thorn shrub). Trees are classified on the basis of leaf form only.

With Warming's (1884) starts a series of classification schemes (*e.g.*, Drude 1887, 1896, Warming 1895, 1909, Raunkiaer 1904, 1907) that use characters supposedly important to describe plant function, not necessarily physiognomic, yet strongly criticized by Du Rietz (1931). The objective of these systems, at least in theory, is not to express the physiognomy of the vegetation, but "simply to express the epharmonic component of the physiognomy" (Du Rietz 1931:14). The reasons are rooted in the evolutionary point of view that explains plant existence by the adaptations they developed through evolution.

The characters considered in the classification created by Warming (1884) are: duration of the whole plant, power of vegetative propagation, duration of tillers, hypogeous or epigeous type of shoots, mode of hibernation, and degree and mode of branching of rhizomes. The classification recognizes 14 groups without any special name. These groups are heterogeneous from the viewpoint of plant architecture, since important morphological features are disregarded (Barkman 1988a).

The classification proposed by Reiter (1885) resembles Grisebach's physiognomic classification, notwithstanding his stated preference for functional characters. The system recognizes plant classes, very much the same as Grisebach's "Vegetationsformen", based upon stem and leaf forms and named after a characteristic genus or family. However, on higher levels the classes are arranged differently, according to being photosynthetic or not, rooted or not, terrestrial or aquatic.

Drude (1887:489) emphasizes the importance of adaptations, especially to climate, and criticizes Humboldt's and Grisebach's physiognomic systems for mixing biological forms (based on functional characters) and morphological forms (based on nonfunctional characters). The scheme recognizes "Vegetationsformen", which is later modified (Drude 1896). However, the scheme is inconsistent in character criteria, for it recognizes groups, such as monocots and dicots. Pound and Clements (1898) adapt Drude's (1896) scheme to prairie vegetation, with plants classified in 7 main groups: (1) woody plants (subdivided into trees, shrubs, undershrubs, and climbers/twiners), (2) half shrubs, (3) perennial herbs (rosettes, mats, succulents, creepers and climbers, sod-formers, bunch grasses, rootstock plants, bulb and tuber plants, and ferns), (4) hapaxanthous herbs (biennials, annuals), (5) water plants (floating plants, submerged plants, amphibious plants), (6) hysterophytes (saprophytes, parasites), (7) thallophytes (mosses, liverworts, foliaceous lichens, fruticulose lichens, crustaceous lichens, geophilous fungi, xylophilous fungi, biophilous fungi, sathrophilous fungi, hydrophilous fungi, entomophilous fungi, filamentous algae, coenobioid algae). There is inconsistency in this classification, since it mixes plant architecture, habitat, and phylogeny as defining characters on the same level.

Raunkiaer's (1904, 1907) classification recognizes life-forms (or "biological types" in the 1904 paper) on the basis of plant adaptation to survive the unfavorable season. These forms are (1) phanerophytes (the survival buds or shoot-apices are borne on negatively geotropic shoots which projects into the air), (2) chamaephytes (the surviving buds or shoot-apices are borne on shoots very close to the ground), (3) hemicryptophytes (the surviving buds or shoot-apices are situated on the soil-surface), (4) cryptophytes (the surviving buds or shoot-apices are buried in the ground), (5) therophytes (plants of the summer or of the favorable season). Additional characters are used on the lower levels. Phanerophytes are subdivided into 15 subtypes by plant height (nanophanerophytes to megaphanerophytes), duration of leaves (evergreen, deciduous), presence of cover on buds, succulence and epiphytism. Chamaephytes are subdivided according to growth-form in suffruticose (erect), passive (erect until the unfavorable season, then bend to the ground due to own weight), active (stem lies on the ground), cushion plants (like the passive chamaephytes, but stems are arranged so close together that they prevent each other from falling over). Hemicryptophytes are subdivided on the basis of leaf distribution in proto-hemicryptophytes (only stem leaves), partial rosette plants (both stem and basal rosette leaves), rosette plants (only basal rosette leaves); these are further subdivided on the basis of presence of stolons. Cryptophytes are subdivided into geophytes (buds buried in the soil), helophytes (buds submersed in water, but vegetative shoots not), and hydrophytes (buds and vegetative shoots submersed in water). Geophytes are subdivided into rhizome, stem-tuber, root-tuber, bulb and root geophytes. Raunkiaer (1908) uses a reduced version of the classification to show that the same "biological spectrum" appears in similar climates in different regions. Raunkiaer's life-form classification, consistent and simple on the character criterion and terminology, has dominated the scene of vegetation studies since then (see Adamson 1939, Cain 1950, Gimingham 1951). Braun-Blanquet (1928:287) and latter Mueller-Dombois and Ellenberg (1974:449) introduce modifications and new subdivisions on the basis of growth-form, resulting in a classification not as simple as the original.

Warming (1909:6) arranges plants in main growth-form groups namely: heterotrophic, aquatic, muscoid, lichenoid, lianoid, and all other autonomous land plants (divided in monocarpic and polycarpic). On this level there is an obvious inconsistency on the character criterion (parasitism, habitat, phylogeny, plant architecture). The main groups are further subdivided. For instance, monocarpic plants are subdivided into aestival annuals, hibernal annuals, and biennial plants. Polycarpic plants are subdivided on the basis of characters which are ranked. Accordingly, the most important character is (1) duration of the vegetative shoot (lignified, herbaceous perennial, herbaceous deciduous); then come the others, (2) length and direction of internodes (short or long internodes, and if the latter are erect or prostrated/creeping), (3) position of the renewal buds during the unfavorable season (the Raunkiaer scheme). Less importance is attached to (4) structure of the renewal buds or of buds in general, (5) size of the plant, (6) duration of the leaves, (7) assimilation (photosynthesis) by leaves and/or by stems, and (8) capacity of "social life" (sociability).

Drude's classification (1913) abandons the unilateral view of using only purely functional, "epharmonic" characters and relies on the following main criteria: (1) the basic form (tree, shrub, annual or perennial), (2) form and duration of leaves, (3) protection of leaf and flower buds during dormancy, (4) position and structure of absorptive organs, and (5) single or recurrent reproduction. Based on this, 55 types are recognized.

Du Rietz (1931) also rejects functional characters as the only basis for plant form classification, but as he states (*loc. cit.* p. 43), to find an optimal classification system is a task for endless discussion:

> "While some authors will use the size, duration, and mode of branching of the stem as the main basis for the life-form system, others will be more inclined to put the size, form and duration of the leaves in the foreground, and still others will insist that the primary division should be based upon none of these points of view, but upon the way in which the plant survives the unfavorable season."

The only way out according to Du Rietz is to admit "not one single life-form system but several parallel ones, based upon different points of view". The scheme he proposes, thus, is unfolded in six classifications, namely: (1) main life-forms (based on the general physiognomy of the plants when completely developed), (2) growth-forms (based on shoot architecture), (3) periodicity life-forms (based on the seasonal physiognomical variation and the distribution of growth and resting periods), (4) bud height life-forms (Raunkiaer's life-forms), (5) bud type life-forms (based on the structure of the buds), and (6) leaf life-forms (based on form, size, duration, texture etc. of leaves). These classifications are arranged in parallel. If instead they were arranged in line or in a hierarchy, the scheme would be similar to others, such as Dansereau's (1951, 1957:118), Orlóci and Orlóci (1985), Halloy (1990), among others, in which the types are defined by the combination of the states of several characters.

Gimingham (1951) points out the limitation of Raunkiaer's life-form classification for expressing vegetation variation under slightly differing conditions. He creates a growth-forms classification, which is to be used on sand dunes vegetation, along two parallel criteria: form of the shoot system, and presence of a means of vegetative spread. Horikawa and Miyawaki (1954) propose two parallel growth-forms classifications for weeds according to the external, morphological features. In one case, the morphology of the shoots is considered (erect, prostrate, rosette, caespitose, and twining form); in the other, the morphology of roots is considered (fibrous, reticulate, straight, branched, and root stock form).

Barkman (1988a) describes a system of growth-forms that excludes characters related to life strategy, life cycle, hibernation level, and size, consistency and inclination of leaves, which according to him should be studied separately. The system is based on gross morphology, better known as plant "architecture" as it affects vegetation structure, "free of hypothesis of environmental adaptation." The growth-form types are named after characteristic plant taxa, such as Piceids, Quercids, etc. The main groups are errant (7 growth-forms), adnate (26 growth-forms) and radicant plants (55 growth-forms). The latter is subdivided into non self supporting plants (water plant, climbers), and self supporting plants (equisetoids, graminoids, herbae, dwarf shrubs, shrubs, trees). As he recognizes, the system is not always consistent on the defining characters, for in some groups some characters are more important for subdivisions than in other groups. For instance, graminoids are classified by height, leaf distribution and growth habit, while trees and shrubs are classified according to the way of branching. A classification of phenological types is also presented, based on vegetative periodicity (evergreen, semi evergreen, winter green, spring green, spring-summer green) and generative periodicity (time of flowering).

Dansereau (1951) is the first to propose a classification in which the types are defined by the combination of character states. The characters in his scheme are: (1) life-form (trees, shrubs, herbs, bryoids, epiphytes, and lianas), (2) size (tall, medium, and low), (3) function (deciduous, semideciduous, evergreen, evergreen-succulent, and evergreen-leafless), (4) leaf shape and size (needle or spine, graminoid, medium or small, broad, compound, and thalloid), (5) and leaf texture (filmy, membranous, sclerophyll, and succulent or fungoid). An estimate of coverage is treated as an additional character. Size is interpreted in conjunction with life-form; for instance, the minimum height for a tall herb is 2m, while it is 25 m for a tree. Not all potential types may actually appear in nature. Dansereau's short-hand symbols are intended to give a parsimonious structural characterization of the plant community. Similarly to Dansereau's, several other schemes for vegetation description have been proposed in which no preconceived types are recognized (see Knight 1965, Knight and Loucks 1969, Parsons 1976, Feoli and Scimone 1984, Orlóci and Orlóci 1985, Orlóci *et al.* 1986, Lausi and Nimis 1985, 1986, Gomez Sal *et al.* 1986, Grime, Hunt and Krzanowski 1987, Lausi, Nimis and Tretiach 1989, Halloy 1990, Guárdia and Ninot 1991, Cabido, Diaz and Acosta 1991). In each case, types are obtained by the combination of character states, or by clustering techniques, or no types are recognized. The characters are simple, or complex based on some of the schemes already described (life-forms, growth-forms etc.).

Knight (1965) presents a character set used to describe species in prairie relevés. The characters are life-form (*sensu* Raunkiaer), plant height, leaf characters (size as in Raunkiaer 1916, arrangement, pubescence, duration), root system, vegetative reproduction, overwintering organ, stem form, plant longevity, flowering season, rate of decay over winter, seed dispersal, grass leaf width, and grass growth-form. No recognition of plant types is sought, but rather a character description of relevés is obtained from the character description of the species (see Section 1.7). Knight and Loucks (1969) adopt a similar approach in Wisconsin forest communities. In this case the character set includes life-form, leaf characters (size, persistence), flowering season, pollinating mechanism, bark thickness, branching pattern, shade tolerance, moisture optimum, seed dispersal, surface fire susceptibility, twig diameter, and vegetative reproduction.

Parsons (1976) describes shrub species in communities of the Mediterranean climatic regions of California and Chile using life-form (*sensu* Raunkiaer), growth-form based on leaf width and duration, plant height, several leaf characters (size, periodicity, shape, margin, inclination, density, pubescence, stomata, odor, reflectivity), presence and location of spines, presence of fire sprouts, seed weight, seed dispersal, branching pattern, bark texture, pollination mechanism, genetical breeding system, reproductive periodicity, and vegetative periodicity. A reduced version of this scheme is applied by Feoli and Scimone (1984). Types are recognized by cluster analysis.

Noble and Slatyer (1980) propose a scheme of "vital attributes" to characterize species from the viewpoint of vegetation dynamics under disturbance. These characters are mainly (1) the method of arrival or persistence during and after a disturbance, (2) the ability to establish and grow to maturity in the developing community. For instance, in (1) ten patterns (character states) are recognized, in which the ones involving propagules are: (D) widely dispersed propagules such that they are available everywhere at any time that a disturbance occurs, (S) long lived propagules that persist viable in the soil at all life stages, (G) as in S, but either all germinate or are lost at the first disturbance, and (C) short lived propagules. In (2) three patterns are recognized: (T) can establish and grow immediately after disturbance and can continue establishing, (I) can establish and grow immediately after disturbance but cannot continue establishing as competition increases, and (R) cannot establish immediately after disturbance, but can do at later stages. Species types are defined by the combination of these 2 characters.

Box (1981) develops a model to predict and characterize the natural vegetation worldwide on the basis of climate. The classification recognizes 77 lifeforms. The defining characters are: (1) structural type (trees, small and dwarf trees, shrubs, etc., 19 states in total), (2) relative plant size, (3) leaf form (broad, narrow, graminoid, absent), (4) relative leaf size, (5) leaf surface structure (malacophyllous, coriaceous, sclerophyllous, succulent, photosynthetic stem, pubescent), (6) seasonal photosynthetic habit (evergreen, etc.).

Orlóci and Orlóci (1985), Orlóci *et al.* (1986), Orlóci and Stofella (1986), Orlóci and Orlóci (1990) describe communities on the basis of *character set types*. These are defined by the states of sets of complex and simple characters. For instance, the defining character set in Orlóci and Orlóci (1990), describing vegetation in the Chihuahua Desert, includes: (1) biological type (bryoid, lichen, pteridophyte, conifer, graminoid, cactoid, other angiosperm), (2) stem tissue type (succulent, herbaceous, woody), (3) stem function (structural support, structural support plus photosynthesis), (4) stem armature type (thorn, spine, prickle/hair), (5) stem arrangement (solitary, branched), (6) leaf longevity (drought deciduous, seasonally deciduous, withering, persistent), (7) leaf tissue type (succulent, herbaceous, fibroid), (8) leaf form (scale, filiform/needle, broad), (9) leaf arrangement (rosette absent, present), (10) leaf epidermal surface (glabrous, glaucous, trichomous), (11) leaf width (in 6 classes), (12) leaf length (in 5 classes), (13) leaf thickness (in 4 classes), (14) plant height (in 8 classes). An important feature in their scheme is the hierarchical, perfectly nested arrangement of the characters.

The character set used by Lausi and Nimis (1985, 1986) and Lausi, Nimis and Tretiach (1989) contains several macro and micromorphological features of leaves for the description of communities in the Canary Islands. The character set comprises: plant height, leaf (area, thickness, degree of dissection, single or in bundles, deciduousness, inclination, convolute or not), leaf pubescence (hair density, hair length, pluri or unicelular), cuticle (thickness, if smooth or rough), stomata (the number on upper and lower surface, in groves or not), epidermal cells (number and shape on upper and lower surfaces).

The character set adopted by Gomez Sal *et al.* (1986) to describe pastures in Central Spain includes characters associated with (1) reproduction, (2) occupation of space, and (3) adaptations to external factors of control or exploitation. The characters of the 1st type, which resemble the "vital attributes" of Noble and Slatyer (1980), are: Location of propagules (buried in the soil, dispersed, or kept attached to the flowering stem), number and relative size of seeds, and type of inflorescence. The characters of the 2nd type are: Density of spatial concentration (low, dense), plant growth (preferably erect, horizontal or creeping, both), and adult plant height. The ones of the 3rd type are: leaf abundance, leaf situation, degree of herbivore consumption, epidermal surface (pubescent, spiny, etc), phenology, stem texture, plant persistence, and leaf texture.

Halloy (1990) conceives a scheme to classify plants on the basis of complex characters. These are: plant silhouette, the "outline form of the plant as defined by the positions and arrangement of its apices" (11 states), leaf group, a sequence of leaves delimited by a stem without leaves (14 states), stem (27 states), root (5 states), and inflorescence (3 states). A plant type description is given by the combination of the states in all these characters or in some of them. The character states have names not always meaningful without an explanation, but keys and drawings are provided to determine the state to which a plant belongs in each one of the characters. As recognized by Halloy, the system does not aim at being complete; other characters, such as leaf size, shape, etc., may be important and could be included without changing the basic scheme.

1.7. Data analytical considerations

The objective of vegetation data analysis is to reveal and explain structures and structural connections. Plant community structures are described by characters. Among these, the intrinsic, organismic level characters constitute an analytical taxonomy, which describe the community components (see Section 1.2). The nature of the data available for analysis depends on the type, number and arrangement of these characters. Species-based descriptions, *i.e.*, those in which no other defining character but the species constitutes the analytical taxonomy, generate a matrix (k, ν) with k species populations and ν relevés. This can be subjected to analysis by the plethora of multivariate analytical techniques already available. The character-based schemes generate a similar k, ν data matrix, but then each population, usually a species, is described by, in the simplest case, one defining character with f states, which yields data for an f, k incidence matrix. Each entry in the k, ν matrix is a presence/absence score (1 or 0) or some other performance value. Each of the ν vectors is a vegetation relevé. The multiplication of matrices (f, k) and (k, ν), as Feoli and Scimone (1984) suggested, generates the matrix (f, ν). This matrix shows the frequency distribution of each character state among the relevés when the k, ν matrix contains presences/absences, or the sum of population performances in each character state and relevé when the k, ν matrix contains other performance values. Each row of the f, ν matrix is in fact a group of populations, defined as a separate taxon by the analytical taxonomy. Raunkiaer's (1908) "biological spectrum" is a typical example. He uses a reduced version of his classification of life-forms to show that a similar "biological spectrum", *i.e.*, the frequency distribution of life-forms, indicates similar climates in different regions (see also Cain *et al.* 1956). Analysis of a matrix (f, ν) can follow procedures already available for species groups by relevés matrices.

A descriptive scheme may have more than one defining character. Orlóci (1988a, 1988b, 1991) recognized sequential and hierarchical nested arrangements. A *sequential* scheme constitutes as many parallel analytical taxonomies as there are defining characters. The maximum number of taxa recognized by all these taxonomies, *i.e.*, the number of rows in the matrix (f, k), is $f = \sum_{h=1}^{m} s_h$, where m is the number of defining characters and s_h is the number of states in character h. To each character corresponds an s_h row submatrix in the matrix (f, k). The multiplication of matrices (f, k) and (k, ν) also generates a matrix (f, ν), but the rows in matrix (f, ν) that do not correspond to the same character (to the same analytical taxonomy) are not independent (Dale and Clifford 1976, Feoli and Scimone 1984). This is so because each row in matrix (f, ν) is actually a population group defined as a separate taxon by the analytical taxonomy corresponding to the character, and the same population performance is contributing simultaneously to m population groups (see Table 1.7.1). A typical example is given by Knight (1965) and Knight and Loucks (1969). In the latter, the matrix (f, ν) is subjected to ordination (method of Bray and Curtis 1957). A similar approach is applied by Gomez Sal *et al.* (1986), but the ordination method is correspondence analysis (Benzécri 1969). In the sequential scheme, the submatrices of the matrix (f, ν) can be analyzed as separate contingency tables, as in the example given by Feoli and Scimone (1984). In this case the relevés are arranged by environmental classes, and canonical contingency table analysis (Feoli and Orlóci 1979) is applied. The analysis can also use the decomposition of the expected frequencies (Feoli and Orlóci 1985) to isolate nonrandom trends of character variation.

Table 1.7.1. Example illustrating a sequential arrangement of characters. The two relevés are from the Caatinga (NE Brazil) and Chaco (NW Argentina) formations. For simplicity, only 5 species populations are considered in each relevé. The matrix (f, k) describes 10 species on the basis of 9 states in 4 defining characters. The matrix (k, ν) gives the species cover (%) in the relevés. The matrix (f, ν) is the product of the other two matrices. The submatrices corresponding to each character in matrix (f, ν) are not independent each other.

Matrix (f, k)												Matrix (f, ν)	
Character states		Species										Relevés	
		1	2	3	4	5	6	7	8	9	10	Caat	Cha
Stem tissue	- succulent	0	1	0	0	0	1	0	0	1	0	1	30
	- woody	1	0	1	1	1	0	1	1	0	1	46	48
Stem function	- regular	1	0	1	1	1	0	1	1	0	1	46	48
	- twin-purpose	0	1	0	0	0	1	0	0	1	0	1	30
Stem armature	- thorn/spine	0	1	0	0	1	1	1	0	1	1	16	46
	- none	1	0	1	1	0	0	0	1	0	0	31	32
Leaf duration	- deciduous	1	0	1	1	1	0	1	0	0	1	46	16
	- persistent	0	0	0	0	0	0	0	1	0	0	0	32
	- leafless	0	1	0	0	0	1	0	0	1	0	1	30

Matrix (k, ν)		
Species	Relevés	
	Caat	Cha
1	15	0
2	1	0
3	1	0
4	15	0
5	15	0
6	0	15
7	0	1
8	0	32
9	0	15
10	0	15

Instead of using the product matrix (f, ν), the analysis can proceed separately based on the matrices (f, k) and (k, ν). Dimensionality in each matrix can be reduced by applying ordination. Groups can be recognized by applying cluster analysis. These generate contingency tables, which may be interpreted in an ordination. The analytical taxonomy in this case has its population taxa defined by cluster analysis applied to the matrix (f, k). The following analytical schemes have been proposed:

(1) Population groups obtained from matrix (f, k) and population groups obtained from matrix (k, ν) define a contingency table. In this table each entry is

the frequency of populations that fall in the correspondent row and column classes. The degree with which one classification can be predicted by the other is then given by a measure of congruence, which may be mutual information (Feoli, Lagonegro and Orlóci 1984). Other methods for comparison of classifications may also apply (see Podani and Dickinson 1984, Podani 1986). An example given by Feoli and Scimone (1987) shows a very significant congruence between classifications of species based on characters and based on relevé composition.

(2) Population groups revealed by cluster analysis applied to the matrix (f, k), and ungrouped, or grouped relevés obtained from matrix (k, ν) or from environmental data define a contingency table. Each entry in the contingency table is the frequency or pooled performance of populations belonging to the same population group in a given relevé or group of relevés. The contingency table can be analyzed by techniques already available, such as canonical contingency table analysis (Feoli and Orlóci 1979), or can simply be visually interpreted as in Parsons (1976).

(3) Matrix (f, k) is subjected to ordination by characters and/or by populations (see *e.g.*, Karr and James 1975 studying bird communities, Lausi and Nimis 1986, Lausi, Nimis and Tretiach 1989). Also, groups of populations and groups of character states obtained from the same matrix define a contingency table. In this table each entry is the frequency of populations belonging to a given group of populations that present the character states belonging to a given group of character states. Analysis of the contingency table can proceed as in (2). The analysis reveal only trends at the population level. Examples are found in Feoli and Scimone (1984), Lausi and Nimis (1986) and Lausi, Nimis and Tretiach (1989).

In Lausi and Nimis (1986) and Lausi, Nimis and Tretiach (1989) the species are grouped based on the defining characters applying cluster analysis to the matrix (f, k). The role of the species groups in explaining vegetation structure in one community (Lausi and Nimis 1986) or along ecological gradients (Lausi, Nimis and Tretiach 1989) is then explored through ordination. In Lausi and Nimis (1986) only one relevé ($\nu = 1$) from high-elevation desert is in focus; the matrix (f, k) is subjected to cluster analysis and ordination after binary multiple coding, defining groups of species, groups of characters, and their configuration in ordination space. The ordination space shows a strong correlation between species position along the first axis and plant height, indicating a possible response to microclimatic changes. In Lausi, Nimis and Tretiach (1989) the data set includes several relevés along a transect in a *Myrica faya-Erica arborea* formation in the Canary Islands. As before, clustering and ordination define species groups, character groups, and their configuration in ordination space. Additionally, canonical contingency table analysis (Feoli and Orlóci 1979) reveals the connections between relevés classified by species composition and species classified by morphological characters. The data analysis in Gomez Sal *et al.* (1986) applies clustering and ordination to the f, k matrix to explore trends on the populational level, and ordination to the matrix product (f, ν) to explore trends in character variation among relevés. Grime, Hunt and Krzanowski (1987) applies cluster analysis to the f, k matrix to support the existence of three groups of plant strategies (competitive, stress tolerant, and ruderal or disturbance tolerant) characteristic of particular conditions. The clustering criterion is the minimization of the within group sum of squares. The matrix (f, k) contains 30 vegetative, reproductive and environmental defining characters.

A descriptive scheme with more than one defining character can also be *hierarchical nested* (Orlóci 1988a, 1988b, 1991, and Chapter 2). In the nested arrangement, Feoli's (1984) "conditional spaces", the taxon is a *character set type* (CST, Orlóci and Orlóci 1985), which "is a population whose individuals are identical with respect to the states of the chosen characters" (Orlóci 1991). In this case the number of potential taxa recognized by the analytical taxonomy is $f = \prod_{h=1}^{m} s_h$. The terms in the equation are already defined. For this reason the hierarchical nested arrangement potentially conveys much more information ($\log_2 f$) than the sequential arrangement with the same number of characters (Orlóci 1991). The method for analysis of hierarchical nested character-based community data (Orlóci and Orlóci 1985, Orlóci *et al.* 1986, Orlóci and Stofella 1986, Orlóci 1991) is reviewed and extended in Chapter 2.

Chapter 2

RESEMBLANCE MEASURES

The analysis of community data is comparative. It reveals data structures by placing the objects in an abstract space according to their resemblances. The objects in focus are plant communities that are compared on the basis of their CST records on different hierarchical levels. Many resemblance functions are available in the literature (see for instance Orlóci 1972, 1978). These can be adapted to character-based analysis. Of special interest are the functions that are partitionable into additive components, each corresponding to a level in the hierarchic nested model. The theory is discussed, partitions are derived, and examples are shown. Computations use the application program SYNCSA (see Chapter 6).

2.1. The hierarchical nested model of characters

Community components are identified as *character set types* (CSTs). These are defined by a sequence of states of plant attributes which can be arranged hierarchically (see Table 2.1.1 and Figure 2.1.1). By convention, the first attribute in the CST record is the top level of the hierarchy and the last attribute is the bottom level. Based on Orlóci (1991), the following characterizes the hierarchy:

1. Levels are numbered from the bottom up. On hierarchic level i there are k_i nodes, that is

$$k_i = \prod_{j=m}^{i} s_j = s_i \, k_{i+1} \tag{2.1}$$

where m is the number of levels, s_j is the number of states of the defining character on level j. Take k_{m+1} as 1. CSTs are mapped into the character hierarchy as runs through the nodes, each node corresponding to a state of its level's defining character. A node indicates the existence of a potential CST as a run from the top of the hierarchy to the level where the node is located. In this respect the terms "CST" and "node" are interchangeable, although the term "node" emphasizes that the model in focus is hierarchical nested. At the lowest level, CSTs are defined by the complete character set. On this level the number of nodes can be very large and not all of the nodes may correspond to an actual CST in nature. At the top level there are as many nodes as there are states of the 1st character in the list, *i.e.*, $k_m = s_m$.

2. Resemblance is computed between relevés from CST performance estimates, such as cover-abundance, local frequency, or density. As the CST number is reduced moving up in the hierarchy, their performance values are summed. The performance value at any node b, level i, is the sum of the performance values at the nodes connected to this node on the next lower level i-1, *i.e.*,

$$X_{ib\gamma} = \sum_{e=1+(b-1)s_{i-1}}^{b(s_{i-1})} X_{(i-1)e\gamma} \qquad (2.2)$$

The incremental limits of the subscripts are: i=2, ...,m (levels), b=1, ...,k_i (nodes), and γ=1, ...,ν (relevés). s_{i-1} is the number of states of the defining character on level i-1. The accumulation of CST performances is illustrated in Figure 2.1.1.

3. If an equal weight is to be given to all CSTs, CST performance estimates must be additive, *i.e.*, if two CSTs are joined analytically, the sum of their performance values should reflect a value that would actually be estimated in the field for the two populations if they were taken as one and the same. This is critical when cover-abundance classes are used. For instance, the scale proposed by van der Maarel (1979) to transform the Braun-Blanquet cover-abundance symbols into numerical values (1-9) gives more weight to scarcity than to dominance. The consequence of this on the hierarchical accumulation of performance values is complex. No effect is expected if scarce CSTs tend to join dominant ones, but assuming a random nesting, and noting that scarce components naturally tend to be more numerous, it is likely that the higher the hierarchical level the more weight is given to scarcity.

Table 2.1.1. Character score matrices of two relevés in the Caatinga (NE Brazil) and Chaco (NW Argentina) vegetation. For simplicity only four characters are used. These are extracted from a larger character set.

	Character Set Types									
	Caatinga				Chaco					
Characters	a	b	c	d	a	b	c	d	e	f
Stem tissue - **s**ucculent, **w**oody	w	s	w	w	w	s	w	w	w	w
Stem function - **r**egular, **t**win-purpose	r	t	r	r	r	t	r	r	t	t
Stem armature - **t**horn/spine, **n**one	n	t	t	n	n	t	t	n	t	n
Leaf duration - **d**eciduous, **p**ersistent, **l**eafless	d	l	d	p	d	l	d	p	d	d
Cover (%)*	6	2	1	1	2	3	3	3	2	1
	0		6			0	4	7		

---

* Digits in the numbers are vertically arranged.

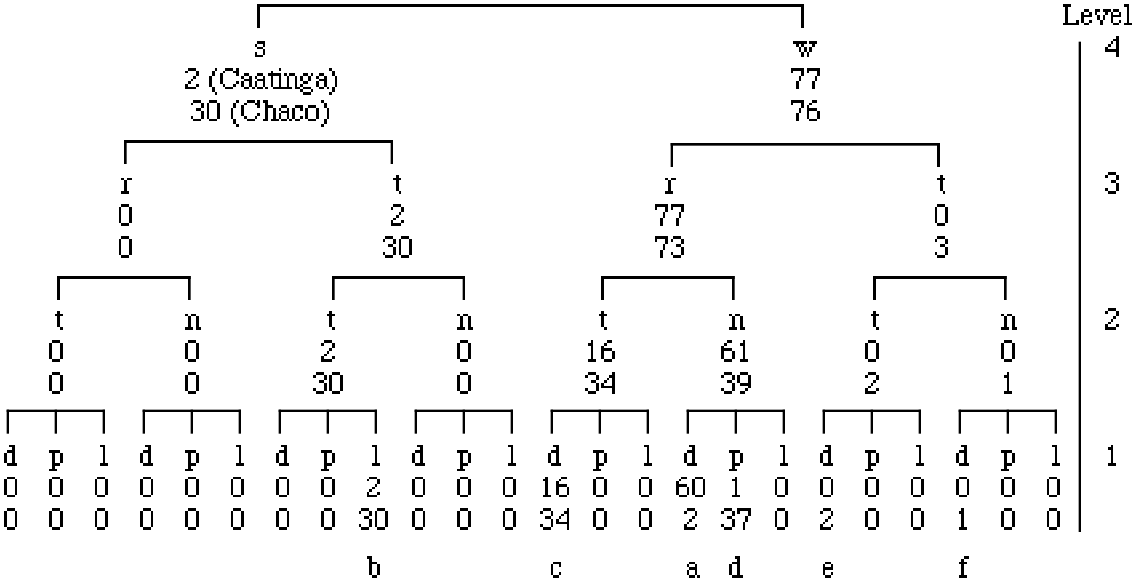


Figure 2.1.1. Hierarchical nested structure for the character set presented in Table 2.1.1. Letters at the nodes identify the states of the defining characters. The upper and lower rows of numbers indicate cumulative CST cover for Caatinga and Chaco. Letters along the bottom identify the materialized CSTs in the communities.

2.2. Hierarchical partitioning of resemblance

In order to take advantage of the hierarchically nested arrangement of characters, community level resemblance measures based on characters should be designed in such a way that an additive analysis is achieved. Under this condition the total resemblance, due to all characters, is partitioned into as many additive components as there are levels in the character hierarchy. Nested partitions are used in pattern analysis (Greig-Smith 1952, Orlóci 1971, Feoli, Lagonegro and Orlóci 1984:19) involving a hierarchical nesting of samples and not taxa as in this case. Nevertheless, the basic principles can be adopted. A partition of cross products has been the basis of the analysis in Orlóci and Orlóci (1985), Orlóci and Kenkel (1985:149) and Orlóci *et al.* (1986). We investigate additional resemblance measures.

In general terms, if a resemblance function is partitionable, the *nominal* resemblance value $D_{h\alpha\beta}$ between relevé α and β, hierarchical level h, can be conceived as a sum of m-h+1 partial resemblances.

$$D_{h\alpha\beta} = \sum_{i=h}^{m} \Delta_{i\alpha\beta} \tag{2.3}$$

where

$$\Delta_{i\alpha\beta} = D_{i\alpha\beta} - D_{(i+1)\alpha\beta}. \tag{2.4}$$

The Δ values are *partial* resemblances corresponding to levels h to m. At i=m (on the highest level), we take $D_{(i+1)\alpha\beta}$ as a zero value. $D_{1\alpha\beta}$, the value on level 1, is the *global* resemblance of relevé α and β. At a given level, the partial resemblance represents the contribution of the defining character to the nominal resemblance of relevé α and β on that level. The partial resemblance reflects what is unique in the nominal resemblance on level i. If partial resemblances of subsequent levels i, i+1, ..., i+t are summed, the pooled value reflects what is added to the nominal resemblance α and β on level i+t+1. The effect of character order is discussed in chapter 4, but here it should be mentioned that even though the partial resemblance on level i indicates the contribution of character i, this contribution is order dependent within the character set. That is, the same character positioned on a different level may define different nominal and partial resemblance. By the same token, different characters placed on the same level, other than level 1, will define potentially different nominal and partial resemblances. On level 1, all characters define the same global resemblance.

Another consequence of a resemblance function being partitionable among levels is that its partial components are within the same range defined for the nominal resemblance. For instance, a dissimilarity measure assuming nominal values larger than zero also implies that $D_{i\alpha\beta} \geq D_{(i+1)\alpha\beta}$.

The question is, given the cumulative CST performances in the hierarchical nested model, how to obtain resemblance measures that are hierarchically partitionable. In this respect, the cumulative CST performance (Eq. 2.2) at a given node b on level i, relevé γ, is viewed as a sum expanded in n_i terms, each of size $\frac{X_{ib\gamma}}{n_i}$. Important to note that n_i is the number of first level nodes fused at a node (any node) on level i. On level 1 $n_1 = 1$, and on level i

$$n_i = \prod_{j=2}^{i} s_{j-1} = s_{i-1} n_{i-1} = \frac{k_1}{k_i} \qquad (2.5)$$

In this equation s_j represents the number of states of the defining character on level j; k_1 and k_i are the numbers of nodes on levels 1 and i (Eq. 2.1). Regarding an example, the performance values on level 3 in Figure 2.1.1 are expanded in groups of $n_3 = 6$ terms:

$$\text{Caatinga:}\quad 0\ 0\ 0\ 0\ 0\ 0\ \frac{2}{6}\ \frac{2}{6}\ \frac{2}{6}\ \frac{2}{6}\ \frac{2}{6}\ \frac{2}{6}\ \frac{77}{6}\ \frac{77}{6}\ \frac{77}{6}\ \frac{77}{6}\ \frac{77}{6}\ \frac{77}{6}\ 0\ 0\ 0\ 0\ 0\ 0$$

$$\text{Chaco:}\quad 0\ 0\ 0\ 0\ 0\ 0\ \frac{30}{6}\ \frac{30}{6}\ \frac{30}{6}\ \frac{30}{6}\ \frac{30}{6}\ \frac{30}{6}\ \frac{73}{6}\ \frac{73}{6}\ \frac{73}{6}\ \frac{73}{6}\ \frac{73}{6}\ \frac{73}{6}\ \frac{3}{6}\ \frac{3}{6}\ \frac{3}{6}\ \frac{3}{6}\ \frac{3}{6}\ \frac{3}{6}$$

The resemblances computed from these expanded values are partitionable among the levels, provided that a suitable resemblance function is used. We show the *modus operandi* in the sequel.

2.3. Effect of data transformations on hierarchical accumulation

Some resemblance functions allow the use of transformed data or, as often, may even perform implicit transformations. When a hierarchical nested model is considered, some transformations may affect the accumulation of performance values, and in turn affect the resemblance function's capacity of being partitioned among hierarchical levels. If Eq. 2.2 is violated by the transformation, the resemblance function will not be partitionable additively. However, data transformations limited to the lowest hierarchical level, before accumulation, will have no effect on resemblance partitioning. Scalar transformations which treat equally all values on all hierarchical levels do not have any effect either.

Vector transformations applied to CST performance values already hierarchically accumulated may affect the partitioning properties. Vector transformations are based on relevé marginal totals (of CST performances within each relevé), CST marginal totals (of CST performances within each CST), or both. Since all transformations using CST marginal totals affect partitioning, only typical transformations using the relevé marginal totals are examined below. In these, the transformation is applied to the CST cumulants $X_{ib\gamma}$ for i=1, ..., m levels, b=1, ..., k_i nodes on level i (Eq. 2.1), and relevés γ=1, ...,ν.

a) Centering has no effect on hierarchical accumulation. To show this, we begin with the centered cumulant on level i:

$$T_{ib\gamma} = n_i \left(\frac{X_{ib\gamma}}{n_i} - \frac{X_{i.\gamma}}{k_1} \right) = X_{ib\gamma} - \frac{n_i X_{i.\gamma}}{k_1} \tag{2.6}$$

In the above equation, $X_{i.\gamma}$ is the total of all CST performances in relevé γ, which is the same for all levels. All symbols have been defined. Using Eq. 2.2,

$$\sum_{e=1+(b-1)s_{i-1}}^{b(s_{i-1})} \left(X_{(i-1)e\gamma} - \frac{n_{(i-1)} X_{(i-1).\gamma}}{k_1} \right) =$$

$$= \left(\sum_{e=1+(b-1)s_{i-1}}^{b(s_{i-1})} X_{(i-1)e\gamma} \right) - s_{i-1} \frac{n_{(i-1)} X_{(i-1).\gamma}}{k_1} =$$

$$= X_{ib\gamma} - \frac{n_i X_{i.\gamma}}{k_1}$$

Note that $s_{i-1} n_{i-1} = n_i$ (Eq. 2.5). Also $X_{(i-1).\gamma} = X_{i.\gamma}$, since the grand total per level does not change. Clearly, centering does not violate Eq. 2.2. b) Adjustment of relevé vectors to unit sum $\frac{X_{ib\gamma}}{X_{i.\gamma}}$ (Whittaker 1952) does not affect hierarchical accumulation. The adjusted cumulant on level i is

$$U_{ib\gamma} = n_i \frac{\dfrac{X_{ib\gamma}}{n_i}}{X_{i.\gamma}} = \frac{X_{ib\gamma}}{X_{i.\gamma}} \tag{2.7}$$

The terms in this equation have already been defined. Since the denominator is unchanged through the levels, this transformation does not violate Eq. 2.2.c) Adjustment of relevé vectors to unit length (normalization) affects hierarchical additivity. The transformation is defined as

$$V_{ib\gamma} = n_i \frac{\dfrac{X_{ib\gamma}}{n_i}}{\left(\sum_{d=1}^{k_i} n_i \left(\dfrac{X_{id\gamma}}{n_i}\right)^2\right)^{1/2}} = X_{ib\gamma} \left(\frac{n_i}{\sum_{d=1}^{k_i} X_{id\gamma}^2}\right)^{1/2}$$

which is different from what is obtained using Eq. 2.2:

$$\sum_{e=1+(b-1)s_{i-1}}^{b(s_{i-1})} X_{(i-1)e\gamma} \left(\frac{n_{i-1}}{\sum_{g=1}^{k_{i-1}} X_{(i-1)g\gamma}^2}\right)^{1/2} =$$

$$= X_{ib\gamma} \left(\frac{n_{i-1}}{\sum_{g=1}^{k_{i-1}} X_{(i-1)g\gamma}^2}\right)^{1/2} \neq X_{ib\gamma} \left(\frac{n_i}{\sum_{d=1}^{k_i} X_{id\gamma}^2}\right)^{1/2} \tag{2.8}$$

Eq. 2.2 is violated.

d) Adjustments implicit in a correlation coefficient affect hierarchical accumulation. Correlation between relevés can be defined as the product $\mathbf{W}_i'\mathbf{W}_i$ (see Orlóci 1967b) where

$$W_{ib\gamma} = n_i \frac{\dfrac{X_{ib\gamma}}{n_i} - \dfrac{X_{i.\gamma}}{k_1}}{\left(\sum_{d=1}^{k_i} n_i \left(\dfrac{X_{id\gamma}}{n_i} - \dfrac{X_{i.\gamma}}{k_1}\right)^2\right)^{1/2}}$$

From Eq. 2.6 it follows that

$$W_{ib\gamma} = n_i \frac{\frac{T_{ib\gamma}}{n_i}}{\left(\sum_{d=1}^{k_i} n_i \left(\frac{T_{ib\gamma}}{n_i}\right)^2\right)^{1/2}} = T_{ib\gamma} \left(\frac{n_i}{\sum_{d=1}^{k_i} T_{ib\gamma}^2}\right)^{1/2} \tag{2.9}$$

It can be seen that this transformation imply in both centering and normalization. Since the latter violates Eq. 2.2, the correlation coefficient is not partitionable.

The fact that not all resemblance measures are partitionable into additive partial components does not preclude any of them to be applied to the hierarchical nested model to define nominal resemblances.

2.4. Euclidean spatial parameters

2.4.1. Cross product

Resemblances are computed based on the k_i cumulants $X_{ib\alpha}$ and $X_{ib\beta}$ for i=1, ..., m levels, and α, β=1, ...,ν relevés. The resulting resemblance matrix is symmetric; thus only $\nu(\nu-1)$ values need to be calculated on each level. The cross product of relevé α and β on the ith level is:

$$q_{i\alpha\beta} = \sum_{b=1}^{k_i} n_i \frac{X_{ib\alpha}}{n_i} \frac{X_{ib\beta}}{n_i} = \sum_{b=1}^{k_i} \frac{X_{ib\alpha} X_{ib\beta}}{n_i} \tag{2.10}$$

As before, k_i represents the number of nodes on level i (Eq. 2.1), and n_i the number of level 1 CSTs fused at any node on level i (Eq. 2.5).

In order to calculate product moment correlations, a centering transformation is applied (Orlóci *et al.* 1986). The centered cross product equation is :

$$q_{i\alpha\beta} = \sum_{b=1}^{k_i} \left[n_i \left(\frac{X_{ib\alpha}}{n_i} - \bar{X}_{1.\alpha} \right) \left(\frac{X_{ib\beta}}{n_i} - \bar{X}_{1.\beta} \right) \right] \tag{2.11}$$

The term $\bar{X}_{1.\gamma}$ represents the mean performance of the CSTs on the first level in $\gamma = \alpha$ or β, that is

$$\bar{X}_{1.\gamma} = \frac{X_{1.\gamma}}{k_1}$$

An equivalent equation is

$$q_{i\alpha\beta} = \sum_{b=1}^{k_i} \frac{X_{ib\alpha} X_{ib\beta}}{n_i} - \frac{X_{1.\alpha}\, X_{1.\beta}}{k_1} \tag{2.12}$$

which allows one to use only the non-zero nodes, reducing computation time in large hierarchies. The first term in the subtraction is a non-centered cross product; the second one is a constant term of correction. The Pearson product moment correlation coefficient is defined in centered cross product terms by

$$r_{i\alpha\beta} = \frac{q_{i\alpha\beta}}{\sqrt{q_{i\alpha\alpha}\, q_{i\beta\beta}}} \tag{2.13}$$

As already verified, this value is not partitionable in hierarchical level components. The partial cross product given by $\theta_{i\alpha\beta} = q_{i\alpha\beta} - q_{(i+1)\alpha\beta}$ is partitionable. Therefore, to calculate partial correlations, the partial cross products are used:

$$r_{i\alpha\beta} = \frac{\theta_{i\alpha\beta}}{\sqrt{\theta_{i\alpha\alpha}\, \theta_{i\beta\beta}}} \tag{2.14}$$

If uncentered cross products are used in equations 2.13 and 2.14, the cosine of the angle subtending to the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in a k_1-dimensional space is computed. Some cross products and product moment correlations are given for the Caatinga and Chaco relevés in Table 2.4.1.1.

2.4.2. Euclidean distance

General properties are discussed in Orlóci (1967a, 1978). For the hierarchical nested model, a squared euclidean distance is computed between relevé α and β on level i, according to

$$d^2_{i\alpha\beta} = \sum_{b=1}^{k_i} n_i \left(\frac{X_{ib\alpha}}{n_i} - \frac{X_{ib\beta}}{n_i}\right)^2 = \sum_{b=1}^{k_i} \frac{(X_{ib\alpha} - X_{ib\beta})^2}{n_i} \tag{2.15}$$

In this equation, $X_{ib\alpha}$ and $X_{ib\beta}$ are the bth node CST performances, level i, which may be transformed according to criteria already mentioned. For instance, an adjustment to unit sum of elements in the relevé vectors (Orlóci 1978:95) is admissible:

$$d^2_{i\alpha\beta} = \sum_{b=1}^{k_i} \frac{\left(\frac{X_{ib\alpha}}{X_{i.\alpha}} - \frac{X_{ib\beta}}{X_{i.\beta}}\right)^2}{n_i} \qquad (2.16)$$

This places the values of the function within the range $0 \leq d^2_{i\alpha\beta} \leq 2$ (see example in Table 2.4.2.1). Squared euclidean distances can also be calculated from the cross products in Eqs. 2.10 or 2.12:

$$d^2_{i\alpha\beta} = q_{i\alpha\alpha} + q_{i\beta\beta} - 2\, q_{i\alpha\beta} \qquad (2.17)$$

The chord distance implicitly includes normalization within relevés, and therefore it cannot be partitioned (see Eq. 2.8). However, chord distances can be calculated from uncentered cross products (Eq. 2.10) using:

$$c^2_{i\alpha\beta} = 2\left(1 - \frac{q_{i\alpha\beta}}{\sqrt{q_{i\alpha\alpha}\, q_{i\beta\beta}}}\right) \qquad (2.18)$$

In this equation $c^2_{i\alpha\beta}$ is the nominal squared chord distance and the nominal products are used; to obtain the partial squared chord distances, the partial products have to be used.

Table 2.4.1.1. Nominal and partial cross products and correlations (r) calculated for the data set in Table 2.1.1. Uncentered and centered cross products are presented. Correlations are calculated from the centered products. Note the correspondence of the nominal resemblances level 1 and the total (global) partial resemblances.

Nominal resemblances:

	Uncentered products			Centered products			r
Level	$q_{\text{Caa Chac}}$	q_{Caa}	q_{Chaco}	$q_{\text{Caa Chac}}$	q_{Caa}	q_{Chaco}	Caa Chaco
4	492.7	494.4	556.3	143.8	234.4	88.17	1.000
3	946.8	988.8	1039.7	597.9	728.8	571.5	.9265
2	994.3	1327.	1194.	645.4	1067.	725.8	.7334
1	761.0	3861.	3434.	412.1	3601.	2966.	.1261

Partial resemblances:

	Uncentered products			Centered products			r
Level	$q_{\text{Caa Chac}}$	q_{Caa}	q_{Chaco}	$q_{\text{Caa Chac}}$	q_{Caa}	q_{Chaco}	Caa Chaco
4	492.7	494.42	556.33	143.8	234.4	88.17	1.000
3	454.2	494.42	483.33	454.2	494.4	483.3	.9291
2	47.50	338.17	154.33	47.50	338.2	154.3	.2079
1	-233.3	2534.	2240.	-233.3	2534.	2240.	-.0979
Total	761.0	3861.	3434.	412.1	3601.	2966.	.1261

Table 2.4.2.1. Nominal and partial squared euclidean distances for the Caatinga and Chaco relevés described in Table 2.1.1. The data were subjected to adjustments to unit sum and unit length of relevé vectors. The latter generate squared chord distances, in which case the partial and global (total) resemblances are obtained from uncentered partial products (Table 2.4.1.1).

	Nominal resemblance			Partial resemblance		
Level	Raw data	Unit sum	Unit length	Raw data	Unit sum	Unit length
4	65.42	.01107	.1212	65.42	.01107	.1212
3	134.8	.02483	.1324	69.41	.01376	.1419
2	532.3	.08143	.4201	397.5	.05660	1.584
1	5773.	.7424	1.582	5241.	.6609	2.196
			Total	5773.	.7424	1.582

2.5. Absolute value function

The absolute value function (Orlóci 1978:55) for relevés α, β hierarchical level i, is defined by

$$a_{i\alpha\beta} = \sum_{b=1}^{k_i} n_i \left| \frac{X_{ib\alpha}}{n_i} - \frac{X_{ib\beta}}{n_i} \right| = \sum_{b=1}^{k_i} |X_{ib\alpha} - X_{ib\beta}| \qquad (2.19)$$

The symbols used in this equation have been defined in the preceding sections. $X_{ib\alpha}$ and $X_{ib\beta}$ may be transformed according to criteria already mentioned. If an adjustment of the relevé vectors to unit sum is applied, a relativized absolute value function is obtained:

$$w_{i\alpha\beta} = \sum_{b=1}^{k_i} \left| \frac{X_{ib\alpha}}{X_{1.\alpha}} - \frac{X_{ib\beta}}{X_{1.\beta}} \right| \qquad (2.20)$$

which assumes values in the range 0 and 2. An example is given in Table 2.5.1.

Table 2.5.1. Nominal and partial absolute value function calculated for the Caatinga and Chaco based on data in Table 2.1.1. Data adjustment to unit sum of relevé vectors is also applied.

	Nominal resemblance		Partial resemblance	
Level	Raw data	Unit sum	Raw data	Unit sum
4	29	.5154	29	.5154
3	35	.5720	6	.05650
2	71	.8085	36	.2366
1	143	1.481	72	.6727
		Total	143	1.481

2.6. Information theoretical measures

Similarity and dissimilarity measures based on information are described in Orlóci (1978:78) and Feoli, Lagonegro and Orlóci (1984:26), and used in Feoli, Orlóci and Scimone (1985). We consider here different formulations. The bases of these are information divergences on each hierarchical level from a diversity array with k_i rows (CSTs) and 2 columns (relevés α and β). The information divergence within rows is a measure of dissimilarity between relevés α and β, that is

$$I(X_{ib\gamma};\overline{X}_{ib.}) = n_i \sum_{b=1}^{k_i} \sum_{\substack{\gamma=\alpha \\ X_{ib\gamma}>0}}^{\beta} \frac{X_{ib\gamma}}{n_i} \ln \frac{\dfrac{X_{ib\gamma}}{n_i}}{\dfrac{\dfrac{X_{ib.}}{2}}{n_i}} =$$

$$= \sum_{b=1}^{k_i} \sum_{\substack{\gamma=\alpha \\ X_{ib\gamma}>0}}^{\beta} X_{ib\gamma} \ln \frac{2X_{ib\gamma}}{X_{ib.}} \qquad (2.21)$$

for any i=1, ..., m levels and α, β=1, ..., ν relevés. $X_{ib\gamma}$ represents a nontransformed CST performance at node b, level i, in relevé $\gamma = \alpha$ or β. Only $X_{ib\gamma}>0$ are included. $X_{ib.}$ is the sum of CST performances in α and β at node b, level i. Another information measure is based on the mutual information in the k_i x 2 array, which happens to measure dissimilarity between α and β:

$$I_{i(\text{CSTs;relevés})} = n_i \sum_{b=1}^{k_i} \sum_{\substack{\gamma=\alpha \\ X_{ib\gamma}>0}}^{\beta} \frac{X_{ib\gamma}}{n_i} \ln \frac{\dfrac{X_{ib\gamma}}{n_i}}{\dfrac{\dfrac{X_{i.\gamma}X_{ib.}}{X_{i..}}}{n_i}} =$$

$$= \sum_{b=1}^{k_i} \sum_{\substack{\gamma=\alpha \\ X_{ib\gamma}>0}}^{\beta} X_{ib\gamma} \ln \frac{X_{ib\gamma}\, X_{i..}}{X_{i.\gamma}\, X_{ib.}} \qquad (2.22)$$

In this equation $X_{i.\gamma}$ is the CST performance in relevé $\gamma = \alpha$ or β summed over all nodes, level i, and $X_{i..}$ is the sum of the CST performances in α and β. The quantity defined in Eq. 2.21 can be decomposed as

$$I(X_{ib\gamma};\overline{X}_{ib.}) = I_{i(\text{CST;relevés})} + I(X_{i.\gamma};\overline{X}_{i..})$$

where

$$I(X_{i.\gamma};\overline{X}_{i..}) = \sum_{\gamma=\alpha}^{\beta} X_{i.\gamma} \ln \frac{2X_{i.\gamma}}{X_{i..}} \tag{2.23}$$

is an information divergence based on the column totals in the k_i x 2 diversity array. This is the same on all hierarchical levels, which is easily verified. In fact, the totals in each relevé ($X_{i.\gamma}$) and the grand total ($X_{i..}$) are constant through the hierarchy. For two communities that differ in the values of $X_{i.\gamma}$ but have the k_i CSTs in the same proportion, the divergence defined in Eq. 2.22 is zero. Therefore, Eq. 2.21 is the least informative divergence measure (Orlóci 1978:78), since it may reflect mere differences in the $X_{i.\gamma}$ values. An example is presented in Table 2.6.1.

In order to be consistent with the fact that the diversity arrays are k_i x 2 contingency tables, the resemblance measures based on information seem only appropriate when CST performances represent counts or frequencies. However, individuals are not always counted. Interestingly, to Pielou (1966) the number of individuals is an amount in whatever units (*e.g.* percentage cover, kilograms) have been chosen for measurement.

Table 2.6.1. Nominal and partial information quantities calculated for the Caatinga and Chaco. Comparison is based on data presented in Table 2.1.1. The information divergences of Eq. 2.21 and mutual information of Eq. 2.22 are given. The difference between the nominal measures on each level is constant and corresponds to the information divergence based on the relevé marginal totals (Eq. 2.23).

	Nominal		Partial	
Level	Information divergence	Mutual information	Information divergence	Mutual information
4	14.70	12.73	14.70	12.73
3	16.83	14.85	2.129	2.130
2	22.53	20.56	5.701	5.701
1	75.95	73.97	53.42	53.42
		Total	75.95	73.97

2.7. Partitioning resemblances between qualitative and quantitative data components

The CST performances on each level can be partitioned into purely qualitative and purely quantitative components following Williams and Dale (1962) and Orlóci *et al.* (1986). The quantitative component exists only as a mathematical abstraction. The isolation of these components is useful to reveal the sources of the similarity or dissimilarity. The k_1 raw CST performances ($X_{1b\gamma}$), level 1, relevé γ, are denoted by the vector $\mathbf{X}$ with k_1 elements. This vector is the sum of a qualitative data vector $\mathbf{M}$, in which the non-zero $X_{1b\gamma}$ values are replaced by their mean M, and a quantitative data vector $\mathbf{D}$, in which the non-zero $X_{1b\gamma}$ values are replaced by their deviations from the mean M; the zero $X_{1b\gamma}$ values remain unchanged. An example is given in Table 2.7.1. Note that, in order to retain hierarchical additivity (Eq. 2.2), data partitioning is performed only on level 1, before hierarchical accumulation of the values.

Table 2.7.1. Data partitioning into mixed, purely qualitative and purely quantitative components for the Caatinga and Chaco example (Table 2.1.1), on hierarchical level 1. The method is described in the main text.

CSTs		Data type					
		Mixed (**X**)		Qualitative (**M**)		Quantitative (**D**)	
#	Level 1	Caatinga	Chaco	Caatinga	Chaco	Caatinga	Chaco
a	w r n d	60	2	19.75	17.67	40.25	-15.67
b	s t t l	2	30	19.75	17.67	-17.75	12.33
c	w r t d	16	34	19.75	17.67	-3.75	16.33
d	w r n p	1	37	19.75	17.67	-18.75	19.33
e	w t t d	0	2	0	17.67	0	-15.67
f	w t n d	0	1	0	17.67	0	-16.67

Resemblances can be computed using the mixed data, or the qualitative or quantitative data components. The quantitative and qualitative resemblance components, plus their interaction, sum to the resemblance value computed with the mixed data. For instance, the uncentered cross product of relevés α and β, level i, using $\mathbf{X}_\alpha$ and $\mathbf{X}_\beta$ can be decomposed:

$$q_{i\alpha\beta|X} = \frac{1}{n_i}\sum_{b=1}^{k_i} X_{ib\alpha}\, X_{ib\beta} =$$

$$= \frac{1}{n_i}\sum_{b=1}^{k_i} (M_{ib\alpha} + D_{ib\alpha})(M_{ib\beta} + D_{ib\beta}) =$$

$$= \frac{1}{n_i}\sum_{b=1}^{k_i} [(M_{ib\alpha}M_{ib\beta}) + (D_{ib\alpha}D_{ib\beta}) + (M_{ib\alpha}\, D_{ib\beta}) + (D_{ib\alpha}\, M_{ib\beta})] =$$

$$= q_{\alpha\beta|M} + q_{\alpha\beta|D} + q_{\alpha|M,\beta|D} + q_{\alpha|D,\beta|M}$$

In this equation $q_{\alpha\beta|M}$ and $q_{\alpha\beta|D}$ are the qualitative and quantitative components respectively. The other two components are interactions. Centering after data type decomposition do not change the relation, since $\sum_{b=1}^{k_i} M_{ib\gamma} = \sum_{b=1}^{k_i} X_{ib\gamma}$ and $\sum_{b=1}^{k_i} D_{ib\gamma} = 0$. An example is shown in Table 2.7.2.

The squared euclidean distance is similarly partitioned:

$$d^2_{i\alpha\beta|X} = \frac{1}{n_i}\sum_{b=1}^{k_i} (X_{ib\alpha} - X_{ib\beta})^2 =$$

$$= \frac{1}{n_i}\sum_{b=1}^{k_i} [(M_{ib\alpha} + D_{ib\alpha}) - (M_{ib\beta} + D_{ib\beta})]^2 =$$

$$= \frac{1}{n_i}\sum_{b=1}^{k_i} [(M_{ib\alpha} - M_{ib\beta}) + (D_{ib\alpha} - D_{ib\beta})]^2 =$$

$$= \frac{1}{n_i}\sum_{b=1}^{k_i} [(M_{ib\alpha}-M_{ib\beta})^2 + (D_{ib\alpha}-D_{ib\beta})^2 + 2(M_{ib\alpha}-M_{ib\beta})(D_{ib\alpha}-D_{ib\beta})]$$

Hence,

$$d^2_{i\alpha\beta|X} = d^2_{i\alpha\beta|M} + d^2_{i\alpha\beta|D} + \frac{1}{n_i}\sum_{b=1}^{k_i} 2\,(M_{ib\alpha}-M_{ib\beta})\,(D_{ib\alpha}-D_{ib\beta})$$

The partition of other euclidean spatial parameters can be derived. Information measures are not partitionable between data types, because the negative values in the quantitative component cannot be handled.

Table 2.7.2. Centered cross products and correlation coefficients between Caatinga and Chaco for mixed, qualitative and quantitative data types. Computations use data presented in Table 2.7.1. Only nominal resemblances are shown. The method is described in the main text.

	Centered cross products			Correlation coefficients		
Level	Mixed (**X**)	Qualitative (**M**)	Quantitative (**D**)	Mixed (**X**)	Qualitative (**M**)	Quantitative (**D**)
4	143.8	116.3	-36.49	1.000	1.000	-1.000
3	597.9	232.6	22.68	.9265	0.7303	0.1356
2	645.4	348.9	-67.11	.7334	0.8018	-0.2315
1	412.1	1047.	-1273.	.1261	0.7746	-0.6718

2.8. Probabilistic indices of resemblance

2.8.1. The problem

The comparison of relevés, using the resemblance functions presented so far, do not reflect how common or uncommon within the data set is the resemblance value obtained. In this respect, completely different data sets may yield the same resemblance value. For instance, suppose that in relevés A and B the shared CSTs are very rare in the set, while in relevés C and D the shared CSTs are very common in the set. Although the nominal dissimilarity may be the same in both cases, in probabilistic terms one would regard relevés A and B closer than C and D. An analogous example is given by Goodall (1966), who proposes a similarity index based on probability. Goodall's index (see Goodall 1966, Goodall, Ganis and Feoli 1987) is applicable to pairs of individuals, relevés in our case, described on the basis of attributes. In a given data set, in each attribute the pairs of individuals are ordered according to their differences. Based on this ordering, a probability is obtained for each pair in each attribute. Goodall's method uses the product of the logarithm of the probabilities over the attributes, which, assuming random sampling and independent attributes, *i.e.*, a homogeneous population, has a chi-square probability distribution. The similarity index is the one complement of the probability of a more extreme chi-squared value. An exact method of combining probabilities is also described by Goodall (1966), but it was computationally impractical when it was proposed.

We note that the similarity index by the chi-squared method relies on underlying assumptions intrinsic to an axiomatic distribution, which are rarely met in ecological data (Orlóci 1990) and are far from being universal (see *e.g.*, Bradley 1968:6, Edgington 1987:6). In the analytical context of the hierarchical nested model, we develop an alternative method. The method uses *randomization tests* to generate the probabilities, and thus, it is free from assumptions about the underlying distribution. The probability of a resemblance value indicates how rare is such a value if the stated null hypothesis were true. This probability can be interpreted as a dissimilarity, and used as such, or can be transformed into a similarity by taking the one complement as in Goodall's index. Any of the resemblance measures shown in the previous sections can have a probability attached by this method. Randomization tests have been described in other contexts, which emphasized the analysis of experimental data (see Fisher 1951:43, Kempthorne 1952:120, 1955, Bradley 1968, Edgington 1969a, 1969b, 1987). Pitman (1937) developed a randomization test for the product moment correlation coefficient. Randomization techniques have been applied in the analysis of taxonomic data (Rohlf 1965) and community data (Orlóci and Beshir 1976, Orlóci and Kenkel 1985:80, Orlóci *et al.* 1986, Podani 1991), which differ from the approach here adopted.

2.8.2. The randomization method

The basis of the method of randomization was first formulated by Fisher (see Fisher 1951:43, Bradley 1968:71). In a data set we can distinguish between (a) the magnitudes (or states in any scale) of the *observations*, and (b) their *locations* in the data table. Any observed data set can thus be interpreted as being one possibility of many different permutations of the same observations among the locations. The different permutations constitute the *reference set* (Hope 1968, Edgington 1987:305). The reference set is a family of data sets delimited by keeping the observations constant and systematically changing their location in ways that agree with the null hypothesis and other restrictions (*e.g.*, sample sizes). Each data set member of the reference set is just as likely as the observed data set to be obtained were the null hypothesis true. The commonness of a given statistic computed in the observed data set is assessed by comparing its value to the values obtained in all (or many) data sets members of the reference set. Since the number of permutations in the reference set may be too large to have all considered in computations, a random sample, usually very large, of the reference set is used instead. Edgington (1987:43) refers to the former as the *systematic data permutation*, and to the latter as the *random data permutation*.

The definition of the reference set, and the underlying null hypothesis, is a critical step. Here, the null hypothesis is that the observed resemblance between relevés α and β is the expected value of the resemblance were the conditions delimiting the reference set true. The locations in the data table, among which the permutation of observations proceeds, are the community components that occur in each relevé. The community components should be listed by relevés in an expanded data matrix, as in the example shown in Table 2.8.1. As for the observations, we can distinguish two spheres for randomization. In one, (a) the observations are the taxa (CSTs) to which the community components belong. The community components given, the null hypothesis is that random dispersion and establishment has defined the combination of CSTs observed in the relevés. We shall call this the *random composition hypothesis*. In another case, (b) the observations are the character states, and the null hypothesis is, in addition to the stated in (a), that the character states are randomly assorted among the CSTs. In other words, under the null hypothesis, CSTs arose in a random process. This we call the *random taxon hypothesis*. The two cases are equivalent when there is only one defining character, such as in species-based descriptions.

The definition of the reference set is further complicated by the fact that a performance value is usually attached to the community component. We describe two approaches to test the null hypothesis stated in (a) or (b). In one, (c) the observed performance value is inherent to the community component in the relevé, and it is not subjected to randomization. The assumption that the relevé has an internal structure in which some community components are usually dominant is implicit. In another approach, (d) the performance attained by a community component is assumed to be the result of a random process. Accordingly, the allocation of the relevé total of performance values among the community components is random. Under this assumption, in the internal structuring of the relevé, each community component has the same probability to occupy the space.

In the context of the hierarchical nested model, the randomization procedure described in Orlóci *et al.* (1986) for the determination of confidence limits of correlation coefficients between relevés, adopts the random taxon hypothesis (b above) combined with hypothesis (d) described above. Here, we develop the methods for the random composition and random taxon hypotheses combined with hypothesis (c). We do not further consider the methods involving hypothesis (d). Henceforth, when we refer to the methods as under the random composition hypothesis or under the random taxon hypothesis, it is also implied that case (c) applies.

The examples in Tables 2.8.1 and 2.8.2 present the complete reference set according to the two methods. For the same data set, the reference set under the random composition hypothesis (Table 2.8.1) is smaller than under the random taxon hypothesis (Table 2.8.2). Also, the probability of obtaining a dissimilarity smaller than the observed is higher under the random composition hypothesis because the randomization is less stringent. This trend becomes the most evident in large character sets, in which case there is much less chance of generating common CSTs in two relevés under the random taxon hypothesis.

Table 2.8.1. Example of systematic data permutation using artificial data. There are 2 dichotomous characters (p, q) and 2 relevés (α and β). The first data set is the observed. The complete reference set is shown, generated under the random composition hypothesis (a in the main text). Relevé α and β have 2 components and 1 community component respectively. The 6 data sets are all the permutations of the CSTs among the community components. The 2nd column under C/A is used when the data set has identical CSTs (row vectors), in which case their performance values are pooled. The absolute value function $a_{\alpha\beta}$ (Eq. 2.19) is computed in each data set. The 1 - α probability of obtaining $a_{\alpha\beta}$ smaller than the observed ($a_{\alpha\beta}$ = 7) is 2/6 = 0.33. Only values on hierarchical level 1 are shown.

Ch. state		C/A		$a_{\alpha\beta}$
p	q	$\alpha\beta$	$\alpha\beta$	
2	*1*	*4 0*	4 0	
1	*2*	*2 0*	2 5	
1	*2*	*0 5*		7
2	1	4 0	4 0	
1	2	2 0	2 5	
1	2	0 5		7

Ch. state		C/A		$a_{\alpha\beta}$
p	q	$\alpha\beta$	$\alpha\beta$	
1	2	4 0	6 0	
1	2	2 0		
2	1	0 5	0 5	11
1	2	4 0	4 5	
2	1	2 0	2 0	
1	2	0 5		3

Ch. state		C/A		$a_{\alpha\beta}$
p	q	$\alpha\beta$	$\alpha\beta$	
1	2	4 0	4 5	
2	1	2 0	2 0	
1	2	0 5		3
1	2	4 0	6 0	
1	2	2 0		
2	1	0 5	0 5	11

Table 2.8.2. Example of systematic data permutation using the data set from Table 2.8.1. The complete reference set, as shown, is generated under the random taxon hypothesis (b in the main text). The first data set is the observed one. The other 35 data sets are generated by permuting the character states within the columns of the character table of the observed data set. The probability of obtaining an $a_{\alpha\beta}$ smaller than the observed ($a_{\alpha\beta} = 7$) is 4/36 = 0.11. Only values on hierarchical level 1 are shown.

Ch. state		C/A		$a_{\alpha\beta}$	Ch. state		C/A		$a_{\alpha\beta}$	Ch. state		C/A		$a_{\alpha\beta}$	Ch. state		C/A		$a_{\alpha\beta}$
p	q	$\alpha\beta$	$\alpha\beta$		p	q	$\alpha\beta$	$\alpha\beta$		p	q	$\alpha\beta$	$\alpha\beta$		p	q	$\alpha\beta$	$\alpha\beta$	
2	1	40	40		2	2	40			1	1	40			1	2	40	45	
1	2	20	25		1	1	20			2	2	20			2	1	20	20	
1	2	05		7	1	2	05		11	1	2	05		11	1	2	05		3
2	1	40	40		2	2	40			1	1	40			1	2	40	45	
1	2	20	25		1	1	20			2	2	20			2	1	20	20	
1	2	05		7	1	2	05		11	1	2	05		11	1	2	05		3
2	2	40			2	2	40			1	2	40			1	2	40		
1	2	20			1	2	20			2	2	20			2	2	20		
1	1	05		11	1	1	05		11	1	1	05		11	1	1	05		11
2	2	40			1	1	40			1	2	40	45		1	1	40		
1	1	20			1	2	20			2	1	20	20		1	2	20		
1	2	05		11	2	2	05		11	1	2	05		3	2	2	05		11
2	2	40			1	1	40			1	2	40	45		1	1	40		
1	1	20			1	2	20			2	1	20	20		1	2	20		
1	2	05		11	2	2	05		11	1	2	05		3	2	2	05		11
2	2	40			1	2	40	60		1	2	40			1	2	40	60	
1	2	20			1	2	20			2	2	20			1	2	20		
1	1	05		11	2	1	05	05	11	1	1	05		11	2	1	05	05	11
2	1	40	40		1	2	40			1	1	40			1	2	40		
1	2	20	25		1	1	20			2	2	20			1	1	20		
1	2	05		7	2	2	05		11	1	2	05		11	2	2	05		11
2	1	40	40		1	2	40			1	1	40			1	2	40		
1	2	20	25		1	1	20			2	2	20			1	1	20		
1	2	05		7	2	2	05		11	1	2	05		11	2	2	05		11
2	2	40			1	2	40	60		1	2	40			1	2	40	60	
1	2	20			1	2	20			2	2	20			1	2	20		
1	1	05		11	2	1	05	05	11	1	1	05		11	2	1	05	05	11

The total number of permutations under the random composition hypothesis is p!. In this, $p = \sum_{\gamma=1}^{\nu} p_\gamma$, where p_γ is the number of CSTs present in relevé γ, and ν is the total number of relevés. The total number of permutations under the random taxon hypothesis is $(p!)^m$, where m is the number of defining characters.

The algorithm for random data permutation do not generate the complete reference set, but at each iteration a random data set is created according to the underlying null hypothesis (see Dwass 1957, Hope 1968 and Edgington 1987:43). The observed data set is part of the reference set and is also counted as an iteration, though not a random one. The probability α of attaining resemblance values $r_{i\alpha\beta}$ larger than or equal to the one observed is the relation

$$P(r_{i\alpha\beta RND} \geq r_{i\alpha\beta}) = \frac{\text{number of iterations in which } r_{i\alpha\beta RND} \geq r_{i\alpha\beta}}{\text{total number of iterations}}$$

where $r_{i\alpha\beta RND}$ is any resemblance function, nominal or partial, between relevés α and β, hierarchical level i, computed at each iteration. $r_{i\alpha\beta}$ is the same resemblance, but measured in the observed data set. Instead of α, the $1-\alpha$ probabilities can be used when the resemblance function is a dissimilarity, in which case $1-\alpha = P(r_{i\alpha\beta RND} < r_{i\alpha\beta})$. The computation of resemblances on the m hierarchical levels is performed in the same data set at each iteration. An example using random data permutations is shown in Table 2.8.3 for the Caatinga and Chaco data set used in previous examples. All computations use the program described in Chapter 6.

The total number of iterations in the random data permutation is chosen. The larger it is, the closer α is to the exact α in complete systematic data permutation. A formal proof is given by Hope (1968). Empirically, for instance, for the data in Table 2.8.1, in a trial with 100 iterations the $1-\alpha$ probability is 0.35, while with 1000 iterations the $1-\alpha$ probability is 0.337. The exact probability is 0.33.

The matrix of between relevés similarities or dissimilarities computed as probabilities can be subjected to ordination or clustering in the same way as any other resemblance matrix. It should, however, be noted that the resemblance value based on probability is only valid within the data set. A different probability may arise for the same relevé comparison, if other relevés are added to or removed from the table (Orlóci 1978:64).

Table 2.8.3. Probabilities (1–α) generated by randomization for the nominal absolute value function, comparing the Caatinga and Chaco relevés (Table 2.1.1). Randomization is performed under the random composition and random taxon null hypotheses explained under (a) and (b) in the main text. The algorithm uses random data permutation. Note the lower probabilities on all hierarchical levels when the random taxon hypothesis applies. The analysis is performed by the application SYNCSA (Chapter 6). Randomization was carried to 1000 iterations. Computation time in a Macintosh SE/30 was 29.8 seconds under the random composition hypothesis, and 64.03 seconds under the random taxon hypothesis.

		1-α probabilities under	
Level	Nominal resemblances	random composition hypothesis	random taxon hypothesis
4	29	0.253	0.246
3	35	0.076	0.052
2	71	0.097	0.045
1	143	0.269	0.136

Chapter 3

FUZZY COMMUNITY COMPONENTS

Plant community analysis has assumed traditionally that the plant taxa represent discrete populations, which is in line with the notion of absolute discreteness in classical taxonomy. In such a taxonomy, the rule is that no organism may belong to more than one taxon. An alternative rule that permits overlapping taxa is more realistic. This is because organisms may show affinities to other organisms not just in their taxa and in their communities. The concept of CSTs opens the possibility of treating the community components as fuzzy entities within a flexible framework. This framework allows the taxonomy to arise as the community description progresses. Characteristically, in the approach here outlined, the field records contain crisp CSTs, but these are replaced in the analysis by their fuzzy set equivalents analytically defined based on CST similarities (Pillar and Orlóci 1991). Fuzzy sets form the basis of resemblance measurements in subsequent community level comparisons. The ability of the fuzzy approach to reveal patterns and pattern connections is evaluated by comparison of the observed fuzzy vegetation structures with underlying environmental structures. The computations are performed by application program SYNCSA.

3.1. The problem of analytical indeterminacy

We suggest that taxonomic discreteness may limit our understanding of nature, and plant communities in particular. One of the manifestations of taxonomic discreteness is the problem of *analytical indeterminacy* in community comparisons. Different taxonomies dissect the multitude of community components differently. Through this, they determine the number of absences (zeros) in the records. We refer to Lambert and Dale (1964) and Orlóci and Stofella (1986) for a discussion of the effects of zeros in data interpretation. Absences, indicated by zeros, generate indeterminacy. We measure an absence-related indeterminacy $\phi_{\alpha,\beta}$ in terms of the number of shared community components in two relevés α and β:

$$\phi_{\alpha,\beta} = 1 - \frac{\text{number of shared components}}{\text{total number of distinct components in } \alpha \text{ and } \beta} \qquad (3.1)$$

This increases as the number of components not shared increases. Total indeterminacy occurs when no community component is shared. For instance, indeterminacy is complete when the comparison involves relevés A and B shown in Table 3.1.1. The example in Table 3.1.1 shows that no matter how different the vegetation is during time 1 to 2 in communities A, B and C, no differences can be detected by the analysis in the community data structure, if the changes do not alter indeterminacy. Thus, indeterminacy will set the limits of the universe within which community comparisons are meaningful. A species-based taxon-

Table 3.1.1. An artificial sample, following Orlóci's (1978:46) example, to illustrate the effect of indeterminacy on community resemblance. Indeterminacy is complete ($\phi_{A,B} = 1$) when comparing A to B at time 1 or time 2. The cross product q_{AB} remains zero irrespective of changes that take place in the communities from time 1 to 2. The distance $d^2_{AB} = 2^2 + 3^2 + 1^2 + 1^2 + 2^2 + 4^2 = 35$ also remains the same provided that the quantities globally, irrespective of component, do not change. Indeterminacy is partial when A and C are compared, but since the changes from time 1 to 2 do not affect the shared component a, the resemblance values remain the same.

Community components	Time 1 A	Time 1 B	Time 1 C	Time 2 A	Time 2 B	Time 2 C
a	2	0	2	2	0	2
b	3	0	0	1	0	0
c	1	0	0	3	0	0
d	0	1	0	0	4	0
e	0	2	0	0	2	0
f	0	4	0	0	1	0
g	0	0	5	0	0	1
h	0	0	1	0	0	5
	$q_{AB} = 0$	$q_{AC} = 4$	$q_{BC} = 0$	$q_{AB} = 0$	$q_{AC} = 4$	$q_{BC} = 0$
	$d^2_{AB} = 35$	$d^2_{AC} = 36$	$d^2_{BC} = 51$	$d^2_{AB} = 35$	$d^2_{AC} = 36$	$d^2_{BC} = 51$

omy narrows these limits to the same floristic region and the same type of habitat within the floristic region. A CST-based taxonomy (see Chapter 2) is likely to broaden the limits, since the community components will likely be less frequently absent than the species based components, and consequently, the absence-related indeterminacy will be lower. Furthermore, hierarchical nesting may reduce indeterminacy at higher hierarchical levels, as identical CSTs have their performances pooled at each level, eliminating some of the zeros. In this process, CSTs have so far been treated as discrete entities (Orlóci and Orlóci 1985, Orlóci *et al.* 1986) and their performances are not pooled if they are not completely identical, notwithstanding that some degree of similarity may exist between them.

3.2. Fuzzy sets

An analysis that accepts the assumption that community components are populations belonging to absolutely discrete taxa is vulnerable, since it treats all absences (zeros) as denoting the same thing, no matter how similar the absent components might be to any of those that are present. To overcome this problem, we suggest to view community components as fuzzy entities and to represent them in the analysis by *fuzzy sets.* Although my proposition is novel, the mathematical theory is well known (Zadeh 1965). It is in fact an extension of classical set theory. The notation and definitions we use here are mainly based on Zadeh (1965) and Bezdek (1987):

a) A set is said to be fuzzy if the *membership function* on this set assumes values in the interval [0,1].

b) The membership function is another way to describe a set. Considering X a set with a generic element x, and B a subset of X, the membership function associates with each object in X a real number in the interval [0,1]. This number represents what Zadeh refers to as the "grade of membership", $u_B(x)$ of x in the subset B.

c) In the particular case when the membership function only assumes 1 if it belongs to B, or 0 if not, the set is called a "hard set" (Bezdek) or "ordinary set"

(Zadeh), which is covered by classical set theory[1]. Here we refer to this case as a *crisp set*. If B is fuzzy[2], as Bezdek notes, "the grade of membership measures the extent to which x possesses the imprecisely defined object properties which characterize B"[3]. The closer the value of $u_B(x)$ to unity, the more clearly x belongs to what is defined as B.

d) Translating these into our problem, in the fuzzy approach a population is no longer identified as an exclusive member of one or another taxon, but as an entity with a grade of membership in each of several taxa, albeit the membership may not be equally strong. Any such grade expresses a population's affinity to a taxon on a 0 to 1 scale. Therefore a fuzzy taxonomy, instead of having crisp taxa, depicts taxa as fuzzy sets. Fuzzy sets have been applied in vegetation ecology (Feoli and Zuccarello 1986, 1988, Roberts 1986, Dale 1988, Banyikwa, Feoli and Zuccarello 1990), but these applications have considered fuzziness at the community level and not at the community component level as we do it here. Fuzzy sets in the quoted vegetation studies are based on the degree of compositional or environmental similarity between entire stands bypassing the population level. In my case, they are based on the similarity of CSTs. Fuzzy clustering algorithms have been used in numerical taxonomy (Bezdek 1974), and also in quantitative ecology (Marsili-Libelli 1989).

Community comparisons are very basic in ecological reasoning and in the practice of data manipulations, in which meaningful patterns of change in space

---

[1] For instance, a set H composed of the real numbers between 6 and 8 inclusive can be represented by $H = \{r \in \mathbf{R} : 6 \le r \le 8\}$ or by the membership function $u_H: \mathbf{R} \rightarrow \{0,1\}$ in which

$$u_H(r) = \left\{ \begin{array}{l} 1 \text{ when } r \in H, \textit{ i.e.}, 6 \le r \le 8 \\ 0 \text{ when } r \notin H, \textit{ i.e.}, r < 6 \text{ or } r > 8 \end{array} \right\}$$

[2] For instance, the membership function to represent B, a fuzzy set of the stands r with very low soil potassium content, can be defined as $u_B(r) = 1 - \frac{K_r - \mathrm{INF}\,K}{\mathrm{SUP}\,K - \mathrm{INF}\,K}$, where INF and SUP indicate infirmum and supremum potassium contents in the set of stands. In this case the stand with the INF K content has $u_B(r) = 1$, an intermediate may have say $u_B(r) = 0.55$, and the one with the SUP K content has $u_B(r) = 0$. Other types of membership functions could be defined.

[3] For fuzzy sets only the membership function defines the set clearly; for hard sets the "verbal" and "membership function" forms of representation are equivalent and also precise (Bezdek 1987).

and time are to be revealed. Hence, the relevant task is how to define operationally fuzzy community components, how to incorporate the fuzziness into resemblance measures, and how to evaluate the effect on the intrinsic structures and on the connection of these with extrinsic structures.

3.3. Fuzzy community components

As already explained, in character-based analysis the community components are populations described as character set types (CSTs). These are defined by a sequence of states of plant attributes arranged in a hierarchy. The CSTs may share character states and this sharing can be expressed by a similarity index. The problem of constructing a fuzzy set equivalent of each CST can thus be handled on the basis of the degree of similarity between the CSTs. In the method proposed here there is no need for a clustering algorithm to classify the CSTs among fuzzy taxa, since CSTs are the taxa. There are as many fuzzy sets as there are CSTs and all CSTs have membership in all of the fuzzy sets. The grade of membership of CST_e in the fuzzy set equivalent of CST_d, *i.e.*, the fuzzy set in which CST_d has a membership equal to 1, is the similarity of CST_e and CST_d. It follows that the grades of membership are symmetric, *i.e.*, membership $u_d(CST_e) = u_e(CST_d)$.

Since the CSTs are mapped into a nested character hierarchy as runs through the nodes, CST similarities will change depending on the hierarchical level. Recalling from Chapter 2, the ith of m levels contains k_i nodes. At a node d, level i, there is defined a CST_{id} composed of the states of the first m-i+1 characters in the set (the ith level plus the m-i higher levels in the hierarchy). The fuzzy set associated with any CST_{id} is the relation,

$$F_{id} = \{CST_{ie}\ ,\ u_{id,ie}\!:\ CST_{ie} \text{ similar to } CST_{id}\}\quad e=1, ..., k_i$$

That is, on level i at node e the CST_{ie} has grade of membership $u_{id,ie} = u_{id}(CST_{ie})$ in the fuzzy set composed of the CSTs similar to the CST_{id} at node d. The grade of membership ($u_{id,ie}$) is any index measuring the similarity between CST_{id} and CST_{ie} in the interval [0, 1]. The index proposed by Gower (1971)[4] is

---

[4] The Gower type similarity between the CSTs at node d and node e, hierarchical level i, is

of this kind. The Gower index is convenient to use, since it can handle mixed character types.

To illustrate the method of computing CST fuzzy set equivalents, we use the data in Table 3.3.1. This table displays the same Caatinga and Chaco relevés of the previous chapter but with an additional character (leaf texture) included. Table 3.3.3 depicts the grades of membership in the fuzzy sets equivalent to the CSTs at different hierarchical levels.

For a given relevé, the CSTs' associated fuzzy set has a performance value ($Y_{id\alpha}$) based on the performance values ($X_{id\gamma}$) of its members weighted by the respective grade of membership. For fuzzy set F_{id} in relevé α, the performance value is

$$Y_{id\alpha} = 0 \qquad \text{if} \sum_{\gamma=1}^{\nu} X_{id\gamma} = 0 \tag{3.2a}$$

or

$$Y_{id\alpha} = 0 \qquad \text{if } X_{id\alpha} = 0 \text{ and } X_{id\beta} = 0 \tag{3.2b}$$

$$S_{ide} = \frac{\sum_{h=1}^{m-i+1} t_{deh}\, \delta_{deh}}{\sum_{j=1}^{m-i+1} \delta_{dej}}$$

In this equation t_{deh} and δ_{deh} are scores assigned according to the type of the hth character:

(1) If character h is dichotomous (the states are presence or absence; +,–) and

	t_{deh}	δ_{deh}
present in both CSTs	1	1
absent from both CSTs	0	0
CSTs disagree on the character	0	1

(2) If character h is qualitative, $t_{deh} = 1$ if CST d and CST e agree on the character, and $t_{deh} = 0$ if they disagree. In either case $\delta_{deh} = 1$.

(3) If character h is quantitative $t_{deh} = 1 - \frac{|v_{dh} - v_{eh}|}{(\max v_h - \min v_h)}$ and $\delta_{deh}=1$. v_{dh} is the state of character h, CST d. The extreme values for character h may be defined *a priori* or as the ones realized within the sample.

Table 3.3.1. Character score matrix of two plant communities from the Caatinga (NE Brazil) and Chaco (NW Argentina) formations. For simplicity only five characters are used, which are extracted from a larger character set.

Characters	Character Set Types												
	a	b	c	d	e	f	g	h	i	j	k	l	m
STEM:													
Tissue. **s**ucculent, **w**oody	w	s	w	w	w	w	w	w	w	w	w	w	w
Function. **r**egular, **t**win-purpose	r	t	r	r	r	r	r	r	r	r	r	t	t
Armature. **t**horn/spine, **n**one	n	t	n	t	n	n	n	n	n	t	t	t	n
LEAF:													
Duration. **d**eciduous, **p**ersistent, **l**eafless	d	l	d	d	d	d	d	p	p	d	d	d	d
Texture. 1-5 (herbaceous to leathery) 6 (leafless)	3	6	2	1	1	4	5	5	4	3	4	4	2
Cover (%) Caatinga	48	2	1	16	1	5	5	1	0	0	0	0	0
Chaco	0	30	0	16	2	0	0	5	32	17	1	2	1

If neither condition is true then

$$Y_{id\alpha} = \sum_{e=1}^{k_i} \left[(u_{id,ie})^{\frac{1}{\xi}} X_{ie\alpha} \right] \tag{3.3}$$

for any i = 1, ..., m level, d = 1, ..., k_i node, and $\alpha = 1, ..., \nu$ relevé. $X_{ie\alpha}$ is the cumulative CST performance at level i, node e, relevé α. Note that ξ is an arbitrarily chosen *fuzziness degree* always larger than 0 but not greater than 1, which behaves similarly to the "m" value in the fuzzy clustering algorithm presented in Marsili-Libelli (1989). The larger the fuzziness degree, the fuzzier is the fuzzy set adjustment of the CST performances. When the fuzziness degree tends to zero $\lim_{\xi \to 0} (u_{id,ie})^{1/\xi} = 0$ if $u_{id,ie} < 1$ or $\lim_{\xi \to 0} (u_{id,ie})^{1/\xi} = 1$ if $u_{id,ie} = 1$; hence a crisp set membership is redefined and $Y_{id\alpha} = X_{id\alpha}$. Equations 3.2a and 3.2b are alternative rules. Instead of using rule 3.2a, the *global adjustment*, which assumes that the CST for which the fuzzy set F_{id} is defined is present in at least one relevé in the collection to be considered as potentially present in relevé α, an alternative approach uses rule 3.2b for the relevé pair α, β, the *pairwise adjustment*. As a consequence of rule 3.2b, different CST performances arise for each relevé pair. In any case, the performance of CSTs completely absent in a data set

Table 3.3.2. Performances (cover-abundances) of crisp and fuzzy CSTs for the Caatinga and Chaco relevés in Table 3.3.1. Hierarchical accumulation of crisp CST performance follows the method explained in Chapter 2. Computation of fuzzy CST performance uses the crisp CST performances, the grades of membership from Table 3.3.3 and a degree of fuzziness $\xi = 1$. For instance, the value for fuzzy CST e, level 2 in the Caatinga is computed in 2 steps: $Y_{2e\ Caat} = 0.5(60) + 0.5(2) + 0.75(16) + 0.25(1) + 1(0) + 0.75(0) = 43.25$, which after computation of all Y_{2*Caat} is then adjusted by $Z_{2e\ Caat} = 43.25\ (60 + 2 + 16 + 1 + 0 + 0)\ /\ (72.75 + 6 + 62 + 54 + 43.25 + 54) = 11.70$.

		Cover-abundance			
CSTs		Crisp CST		Fuzzy CST	
#	Level 1	Caatinga	Chaco	Caatinga	Chaco
a	w r n d 3	48	0	7.57	8.4
b	s t t l 6	2	30	1.12	6.2
c	w r n d 2	1	0	7.39	8.02
d	w r t d 1	16	16	6.29	8.36
e	w r n d 1	1	2	7.2	7.63
f	w r n d 4	5	0	7.35	8.59
g	w r n d 5	5	0	7.08	8.39
h	w r n p 5	1	5	5.5	8.33
i	w r n p 4	0	32	5.76	8.54
j	w r t d 3	0	17	6.66	9.13
k	w r t d 4	0	1	6.44	9.32
l	w t t d 4	0	2	4.85	8.2
m	w t n d 2	0	1	5.8	6.9
	Level 2				
a	w r n d	60	2	19.7	19
b	s t t l	2	30	1.62	13.3
c	w r t d	16	34	16.8	21.2
d	w r n p	1	37	14.6	18.9
e	w t t d	0	2	11.7	17.9
f	w t n d	0	1	14.6	15.7
	Level 3				
a	w r n	61	39	26.3	23.4
b	s t t	2	30	2.69	16
c	w r t	16	34	21	26.6
d	w t t	0	2	11.9	21.6
e	w t n	0	1	17.1	18.4
	Level 4				
a	w r	77	73	51.3	49.2
b	s t	2	30	1.33	20.8
c	w t	0	3	26.3	36
	Level 5				
a	w	77	76	77	76
b	s	2	30	2	30

Table 3.3.3. Grades of membership in the fuzzy sets corresponding to the CSTs described in Table 3.3.2 for the different hierarchical levels. There are as many fuzzy sets as there are CSTs, and all CSTs have membership in all fuzzy sets. The values are similarity indices of the CSTs using Gower's (1971) method. For instance, at level 1 the grade of membership of CST a in the fuzzy set equivalent to CST d is $u_{1d}(CST_{1a}) = S_{1d,1a} = S_{1a,1d} = [(1)(1) + (1)(1) + (0)(1) + (1)(1) + (1- |3-1| / (6-1)] / 5 = 0.72$. Note that leaf texture is treated as a quantitative character. At level 2, the grade of membership of CST d in fuzzy set a is $u_{2a}(CST_{2d}) = S_{2a,2d} = S_{2d,2a} = [(1)(1)+(1)(1)+(1)(1)+(0)(1)] / 4 = 0.75$. Computations use application SYNCSA.

Fuzzy CSTs level 1	Grade of membership												
	a	b	c	d	e	f	g	h	i	j	k	l	m
a	1	0.08	0.96	0.72	0.92	0.96	0.92	0.72	0.76	0.8	0.76	0.56	0.76
b	0.08	1	0.04	0.2	0	0.12	0.16	0.16	0.12	0.28	0.32	0.52	0.24
c	0.96	0.04	1	0.76	0.96	0.92	0.88	0.68	0.72	0.76	0.72	0.52	0.8
d	0.72	0.2	0.76	1	0.8	0.68	0.64	0.44	0.48	0.92	0.88	0.68	0.56
e	0.92	0	0.96	0.8	1	0.88	0.84	0.64	0.68	0.72	0.68	0.48	0.76
f	0.96	0.12	0.92	0.68	0.88	1	0.96	0.76	0.8	0.76	0.8	0.6	0.72
g	0.92	0.16	0.88	0.64	0.84	0.96	1	0.8	0.76	0.72	0.76	0.56	0.68
h	0.72	0.16	0.68	0.44	0.64	0.76	0.8	1	0.96	0.52	0.56	0.36	0.48
i	0.76	0.12	0.72	0.48	0.68	0.8	0.76	0.96	1	0.56	0.6	0.4	0.52
j	0.8	0.28	0.76	0.92	0.72	0.76	0.72	0.52	0.56	1	0.96	0.76	0.56
k	0.76	0.32	0.72	0.88	0.68	0.8	0.76	0.56	0.6	0.96	1	0.8	0.52
l	0.56	0.52	0.52	0.68	0.48	0.6	0.56	0.36	0.4	0.76	0.8	1	0.72
m	0.76	0.24	0.8	0.56	0.76	0.72	0.68	0.48	0.52	0.56	0.52	0.72	1

CSTs level 2	a	b	c	d	e	f
a	1	0	0.75	0.75	0.5	0.75
b	0	1	0.25	0	0.5	0.25
c	0.75	0.25	1	0.5	0.75	0.5
d	0.75	0	0.5	1	0.25	0.5
e	0.5	0.5	0.75	0.25	1	0.75
f	0.75	0.25	0.5	0.5	0.75	1

CSTs level 4	a	b	c
a	1	0	0.5
b	0	1	0.5
c	0.5	0.5	1

CSTs level 3	a	b	c	d	e
a	1	0	0.67	0.33	0.67
b	0	1	0.33	0.67	0.33
c	0.67	0.33	1	0.67	0.33
d	0.33	0.67	0.67	1	0.67
e	0.67	0.33	0.33	0.67	1

CSTs level 5	a	b
a	1	0
b	0	1

remains zero after fuzzy adjustment. Thus, only similarities of CSTs materialized in the data set have to be computed.

In order not to change the total CST performance in each relevé, a correction is applied:

$$Z_{id\alpha} = Y_{id\alpha} \frac{\sum_{e=1}^{k_i} X_{ie\alpha}}{\sum_{e=1}^{k_i} Y_{ie\alpha}} \tag{3.4}$$

Table 3.3.2 shows nonadjusted (crisp CSTs) and fuzzy set adjusted (fuzzy CSTs) performances for the Caatinga and Chaco example.

Resemblances of relevés are thus computed based on the fuzzy adjusted CST performances ($Z_{id\alpha}$) using the resemblance functions already described in Chapter 2. Results for the Caatinga and Chaco example are in Table 3.3.4. Because fuzzy adjustments affect hierarchical additivity, the resulting resemblance values are not partitionable into hierarchical levels and so only nominal resemblances are interpretable. Although in this case it seems unnecessary to interpret the problem in a hierarchical framework, the fact that when the degree of fuzziness (ξ) is set to zero a crisp membership, and also the hierarchical partitioning is recovered, indicates the application of the same hierarchical approach for the sake of coherence.

Table 3.3.4. Cross products and correlation coefficients at various hierarchical levels using crisp and fuzzy CSTs (Table 3.3.2) for the Caatinga and Chaco relevés (Table 3.3.1). Computations follow equations presented in Chapter 2.

Level	Crisp CSTs used				Fuzzy CSTs used			
	q_{ChCa}	q_{Chaco}	q_{Caat}	Correl.	q_{ChCa}	q_{Chaco}	q_{Caat}	Correl.
5	23.96	14.69	39.06	1.	23.96	14.69	39.06	1.
4	99.65	95.25	121.5	0.9265	39.11	37.23	49.17	0.9140
3	107.6	121.0	177.8	0.7334	41.25	50.66	44.15	0.8722
2	68.68	494.3	600.2	0.1261	186.3	240.6	162.5	0.9421
1	264.8	2426.	2574.	0.1060	597.2	794.8	471.6	0.9754

3.4. Effect of fuzzy adjustments on data structure

Depending on the character set, fuzzy adjustments will change the data structure to different extents. The question is if this leads to a better understanding of the vegetation's structure and its environmental connections. We have pointed out the problem of analytical indeterminacy, but beyond this we ask: (1) What really is the effect of indeterminacy upon the ecological reliability of resemblance measures in describing vegetation structure? (2) How much improvement does result from using fuzzy taxa?

We propose to answer these questions in relative terms based on *structural evaluation functions.* Several of these functions are applied in Chapter 4. One of these measures the correlation between the vegetation data structure, here described by crisp or fuzzy CSTs, and an extrinsic data structure, which usually is environmental. This idea of relating data structures is present in methods to evaluate and compare dendrograms (see Sokal and Rohlf 1962, Sneath and Sokal 1973:277, Orlóci 1978:264, Podani and Dickinson 1984), to choose species number and type of data (Orlóci and Mukkattu 1973, Orlóci 1978:34) and to evaluate structural stability of mappings in process sampling (Orlóci and Pillar 1989). In these terms, on each hierarchical level i, a $\nu \times \nu$ symmetric matrix $\mathbf{D}_i$ of relevé dissimilarities, calculated from CST performances or from ordination scores, defines the vegetation structure. If the resemblances are correlation coefficients, they are transformed into squared distances using the relation $d^2_{i\alpha\beta} = 2(1 - r_{i\alpha\beta})$. Another $\nu \times \nu$ matrix $\mathbf{\Delta}$ of squared relevé distances based on environmental variables defines environmental structure. The structural evaluation function of interest is $\rho(\mathbf{D}_i;\mathbf{\Delta})$, which is a matrix correlation (Sokal and Rohlf 1962, Sneath and Sokal 1973:280) between $\mathbf{D}_i$ and $\mathbf{\Delta}$, *i.e.*, a product moment correlation involving the $1/2\ \nu(\nu - 1)$ off-diagonal elements in the half distance matrices. The function $\rho(\mathbf{D}_i;\mathbf{\Delta})$ ranges from –1 to +1, but only positive values not close to zero indicate agreement, or *congruence* (Sneath and Sokal 1973:97) between community and environmental data structures. This distinction is important, for when closer relevés in the community data structure tend to be farther apart in their environmental data structure, or vice-versa, there is a negative and strong correlation coefficient, but there is no congruence.

As an example, we use a data set from sub-boreal vegetation on L. Orlóci's recovery research site near Elk Lake, Ontario, Canada. The vegetation is secondary, about 3 years after logging. Quadrat size is 5 m. sq. The vegetation description is by score matrix relevés, in which community components are CSTs.

The character set is listed in Table 3.4.1 and a partial data set is given in Table 3.4.2. The matrix of nominal correlation coefficients using fuzzy CSTs is displayed in Table 3.4.3. Only results on hierarchical level 1 are presented. Figure 3.4.1 displays results emitting from a Q-type eigenordinationof that matrix (Section 5.2). Table 3.4.4 displays nominal correlation coefficients using crisp CSTs, along with values of the indeterminacy indexes. Note that indeterminacy is very high in some comparisons, particularly with quadrat 9. Figure 3.4.2 displays the ordination scattergram based on this set of correlations. Quadrat 9 appears as an extreme outlier in the ordination based on crisp CSTs.

Table 3.4.1. Character set used in community description on the Elk Lake transect site. For characters 1 and 2 only the states materialized in the data are presented. References identify the general class of notions; the system adopted elsewhere may differ. 'Stem' and 'leaf' refers to stem and stem-like or leaf and leaf-like structures.

Life-form (Mueller-Dombois & Ellenberg 1974:449)

1. Form (lf). 1:Phanerophytes, 2:Chamaephytes, 3:Hemicryptophytes, 4:Geophytes, 9:Thallo-chamaephytes

Growth-form (Barkman 1988a)

2. Form (gf). 6:polythrichid, 8:pleuroziid, 20:caespitose graminid, 23:decumbent herb, 24:arching herb, 26:scapos-rosulate herb, 27:erect scapose herb, 30:arctostaphyllid shrub, 31:andromedid (vacciniid), 34: sambucid, 38:piceid, 39:betulid

Stem (Shreve 1942)

3. Type (st). 1:short caudex, 2:long caudex, 3:truncus, 4:caulis, 5:culm, 6:stipe, 7:no stem
4. Consistence (co). 1:succulent, 2:herbaceous, 3:semi-ligneous, 4:ligneous, 5:no stem
5. Direction (di). 1:erect, 2:erect diffuse, 3:creeping, 4:climbing, 5: no stem

Leaf (Dansereau 1957:148)

6. Type (lt). 1:deciduous, 2:withering, 3: evergreen, 4:no leaf
7. Shape (sh). 1:needle, 2:graminoid, 3:broad, simple, 4:compound, 5:thalloid**,** 5:no leaf
8. Texture (tx). 1:filmy, 2:membranous 3:sclerophyll, 4:succulent, 5:no leaf

Leaf (Shreve 1942)

9. Epidermal surface (ep). 1:glabrous, 2:glaucous, 3:trichomous dense, 4:trichomous sparse, 5:no leaf

Leaf (Orlóci and Orlóci 1985)

10. Width (wi). 1:< 2.5 mm, 2:2.5-5, 3:5-10, 4:10-50, 5:50-100, 6:100<, 0:no leaf
11. Length (le). 1:<5mm, 2:5-25, 3:25-75, 4:75-125, 5:125<, 0:no leaf
12. Thickness (th). 1:< 1 mm, 2:1-3, 3:3-5, 4:5<, 0:no leaf

Plant height

13. Height class (he). 1: < 0.1m, 2: 0.1 - 0.5 , 3: 0.5 - 2, 4: 2 - 8 , 5: 8 - 10, 6: 10 - 25, 7: >25m

Table 3.4.2. CST cover-abundances in 9 quadrats on the Elk Lake recovery transect site. The characters are described in Table 3.4.1 and identified by their labels. The CSTs are defined by vectors of the states of the 13 characters of Table 3.4.1.

Characters													Cover-abundance values in quadrats								
lf	gf	st	co	di	lt	sh	tx	ep	wi	le	th	he									
CSTs													1	2	3	4	5	6	7	8	9
1	38	3	4	1	3	1	3	2	1	2	2	2	0	.01	0	0	0.1	0	0	0	0
1	38	3	4	1	3	1	3	1	1	3	2	2	0.1	0.1	0.1	1	1	0	1	0	0
1	39	3	4	2	1	3	2	1	4	2	1	3	2	2	1	0	1	0	0	0	0
1	39	3	4	2	1	3	2	1	4	3	1	3	.01	0.1	0	2	1	2	2	0.1	0.1
2	31	3	4	2	1	3	2	4	2	4	1	2	2	3	2	1	1	0.1	1	0	0
3	24	4	2	2	2	3	2	1	4	3	1	2	0	0	0	0	0	1	1	0	0
2	30	3	3	3	3	3	3	3	4	3	2	1	0	0	0	0	0	0	0	.01	0
2	30	3	3	3	3	3	3	1	4	3	1	1	1	0	.01	0	0	0	1	0	0
2	31	3	4	2	1	3	3	1	4	2	1	2	2	2	0.1	0.1	0.1	0	0	0.1	3
1	39	3	4	2	1	3	2	1	4	3	1	2	0	0.1	2	0	0	2	1	0	0
1	34	3	4	2	1	4	2	3	4	3	1	2	0	0	0.1	0	0	0	0	0	0
1	39	3	4	2	1	3	2	3	4	3	1	3	0	2	0	0	0	0	0	0	0
2	31	3	4	2	1	3	2	1	3	2	1	2	2	1	1	3	2	1	1	3	0
2	31	3	4	2	1	3	2	3	3	2	1	2	1	1	1	0	2	0	0	0	0
2	30	3	3	3	3	3	2	4	4	2	1	1	0.1	1	0.1	0	0	0	0	0	0
2	30	3	3	3	3	3	3	3	2	2	1	1	1	0	0	0	0	0	0	0	0
2	31	3	3	2	3	4	3	4	3	3	1	2	0	0	0	0	1	0	0.1	0.1	1
4	27	4	2	1	2	3	2	3	4	3	1	2	0	0	0	0	0	0	0	0	0.1
3	26	4	2	1	2	3	2	3	6	5	1	2	2	2	1	.01	0	0	0	0	0
3	24	4	3	2	2	3	2	1	4	5	1	2	.01	1	0.1	.01	0	0	0	0	0
4	27	4	2	1	2	3	2	4	4	4	1	1	1	1	0.1	0	0	0	0	0	0
2	23	4	2	2	2	3	2	1	3	3	1	1	.01	0.1	0.1	0	0	0	0	0	0
3	20	5	2	1	2	2	3	4	3	5	1	2	1	2	2	1	1	0	0.1	.01	0
3	20	5	2	1	2	2	3	4	2	3	1	1	0.1	0.1	0	0	1	0.1	0	0.1	.01
3	20	5	2	2	2	2	3	1	1	4	1	2	1	0.1	0	1	2	0	0	0	0
9	8	6	2	3	3	5	1	1	1	1	1	1	1	1	0.1	0	1	0	0	0	0
9	6	6	2	1	3	1	1	1	1	1	1	1	0	0.1	0.1	1	1	0	0.1	0	0.1

The environmental variables considered in this example are elevation, exposure, slope, soil depth, and soil texture (Table 3.4.5). The distance matrix representing environmental data structure is in Table 3.4.6. The congruence between environmental and vegetational structure on hierarchical level 1 is higher when fuzzy taxa is adopted as indicated in Table 3.4.6 and Figures 3.4.1, 3.4.2.

It should be noted that the analysis is much dependent on the character set and environmental variables used. A high congruence between community and environmental structures may indicate, in a small scale study, how well the taxonomy is discriminating between populations that vary compositionally according to the local environmental change or, when a broader geographical and floristic range is studied, how well the characters are reflecting environmental selection. In both cases, the virtual elimination of analytical indeterminacy may increase the understanding of spatial or temporal plant community patterns.

Table 3.4.3. Nominal correlation coefficients between quadrats described by fuzzy CSTs, hierarchical level 1, based on the Elk Lake data, using pairwise fuzzy adjustments and fuzziness degree 1.

	Quadrats								
	1	2	3	4	5	6	7	8	9
1	1	0.998	0.996	0.995	0.995	0.981	0.992	0.981	0.983
2		1	0.999	0.994	0.991	0.986	0.992	0.983	0.980
3			1	0.995	0.992	0.989	0.994	0.985	0.980
4				1	0.997	0.984	0.995	0.979	0.984
5					1	0.966	0.985	0.965	0.980
6						1	0.993	0.986	0.970
7							1	0.990	0.986
8								1	0.989
9									1

Table 3.4.4. Nominal correlation coefficients between quadrats described by crisp CSTs, and values for the indeterminacy index, hierarchical level 1, based on the Elk Lake data.

	Quadrats								
	1	2	3	4	5	6	7	8	9
Nominal correlations:									
1	1	0.820	0.644	0.469	0.605	0.136	0.317	0.398	0.365
2		1	0.708	0.347	0.471	0.091	0.246	0.182	0.321
3			1	0.424	0.506	0.410	0.419	0.251	0.024
4				1	0.720	0.529	0.685	0.723	0.045
5					1	0.289	0.434	0.458	0.104
6						1	0.808	0.337	0.020
7							1	0.337	0.031
8								1	0.043
9									1
Indeterminacy index:									
1	0	0.27	0.33	0.53	0.48	0.8	0.73	0.75	0.86
2		0	0.32	0.5	0.38	0.76	0.70	0.77	0.82
3			0	0.58	0.59	0.85	0.65	0.86	0.90
4				0	0.5	0.77	0.57	0.69	0.77
5					0	0.75	0.59	0.6	0.67
6						0	0.55	0.7	0.8
7							0	0.69	0.77
8								0	0.56

9	0

How can we be sure that all these data manipulations are ecologically consistent? The answer to this question requires a probabilistic assessment of the structural evaluation function, which in this case involves finding how large a given value of congruence is if it were a complete random phenomenon. In this case, a randomization test under the random composition null hypothesis is applied (Section 2.8), but congruence is also computed at each iteration. The α probability of getting congruence values larger than or equal to the one observed

Table 3.4.5. Environmental information recorded from 9 quadrats, Elk Lake recovery transect site, referred in Table 3.4.2. The variables are: elevation (ele, meters above lowest point in transect), soil depth (dpt, 1.<=3cm, 2. 3-10cm, 3. 10-15cm, 4. 15-25cm, 5.>25cm), soil texture (txt, 1.sand, 2.gravel, 3.broken rocks, 4.solid rocks), slope (slo, degrees), and exposure (exp, 1.N, 2.S, 3.neutral).

	Quadrats								
Variables	1	2	3	4	5	6	7	8	9
ele	3.2	1.7	1.3	0.45	0.28	6.17	9	10.25	10.38
dpt	5	5	5	5	5	1	2	2	4
txt	1	2	2	1	1	4	4	3	3
slo	7.7	3.6	3.6	3.6	0	35	45	5	5
exp	1	1	1	1	3	2	2	3	3

Table 3.4.6. Chord distances between quadrats described by environmental variables, in the Elk Lake example. The variables considered are elevation, exposure, slope, soil depth, and soil texture. The congruence $\rho(\mathbf{D}_i;\mathbf{\Delta})$ is 0.273 ($\alpha = 0.126$) using crisp CSTs (Table 3.4.4) and 0.618 ($\alpha = 0.001$) using fuzzy CSTs (Table 3.4.3). Randomization was carried in 1000 iterations under the random composition null hypothesis. See main text for method.

	Quadrats								
	1	2	3	4	5	6	7	8	9
1	1	0.402	0.419	0.443	0.986	0.543	0.524	0.748	0.678
2		1	0.0580	0.235	0.693	0.867	0.859	0.834	0.722
3			1	0.195	0.682	0.869	0.862	0.882	0.773
4				1	0.669	0.877	0.869	1.01	0.900
5					1	1.36	1.36	1.15	1.04
6						1	0.0392	0.912	0.934
7							1	0.899	0.919
8								1	0.152
9									1

is the relation

$$\frac{\text{number of iterations in which } \rho(\mathbf{D}_{i\mathrm{RND}};\mathbf{\Delta}) \geq \rho(\mathbf{D}_i;\mathbf{\Delta})}{\text{total number of iterations}} \tag{3.5}$$

where, on hierarchical level i, $\rho(\mathbf{D}_{i\ \mathrm{RND}};\mathbf{\Delta})$ is the congruence computed at each iteration and $\rho(\mathbf{D}_i;\mathbf{\Delta})$ is the congruence observed in the original data set. In the example, the probability of getting a congruence at least as large as the observed under the random composition null hypothesis is 0.126 using crisp CSTs and 0.001 using fuzzy CSTs for 1000 iterations (Table 3.4.6). Indeed, the use of fuzzy CST produce a higher congruence between community and environmental structures that is very unlikely to happen by chance, compared with the congruence produced by using crisp CSTs. From this, it can be concluded that the environmental factors in consideration do have an influence, or represent other factors that do so, on the survival and performance of the populations that form the communities, as described by fuzzy CSTs.

For the sake of simplicity the example considers only the first hierarchical level. Unless an optimal character order is found (see Chapter 4), the congruence on the other levels is unpredictable. However, in the fuzzy CST analysis the problem of character order is not so critical as it is in the crisp CST analysis. This is so because a fuzzy CST analysis reveals on level i an average of the data structures that would be revealed by a crisp CST analysis on and above level i. An example is given in Section 4.4.

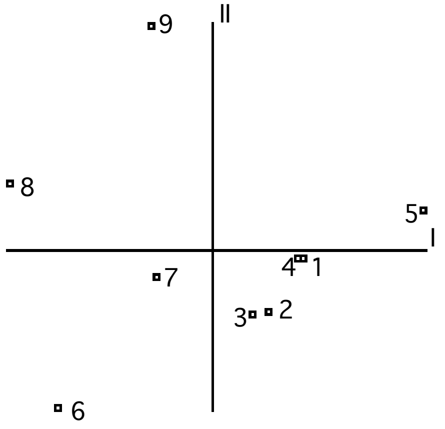


Figure 3.4.1. Scattergram of the first two eigenaxes of the matrix of fuzzy-based nominal correlation coefficients (Table 3.4.3), hierarchical level 1. The eigenvalues for I and II account respectively for 49.8% and 32.8% of the total. The congruence $\rho(\mathbf{D}_i;\mathbf{\Delta})$ between the vegetation

structure so defined and environmental structure (see main text for method) is 0.660 based on the 2 axes.

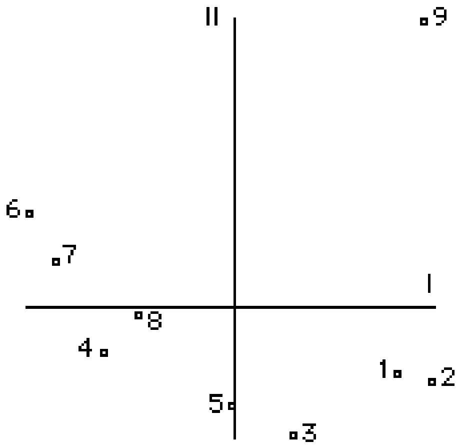


Figure 3.4.2. Scattergram of the first two eigenaxes of the matrix of nominal correlations using crisp CSTs (Table 3.4.4), hierarchy level 1. The eigenvalues for I and II account respectively for 34.6% and 22.3% of the total. The congruence $\rho(\mathbf{D}_i;\Delta)$ between vegetation structure so defined and environmental structure (see main text for method) is 0.125 based on the 2 axes, and 0.304 considering all eigenaxes.

Chapter 4

CHARACTER ORDERING

Since our perception of community level relationships is dependent on the taxonomy, the selection of characters to erect this taxonomy is a critical step. In this chapter an a priori set of defining characters on the first hierarchical level is assumed. The problem addressed is the order for arranging them hierarchically. This is equivalent to selecting character subsets on hierarchical levels above the first level. Since varying character order will place different character subsets on the same hierarchical level, the vegetational structure and structural connections perceived on that level may be affected. Optimization of order is approached by character ranking according to character relevance in unfolding data structures. We consider several quantitative criteria for ranking. By these, character selection within a character set is an analytical exercise helped by computer algorithms, which are implemented in application SYNCSA.

4.1. The relevance of character order

If the complete character set is considered, the order in which characters are arranged will not affect the discrimination of populations (CSTs). For instance, we can say that a plant population has thorny stems and broad, deciduous leaves or that it has deciduous, broad leaves and thorny stems; the order of the "characters" has no effect on the logic of the statement. However, when the analysis uses the hierarchical nested model, the taxonomy is analytically decomposed according to hierarchical level and the character order will affect the results above the first hierarchical level. The result of this is similar to the case of orthogonal functions in statistical analysis (Rao 1952:345). On the first hierarchical level, community resemblances are calculated using the populations as described by the entire character set. As the comparison moves to a level above the first, characters below this level are not considered, populations with a common set of character states are merged and their performances are pooled. Consequently, the populations we perceive may differ and different community resemblance values may arise. Following the example given above, two characters out of three will define the taxonomy on hierarchical level 2. Thus, on this level, the character order of the complete character set will determine whether the CST will be recognized as plants with thorny stems and broad leaves, plants with deciduous, broad leaves, or plants with thorny stems and deciduous leaves.

4.2. Optimal character order and optimal taxonomy

Optimization of character order is viewed as a process of automatic exploration of data structures. A recursive algorithm is devised to identify the less relevant characters and place them lower in the hierarchy. Character relevance is relative to the structures and relationships of interest, expressed by means of a specific *structural evaluation function*. One of these functions is described in Section 3.4. The approach is similar in spirit to successive approximation (Poore 1956, 1962, Orlóci 1990), but strictly analytical.

Having the optimal character order, profiles of the function are drawn (as in Orlóci and Mukkattu 1973, Orlóci and Pillar 1989). Probabilities of the function can be obtained by means of randomization (see Section 3.4). The optimal taxonomy, for the specified objective and a priori set of characters, will be the taxonomy defined at the hierarchical level that shows optimal value or probability for the structural evaluation function.

4.3. Character ranking criteria

4.3.1. Congruence with environmental structure

The congruence between community data structure and underlying environmental structure is used in Chapter 3 to compare results using crisp and fuzzy CSTs. Here congruence is used to rank characters. When the objective is to reveal the connections of community data structure with environmental data structure, ranking should establish a character order that maximizes the structural congruence on all levels.

We refer to Section 3.4 for some details. Accordingly, community structure $\mathbf{D}_i$, hierarchical level i, is defined by a $\nu \times \nu$ matrix of nominal dissimilarities between relevés. The underlying environmental structure is defined by a $\nu \times \nu$ matrix $\mathbf{\Delta}$ of relevé dissimilarities based on environmental variables. Using a recursive algorithm, starting at the highest level, the structural congruence function is the correlation

$$\rho(\mathbf{D}_i; \mathbf{\Delta}) \qquad (4.1)$$

This is defined here as a product moment correlation coefficient involving the $\nu(\nu-1)/2$ off-diagonal elements in $\mathbf{D}_i$ and $\mathbf{\Delta}$. At each level the task is to find the character amongst the i characters at and below level i that maximizes $\rho(\mathbf{D}_i;\mathbf{\Delta})$.

After ranking, it is expected that the characters at the top of the hierarchy will express sharper response to, or higher predictiveness of the environmental conditions, than expressed by the characters at the bottom. An important point is that based on characters so ordered, we are allowed to find compositional trends (in CST terms) that are explained by environmental variation. Related to these, whether the defining characters of the CSTs will express plant survival directly or properties correlated with some real survival characteristic not actually measured, is a question of considerable importance.

The example (Table 4.3.1.1) uses a subset of characters from the Elk Lake data (Tables 3.4.1 and 3.4.2). The environmental information is in Table 3.4.5. The highest rank is given to stem direction (di), for this is the one among the 8 characters that maximizes the congruence between community and environmental structures (0.534) when placed at the top of the hierarchy. At the next lower hierarchical level, the character stem consistency (co) is the one among 7 remaining characters that achieves maximum congruence (0.701) when placed on level 7. The same procedure for the following lower levels establishes the optimal order as being leaf length (le), plant height (he), growth-form (gf), leaf shape (sh), leaf epidermal surface (ep), until the character left is leaf width (wi).

The profiles in Fig. 4.3.1.1 illustrate the effect of character order. At hierarchical levels above level one, the taxonomies defined by a suboptimal character order reveal much weaker environmental congruence than do the ones defined by the optimal character order. The probability profiles (Sections 2.8 and 3.4) indicate congruence in an inverse sense. Considering the structural evaluation function and its probabilities, one could decide that the optimal taxonomy is defined by the character subset on hierarchical level 6 (Fig. 4.3.1.1a).

The algorithm is *agglomerative*: characters are added to the subset according to declining maximum congruence. However, a *divisive* algorithm could be adopted, in which case, starting on level one with the complete character set and moving to each subsequent level i, a character is dropped such that the remaining character subset results in maximum congruence $\rho(\mathbf{D}_i;\mathbf{\Delta})$. It is not expected that the two strategies will give identical ranking for the same data.

Table 4.3.1.1. Ranking of characters on the basis of decreasing congruence $\rho(\mathbf{D}_i;\mathbf{\Delta})$ between vegetational and environmental structures in a subset of the Elk Lake data (Tables 3.4.1, 3.4.2). The characters are growth-form (gf), stem consistency (co), stem direction (di), leaf shape (sh), leaf epidermal surface (ep), leaf width (wi), leaf length (le), and plant height (he). Vegetation structure $\mathbf{D}_i$ is defined as a matrix of distances $d^2_{i\alpha\beta} = 2(1 - r_{i\alpha\beta})$, where $r_{i\alpha\beta}$ are nominal correlation coefficients between relevés, using the defining character subset indicated in the table. Environmental structure $\mathbf{\Delta}$ is defined by relevé squared distances using elevation, exposure, slope, soil depth, and soil texture, after centering within variables and normalization. The asterisk indicates the character subset with maximum congruence. The final character ranking is di co le he gf sh ep wi.

$\rho(\mathbf{D}_i;\mathbf{\Delta})$	Order in character subset considered
0.524235	gf
0.445153	co
0.534043	*di
0.227612	sh
0.120588	ep
0.047395	wi
0.37611	le
0.257664	he
0.459472	di gf
0.70103	di *co
0.108908	di sh
0.128919	di ep
0.12804	di wi
0.604726	di le
0.431131	di he
0.480543	di co gf
0.441492	di co sh
0.326616	di co ep
0.211398	di co wi
0.580569	di co *le
0.486461	di co he
0.554961	di co le gf
0.544541	di co le sh
0.521539	di co le ep
0.33591	di co le wi
0.577934	di co le *he
0.537012	di co le he *gf
0.536784	di co le he sh
0.517702	di co le he ep
.308429	di co le he wi
0.537138	di co le he gf *sh
0.518702	di co le he gf ep
0.31246	di co le he gf wi
0.518772	di co le he gf sh *ep
0.312416	di co le he gf sh wi
0.336037	di co le he gf sh ep wi

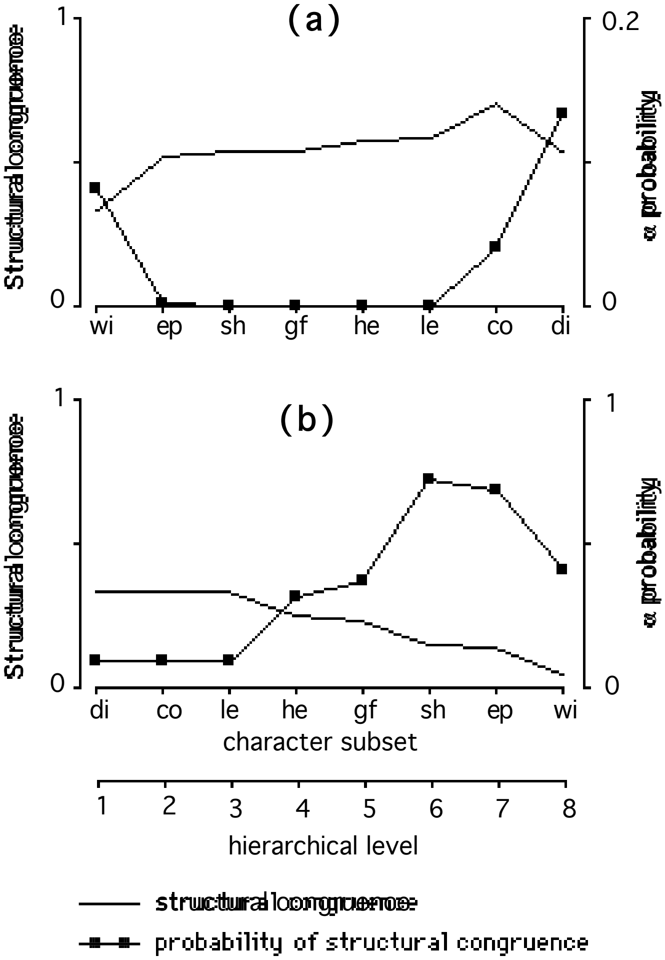


Figure 4.3.1.1 Profiles of congruence ρ($\mathbf{D}_{\mathrm{i}}$;Δ) between vegetational and environmental data structures for the Elk Lake data and the corresponding α probabilities. Vegetation and environmental data structures are defined as in Table 4.3.1.1. The character order in (a) is the optimal character order shown in Table 4.3.1.1 (di co le he gf sh ep wi). The character order in (b) is suboptimal (wi ep sh gf he le co di). The defining character subset on each level is cumulative from the top down the hierarchy. The character labels are identical as used in Table 4.3.1.1. Note that on all levels the optimal character order reveals higher congruence, with much lower α probabilities, than the suboptimal character order does. Maximum congruence with minimum probability would indicate the character subset on level 6 in (a) as optimal.

4.3.2. Structural convergence and divergence

The hypothesis of convergent evolution has been a topic in much of the ecological literature. Work by Cody and Diamond (1975), Orians and Solbrig (1977a), and Orians and Paine (1983) are typical examples. But in these, we see a lack of a more quantitative approach. Questions such as "how convergent" or "how significantly convergent" could have been addressed by quantitative methods as shown in this section. When we apply a character-based approach, it is expected that characters will converge to various degrees. If not isolated, characters that do not converge may obscure the expression of community convergence in the data if placed high in the hierarchy. There should be a character order that maximizes the manifested convergence in inverse proportions to the degree of structural dissimilarity between communities or groups of communities not explained by a common flora.

Divergence d is sometimes viewed as the lack of convergence, no matter whether the flora is common or not. Perhaps it would be better not to view divergence as a complement of convergence c, but as an independent dimension, $d^2 = 1 - c^2$, in the comparison space of the communities. The correlation coefficient (Eq. 2.13) is directly interpretable in terms of convergence (positive side), divergence (negative side), and lack of both (zero). In the case of dissimilarity functions, values that are extremely high indicate convergence, and extremely low values indicate divergence, but there is no natural point indicating the lack of both. We believe that this point can be found by randomization, as the middle point between probability limits on both extremes.

The iterative algorithm that follows evaluates different character subsets on each level with respect to the degree of dissimilarity values in $\mathbf{D}_i$, i=1, ..., m, between relevé groups. A correction factor applied to the resemblances to discount the dissimilarity reduction due to overlapping flora, used by Orlóci *et al.* (1986), is not needed here since it is constant over all iterations. The set of relevés is *a priori* partitioned into groups. Their potential convergence is maximal when the dissimilarities are minimal. Different functions can define the ranking criterion. In a selected case, which we shall call *average dissimilarity*, it accords with

$$\sigma_i = \frac{1}{\sum_{e=1}^{n-1} \sum_{f=e+1}^{n} \nu_e \nu_f} \sum_{j=1}^{\nu-1} \sum_{\substack{k=j+1 \\ g_j \neq g_k}}^{\nu} d_{ijk} \qquad (4.2)$$

This is minimized on successive hierarchical levels i in the course of iterations with different character subsets. σ_i is a stress function and defines an average dissimilarity between the n groups of relevés. Regarding other symbols, d_{ijk} is an element in the dissimilarity matrix $\mathbf{D}_i$, ν is the total number of relevés in the n groups, ν_e and ν_f are the sizes of group e and f, and g_j and g_k identify the group to which relevés j or k belong.

In another case, σ_i is a *average nearest neighbor dissimilarity.* Its use here is akin to that in a similarly named clustering algorithm, and is limited to cases where there are only two relevé groups. In this case, instead of taking the average dissimilarity between all the relevés of one group to *all* the relevés in other group, as in the average dissimilarity method, the structure evaluation function takes the average of the dissimilarities of all relevés of one group to their *closest* relevés in the other group. The quantity to be minimized is

$$\sigma_i = \frac{1}{\nu} \sum_{j=1}^{\nu} \quad \text{INF} \left[d_{ijk}, k=1, ..., \nu \text{ and } g_j \neq g_k\right] \qquad (4.3)$$

σ_i measures stress between the two groups on level i. INF is the lowest value function. All other symbols have already been defined. This ranking method is appropriate when the groups are heterogeneous. The algorithm is agglomerative in both cases. Starting from the top and moving down through the hierarchy, on each level i the character added to the subset is such that the value found for σ_i is minimum.

The example (Table 4.3.2.1) applies the algorithm to the Caatinga and Chaco data set. The algorithm finds the optimal character order (Table 4.3.2.2). Interestingly, the two criteria produced exactly the same character ranking. The same stress function is applied in another analysis (Table 4.3.2.3), but the divergence was maximized. Fig. 4.3.2.1 displays the structural evaluation function profiles. The same figure also shows the effect of optimal character order on the relevé trajectories in an eigenordination.

Table 4.3.2.1. Data of 6 relevés described by 25 CSTs from the Caatinga (Brazil) and Chaco (Argentina) vegetation. The characters are: (bt) biological type 6. cactoid, 7. other; (st) stem type 1. succulent, 2. herbaceous, 3. woody, 4. no stem; (fu) stem function 1. regular, 2. twin-purpose, 3. no stem; (at) armature type 1. thorn/spine, 2. none, 3. no stem; (tx) leaf texture in a scale 1(herbaceous) to 5(leathery) 6. plant leafless; (sh) leaf shape 2. linear, 3. other 4. plant leafless; and (ar) leaf arrangement 1. simple, 2. compound, 3. plant leafless.

							Relevés					
CSTs							Caatinga			Chaco		
bt	st	fu	at	tx	sh	ar	b3	b4	b7	a1	a4	a8
7	3	1	1	1	3	2	1	16	6	6	22	21
6	1	2	1	6	4	3	2	2	2	11	4	2
7	3	1	2	3	3	2	1	16	0	0	0	0
7	3	1	2	2	3	2	32	0	15	5	5	15
7	3	1	2	4	3	1	3	5	5	15	5	0
7	3	1	2	3	3	1	16	32	16	0	0	0
7	3	1	2	4	3	2	2	0	0	0	0	0
7	3	1	2	2	3	1	1	1	1	0	0	0
7	3	2	2	1	3	1	1	0	0	0	0	6
7	3	1	2	1	3	1	1	1	0	0	0	5
7	3	1	2	5	3	1	1	5	0	5	15	1
7	2	1	2	1	3	1	0	15	0	0	0	0
7	3	1	2	5	2	2	0	1	0	0	0	5
7	3	1	1	3	3	1	0	0	0	15	15	15
7	3	1	1	4	3	1	0	0	0	15	15	15
7	3	2	1	5	3	1	0	0	0	1	0	6
7	4	3	3	5	3	1	0	0	0	1	0	0
7	3	2	1	4	3	1	0	0	0	1	5	0
7	3	2	1	1	3	2	0	0	0	1	1	0
7	3	1	1	1	2	1	0	0	0	1	1	0
7	2	2	2	1	3	1	0	0	0	0	5	0
7	3	2	1	1	3	1	0	0	0	0	0	1
7	2	2	2	1	3	2	0	0	0	0	0	5
7	3	1	1	2	3	1	0	0	0	0	0	5
7	2	2	1	1	3	1	0	0	0	0	0	1

Table 4.3.2.2. Ranking of characters by the maximum convergence criterion. The relevé groups compared are from the Caatinga and Chaco (Table 4.3.2.1). The community data structure is defined by a matrix of nominal squared chord distances (Eq. 2.18) between relevés using different character subsets at different hierarchical levels, starting on level 7. The analysis uses crisp CSTs. Asterisks indicate the character set responsible for minimum average distance between the relevé groups on the hierarchical level considered. The final character ranking is bt sh st fu ar tx at.

σ_i	Character subset considered
0.005931	*bt
0.015619	st
0.033512	fu
0.816232	at
0.813563	tx
0.007005	sh
0.174443	ar
0.015619	bt st
0.029937	bt fu
0.805232	bt at
0.813563	bt tx
0.007005	bt *sh
0.174443	bt ar
0.017036	bt sh *st
0.033767	bt sh fu
0.828185	bt sh at
0.8139	bt sh tx
0.176464	bt sh ar
0.03513	bt sh st *fu
0.880767	bt sh st at
0.83422	bt sh st tx
0.206312	bt sh st ar
0.882241	bt sh st fu at
0.821455	bt sh st fu tx
0.225749	bt sh st fu *ar
1.08819	bt sh st fu ar at
0.857027	bt sh st fu ar *tx
1.4439	bt sh st fu ar tx at

Table 4.3.2.3. Ranking of characters by the maximum divergence criterion. The analysis proceeds as in Table 4.3.2.2, except that the asterisk indicates maximum average distance between the relevé groups. The final character ranking is at tx fu sh st bt ar.

σ_i	Character subset considered
0.005931	bt
0.015619	st
0.033512	fu
0.816232	*at
0.813563	tx
0.007005	sh
0.174443	ar
0.805232	atbt
0.85486	atst
0.867879	atfu
1.4384	at*tx
0.828185	atsh
1.01775	atar
1.4384	attx bt
1.45036	attx st
1.45304	at tx *fu
1.44695	at tx sh
1.41888	at tx ar
1.45304	at tx fu bt
1.45711	at tx fu st
1.46219	at tx fu *sh
1.43373	at tx fu ar
1.46219	at tx fu sh bt
1.46639	at tx fu sh *st
1.43373	at tx fu sh ar
1.46639	at tx fu sh st *bt
1.4439	at tx fu sh st ar
1.4439	at tx fu sh st bt ar

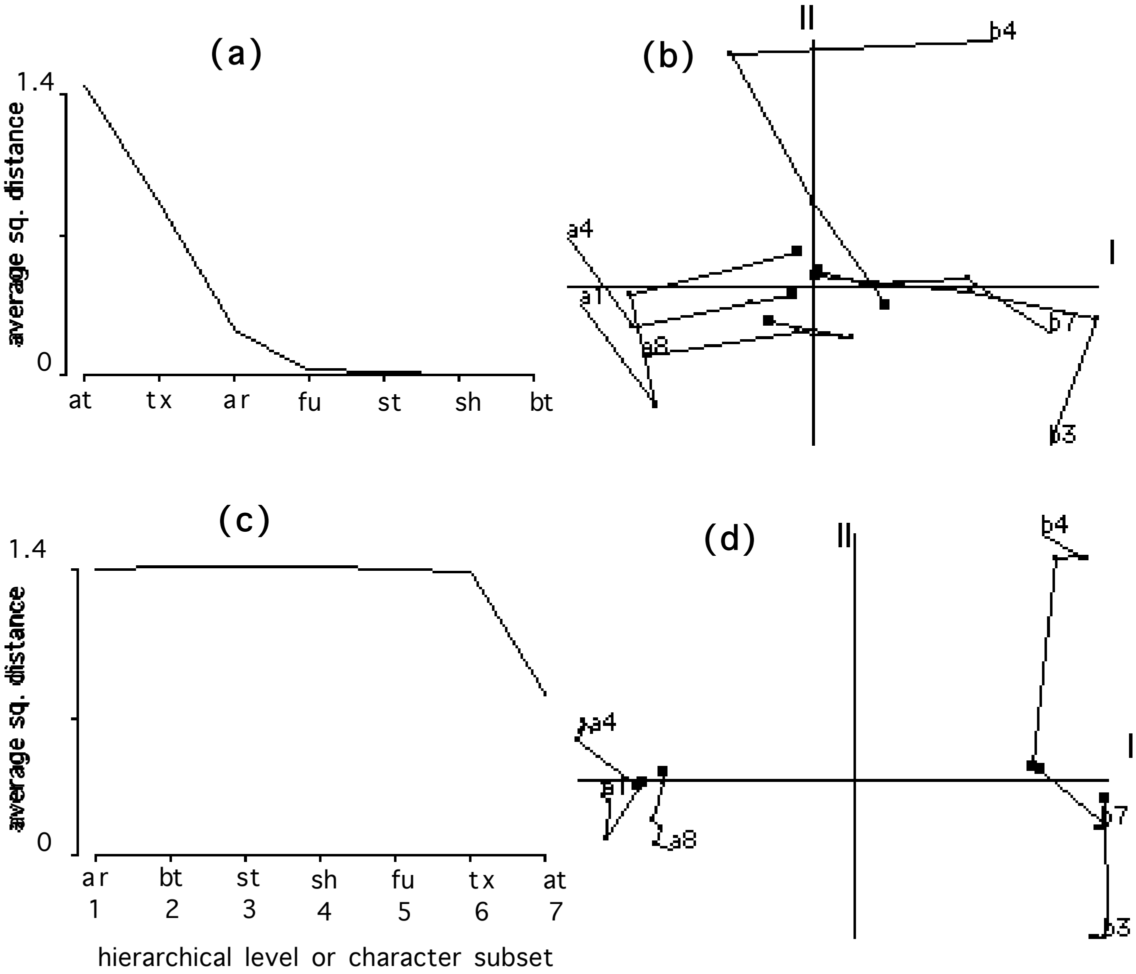


Figure 4.3.2.1. Comparison of the Caatinga and Chaco samples (Table 4.3.2.1) by profiles (a, c) and their relevé trajectories (b, d) through the hierarchical levels. The character order maximizes convergence (Table 4.3.2.2) in (a, b) and divergence (Table 4.3.2.3) in (c, d). The defining character subset on each level is composed of the characters specified on and above that level. The dissimilarity function is the nominal squared chord distance using crisp CSTs. The profiles are based on the average distance between the Chaco and Caatinga relevé groups. The ordination method is eigenanalysis. Components 1 and 2 are plotted in graphs b, d. The total variation accounted for by the individual axes is given in Table 4.3.2.4. The ordinations on the different hierarchical levels are made comparable by using the Procrustes method (Chapter 5). In the scattergrams, each relevé (a1 a4 a8 for the Chaco, and b3 b4 b7 for the Caatinga) is represented by points on a trajectory through 4 (b) and 7 (d) hierarchical levels. Points on hierarchical level 1 are labelled. The larger squares indicate the end of the trajectory. Note the appearance of increasing convergence (declining distance values between the Chaco and Caatinga relevés) in (a, b) in contrast with the one in (c, d) with increasing hierarchical level. This accords with the criteria adopted to optimize the character order. The degree and direction of the slope in the profiles is reflected partially in the extent to which the groups converge or diverge between hierarchical levels in the scattergrams (b, d). Also, note the identical average distance in a, c and relevé configuration in b, d on level 1.

Table 4.3.2.4. Partitioning of the total variation among the eigenvalues on each hierarchical level for the ordination of the Caatinga and Chaco data set (Table 4.3.2.1). Scattergrams are shown for the 1st and 2nd components in Fig. 4.3.2.1 (b, d). The analyses uses optimal character order for maximum convergence (Table 4.3.2.2) and maximum divergence (Table 4.3.2.3) between the Caatinga and Chaco relevés. On level 1 analyses find identical eigenvalues.

Level	% of total variation Component				
	1	2	3	4	5
	Analysis with optimal character order for maximum convergence (Fig.4.3.2.1b)				
1	55.37	25.20	12.65	4.87	1.91
2	48.23	28.62	14.19	6.56	2.40
3	73.27	14.83	8.88	2.45	0.57
4	52.34	31.13	15.23	1.25	0.05
5	71.66	24.07	4.22	0.04	
6	89.49	10.50	0.01		
7	99.98	0.02			
	Analysis with optimal character order for maximum divergence (Fig.4.3.2.1d)				
1	55.37	25.20	12.65	4.87	1.91
2	59.23	22.52	12.94	4.24	1.07
3	59.23	22.52	12.94	4.24	1.07
4	59.06	21.95	13.54	4.30	1.14
5	60.67	22.24	12.97	3.01	1.11
6	61.16	21.96	13.09	2.70	1.09
7	99.59	0.39	0.02		

4.3.3. Structural redundancy

Redundancy is the repetition of information by different community elements in the data (see, *e.g*., Sokal 1965, Jardine and Sibson 1971:171, Sneath and Sokal 1973:103). Redundancy has been used for ranking (see Orlóci 1975, 1978:17), in the sense that a redundant species, or character, can be removed from the analysis without drastically changing the population, or community structure in the data.

For comparison purposes, a population data structure is defined as a symmetric matrix $\mathbf{S}_i$ of resemblance values relating the k_1 CSTs defined on hierarchical level 1. Resemblances are computed on any level i using the $m-i+1$ characters on and above level i. That is, the dimension of matrix $\mathbf{S}_i$ is constantly k_1 x k_1 to allow comparisons between hierarchical levels; what changes is the character subset considered. Community data structure is defined as before, as a ν x ν symmetric matrix $\mathbf{D}_i$ of nominal relevé resemblances. With these definitions in mind, it appears intuitively correct that if the analysis is aimed at revealing data structures with a parsimonious number of characters that form the CSTs, there should be a character order established according to increasing redundancy from top to the bottom of the hierarchy. To achieve this, we offer an algorithm which rearranges the characters in an order which minimizes the changes in data structure on the consecutive levels of the hierarchy from the bottom up. The ranking should be such that the lower the level, the smaller is the difference between the data structures. It is important to note that, in this context, invariant characters are completely redundant, since they add no information.

The ranking algorithm is divisive. On hierarchical level 1 the data structure is unaffected by character order, thus the algorithm starts on level 2. On any level $i > 1$, there are

$$\frac{(m-i+2)!}{(m-i+1)!\,(m-i+2-(m-i+1))!} = m-i+2$$

different character subsets to form a taxonomy with $m-i+1$ characters, *i.e.*, with one character less than on level $i-1$. One of these subsets is found that maximizes redundancy with regard to the structure defined on level $i-1$, and the character removed from the subset is given the rank $i-1$. The remaining characters are carried to the next level. In other words, the question is reduced to finding the character amongst the remaining $m-i+2$ characters that should be dropped. This requires the evaluation of $m-i+2$ redundancy values.

4.3.3.1. Population level redundancy

Redundancy at the population level can be assessed based on the structural evaluation function

$$\rho(\mathbf{S}_i; \mathbf{S}_{i-1}) \tag{4.4}$$

This is conveniently defined as a product moment correlation on the off-diagonal elements in matrices $\mathbf{S}_i$ and $\mathbf{S}_{i-1}$. In fact, $\rho(\mathbf{S}_i; \mathbf{S}_{i-1})$ is also a measure of congruence (Section 4.3.1). $\mathbf{S}_i$ is the population data structure on level i and $\mathbf{S}_{i-1}$ is the population data structure on level $i-1$. The $(k_1 \text{ x } k_1-1)/2$ CST resemblances in

$\mathbf{S}_{i-1}$ are computed from the m–i+2 characters that define level i–1, while in $\mathbf{S}_i$ they are based on the m–i+1 defining characters on level i.

The data structure at the population level may also be defined as an m x m matrix of resemblances between m characters, in which case the method described by Orlóci (1975, 1978:17) and Rohlf (1977) to measure redundancy in species collections can be applied. However, when the set contains characters of different types, a covariance matrix between characters is undefined. Therefore, the method here described relies on resemblances between CSTs. For this, Gower's similarity index (Gower 1971) is used (Section 3.3). This index has the advantage of being able to handle mixed characters (dichotomous, qualitative, quantitative). The method does not preclude the use of other resemblance functions such as Goodall's (1966) probability index for which Goodall, Ganis and Feoli (1987) described algorithms.

Following the recursive algorithm explained above, the character eliminated at the ranking step on level i is the one that maximizes $\rho(\mathbf{S}_i; \mathbf{S}_{i-1})$. Therefore, at each level i, m–i+2 values of $\rho(\mathbf{S}_i; \mathbf{S}_{i-1})$ are compared. A similar method to rank species is mentioned in Orlóci and Mukkattu (1973), but it was considered computationally impractical at that time because of the large number of species involved.

An alternative approach can be sketched in terms of resemblance matrix comparisons on a specific hierarchical level. On hierarchical level 1, the pairwise structural evaluation function is defined as a matrix covariance of the CSTs' resemblance matrices $\mathbf{S}_{1p}$ computed on the basis of m–1 characters without character p and and $\mathbf{S}_{1q}$ computed without character q, for p = 1 , ..., m and q = p, ..., m characters in the set. The resulting m x m covariance matrix can be subjected to Orlóci's (1975, 1978:17) and Rohlf's (1977) ranking method.

The data structure defined by $\mathbf{S}_i$ is relevant only in the context of a fuzzy CST analysis (Chapter 3), because it is the data structure that defines grades of membership in the fuzzy sets of the CSTs. In the context of a crisp CST analysis, the population data structure should, instead, be defined by a matrix of similarities assuming 1 if for a given character subset the CSTs are completely identical, or 0 if they are otherwise. In this case it may happen that $\mathbf{S}_i = I$, *i.e.*, the data structure is an identity matrix. Since the correlation coefficient is undefined between identity matrices, another measure of congruence would have to be used instead of $\rho(\mathbf{S}_i; \mathbf{S}_{i-1})$ in the case of crisp CSTs.

The method outlined above for ranking based on structural redundancy is tested on data in Table 4.3.3.1 constructed to include a redundant character. The

CSTs are taken from the Caatinga and Chaco samples (Table 2.1.1), with the character "stem tissue" repeated at the end. As expected, the ranking algorithm gives the lowest rank to one of the copies of this character (Tables 4.3.3.2 and 4.3.3.3).

Table 4.3.3.1. Character score matrix taken from Table 2.1.1 with the character "stem tissue" (1) repeated (5).

CSTs	Characters					Caatinga	Chaco
#	1	2	3	4	5		
a	w	r	n	d	w	60	2
b	w	r	t	d	w	16	34
c	w	r	n	p	w	1	37
d	s	t	t	l	s	2	30
e	w	t	t	d	w	0	2
f	w	t	n	d	w	0	1

4.3.3.2. Community level redundancy

It may be more relevant to examine character redundancy in community level structures. This would take into account the influence of the character on the hierarchical accumulation of CST performances. Extreme cases of complete redundancy are illustrated by the example in Fig. 4.3.3.1. When a character is completely reproducing another character (Fig. 4.3.3.1 a, b), no accumulation on CST performances results after the character is eliminated. The same is also true for invariant characters (Fig. 4.3.3.1 c). In all these cases an identical set of CSTs is perceived no matter whether the redundant character is considered or not.

The objectives of the ranking algorithm are (1) to achieve reduction of the character subset to minimum size, one that still can discriminate the same set of CSTs, and within that subset, (2) to rank the characters based on their incomplete redundancies. It is interesting to note that there may be more than one answer which satisfies those conditions. In fact the algorithm can identify different character subsets in (1) and ranks in (2) depending on the initial character order.

Table 4.3.3.2. Similarity matrices of CST comparisons. The CSTs are described in Table 4.3.3.1. The similarity index (Gower 1971) applies to different subsets of four characters. Note that the matrices for character subsets 1, 2, 3, 4 (5 is out) and 2, 3, 4, 5 (1 is out) are identical and having the highest correlation with the matrix using all characters. Either character 5 or character 1 may be placed in the lowest rank. If character 5 is chosen, it is dropped from further consideration.

Level 1, all characters,
$\mathbf{S}_i$:

#	CSTs					
	a	b	c	d	e	f
a	1	0.8	0	0.8	0.8	0.6
b		1	0.2	0.6	0.6	0.8
c			1	0	0.2	0.4
d				1	0.6	0.4
e					1	0.8
f						1

Level 2, character subset 1, 2, 3, 4,
$\rho(\mathbf{S}_i; \mathbf{S}_{i-1}) = 0.9740$:

#	CSTs					
	a	b	c	d	e	f
a	1	0.75	0	0.75	0.75	0.5
b		1	0.25	0.5	0.5	0.75
c			1	0	0.25	0.5
d				1	0.5	0.25
e					1	0.75
f						1

Level 2, character subset 1, 2, 3, 5,
$\rho(\mathbf{S}_i; \mathbf{S}_{i-1}) = 0.9468$:

#	CSTs					
	a	b	c	d	e	f
a	1	0.75	0	1	0.75	0.5
b		1	0.25	0.75	0.5	0.75
c			1	0	0.25	0.5
d				1	0.75	0.5
e					1	0.75
f						1

Level 2, character subset 1, 2, 3, 5,
$\rho(\mathbf{S}_i; \mathbf{S}_{i-1}) = 0.9379$:

#	CSTs					
	a	b	c	d	e	f
a	1	1	0	0.75	0.75	0.75
b		1	0	0.75	0.75	0.75
c			1	0	0.25	0.25
d				1	0.5	0.5
e					1	1
f						1

Level 2, character subset 1, 3, 4, 5,
$\rho(\mathbf{S}_i; \mathbf{S}_{i-1}) = 0.9379$:

#	CSTs					
	a	b	c	d	e	f
a	1	0.75	0	0.75	1	0.75
b		1	0.25	0.5	0.75	1
c			1	0	0	0.25
d				1	0.75	0.5
e					1	0.75
f						1

Level 2, character subset 2, 3, 4, 5,
$\rho(\mathbf{S}_i; \mathbf{S}_{i-1}) = 0.9740$:

#	CSTs					
	a	b	c	d	e	f
a	1	0.75	0	0.75	0.75	0.5
b		1	0.25	0.5	0.5	0.75
c			1	0	0.25	0.5
d				1	0.5	0.25
e					1	0.75
f						1

Table 4.3.3.3. Results from ranking characters based on redundancy in the population structure. The algorithm is divisive. Table 4.3.3.2 explains part of the process. The asterisk in the first column indicates the character subset that is analyzed on the next higher level. Using subsets with 4 characters, the lowest rank is assigned to character 5. Using subsets with 3 characters the next lowest rank is assigned to character 1, and so on. The highest rank is assigned to character 4.

Level	$\rho(\mathbf{S}_i; \mathbf{S}_{i-1})$		Character subset	Character out	Characters ordered by rank
2	0.974011	*	1 2 3 4	5	5
2	0.946771		1 2 3 5	4	
2	0.937923		1 2 4 5	3	
2	0.937923		1 3 4 5	2	
2	0.974011		2 3 4 5	1	
3	0.883510		1 2 3	4	
3	0.879957		1 2 4	3	
3	0.879957		1 3 4	2	
3	0.904656	*	2 3 4	1	1 5
4	0.763763		2 3	4	
4	0.763763		2 4	3	
4	0.763763	*	3 4	2	2 1 5
5	0.666667		3	4	
5	0.666667	*	4	3	4 3 2 1 5

When redundancy is examined on the community level, a ν x ν symmetric resemblance matrix $\mathbf{D}_i$ of nominal relevé resemblances defines vegetation structure on hierarchical level i. In this, ν indicates the number of relevés in the data set. Resemblance functions under the hierarchical nested model of analysis are described in Chapter 2. As in the population case, at each level i, the character dropped is the one that maximizes the congruence

$$\rho(\mathbf{D}_i; \mathbf{D}_{i-1}) \qquad (4.5)$$

In this, ρ is a product moment correlation coefficient involving $\nu(\nu-1)/2$ off-diagonal elements in the half matrices $\mathbf{D}_i$ and $\mathbf{D}_{i-1}$. $\mathbf{D}_{i-1}$ is the matrix of nominal resemblances on level i–1. $\mathbf{D}_i$ is the matrix of nominal resemblance values on level i, with one of the m–i+2 characters of level i–1 removed from the set. On each level i, m–i+2 values of ρ are compared. For data sets with only two relevés the matrix correlation coefficient is undefined; in that case

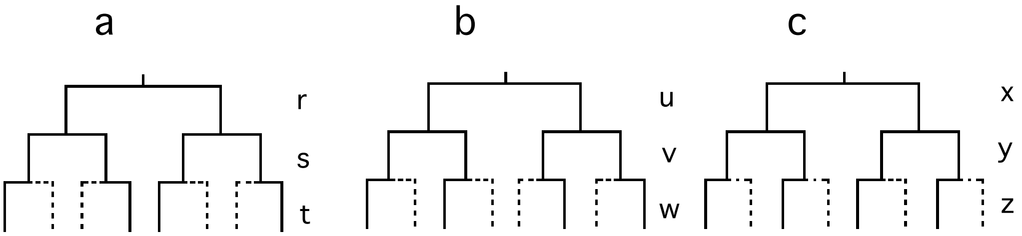


Figure 4.3.3.1. Hierarchical nested structures illustrating three cases of complete character redundancy. Letters on the right identify the binary characters. Position of the branches at each level (left or right) identify the character state. Dashed lines indicate that the state for that branch do not materialize in the data. In structure (a) character t is consistently reflecting character s and vice-versa; when s assumes its 1st state, t assumes its 1st state; when s assumes its 2nd state, t assumes its 2st state. The same is the case when t assumes its 2nd state and s assumes its 1st state, or t assumes its 1st state and s assumes its 2nd state. In structure (b) character w is consistently reflecting character u and vice-versa. In structure (c) character z is invariant. In the 3 cases the elimination of the character from the bottom level will not cause CST performance accumulation at level 2. Hence, no change in community data structure is perceived between level 1 and 2.

$$\rho(\mathbf{D}_i; \mathbf{D}_{i-1}) = 1 - \frac{|D_{i\alpha\beta} - D_{(i-1)\alpha\beta}|}{SUP(D_{i\alpha\beta}, D_{(i-1)\alpha\beta})} \tag{4.6}$$

where $D_{i\alpha\beta}$ and $D_{(i-1)\alpha\beta}$ are the dissimilarities between relevés α and β at levels i and i–1, and SUP is the largest value between them.

Two examples are considered. In the first, the ranking algorithm is applied to the Caatinga and Chaco data set in which one of the characters has been made redundant. The results in Table 4.3.3.4 show that the algorithm is capable of identifying the redundant character. Another example of redundancy-based ranking involves the Elk Lake recovery transect (Chapter 3). The results of this are displayed in Table 4.3.3.5. Clearly, 7 characters out of the 13 show complete redundancy. Therefore, the cover-abundance cumulants remain unchanged after reducing the character set to 6 of the least redundant characters.

A weighting factor can be incorporated in the structural evaluation function, in which case the expression becomes:

$$\rho(\mathbf{D}_{ih}; \mathbf{D}_{i-1})\, \omega_h \tag{4.7}$$

In this, $\mathbf{D}_{ih}$ is the community data structure when character h is removed from the set, and ω_h is a weight defined for character h. The structural evaluation

Table 4.3.3.4. Ranking of characters on the basis of redundancy in community level data structure in terms of the nominal squared chord relevé distance (Eq. 2.18) in the Chaco and Caatinga sample (Table 4.3.3.1). The character subsets are similarly defined as before on the different hierarchical levels (Table 4.3.3.3). Asterisks indicate the character subset carried to the next level up. Since there are only two relevés the structural evaluation function accords with Eq. 4.6. Using subsets of 4 characters, the lowest rank is assigned to character 5. Using subsets of 3 characters the next lowest rank is assigned to character 1, and so on. The highest rank is assigned to character 4. Note that on level 1, despite the same $\rho(\mathbf{D}_i; \mathbf{D}_{i-1}) = 1$ is found for the subsets having characters 5 or 1 out, the algorithm arbitrarily chooses the first subset. Therefore, the outcome is dependent on the initial order of characters.

Level	Sq. chord distances	$\rho(\mathbf{D}_i; \mathbf{D}_{i-1})$		Character subset	Character out	Characters ordered by rank
1	1.5820		*	1 2 3 4 5		
2	1.5820	1.0000	*	1 2 3 4	5	5
2	0.4201	0.2656		1 2 3 5	4	
2	0.7535	0.4763		1 2 4 5	3	
2	1.5407	0.9739		1 3 4 5	2	
2	1.5820	1.0000		2 3 4 5	1	
3	0.4201	0.2656		1 2 3	4	
3	0.7535	0.4763		1 2 4	3	
3	1.5407	0.9739		1 3 4	2	
3	1.5820	1.0000	*	2 3 4	1	1 5
4	0.4439	0.2806		2 3	4	
4	0.7535	0.4763		2 4	3	
4	1.5407	0.9739	*	3 4	2	2 1 5
5	0.5217	0.3386		3	4	
5	0.6921	0.4492	*	4	3	4 3 2 1 5

function is minimized in the way already explained. The weights can be based on congruence with the environmental structure, in which case each character weight is the one complement of the congruence produced when character h is placed at the highest hierarchical level m:

$$\omega_h = 1 - \rho(\mathbf{D}_{m|h}; \mathbf{\Delta}) \tag{4.8}$$

The weights can also be based on the divergence between community groups, in which case each character weight is the average or nearest neighbor dissimilarity under similar considerations:

$$\omega_h = \sigma_{mh} \tag{4.9}$$

Table 4.3.3.5. Ranking of characters on the basis of redundancy in community level data structure based on the Elk Lake recovery transect (see Tables 3.4.1 and 3.4.2). Community data structure is defined by matrices of correlation coefficients (Eq. 2.13) between relevés. Different character subsets are used on different hierarchical levels in accordance with the ranking procedure already described. An asterisk indicates the character most redundant (or the first of two or more identical maxima $\rho(\mathbf{D}_i;\mathbf{D}_{i-1})$). The corresponding character subset defines $\mathbf{D}_{i-1}$ on the next higher level. After each level, characters are listed whose removal produce identical $\rho(\mathbf{D}_i;\mathbf{D}_{i-1})$. The final character ranking is tx le he ep gf lf st co di lt sh wi th.

Level	$\rho(\mathbf{D}_i;\mathbf{D}_{i-1})$	Character out	Defining character subset
2	0.970105	he	lf gf st co di lt sh tx ep wi le th
2	1	* th	lf gf st co di lt sh tx ep wi le he
2	0.938647	le	lf gf st co di lt sh tx ep wi th he
2	0.928772	ep	lf gf st co di lt sh tx wi le th he
2	1	wi, tx, sh, lt, di, co, st, gf, lf	
3	0.970105	he	lf gf st co di lt sh tx ep wi le
3	0.938647	le	lf gf st co di lt sh tx ep wi he
3	1	* wi	lf gf st co di lt sh tx ep le he
3	0.928781	ep	lf gf st co di lt sh tx wi le he
3	1	tx, sh, lt, di, co, st, gf, lf	
4	0.970104	he	lf gf st co di lt sh tx ep le
4	0.938662	le	lf gf st co di lt sh tx ep he
4	0.928781	ep	lf gf st co di lt sh tx le he
4	0.577034	tx	lf gf st co di lt sh ep le he
4	1	* sh	lf gf st co di lt tx ep le he
4	1	lt, di, co, st, gf, lf	
5	0.970104	he	lf gf st co di lt tx ep le
5	0.938664	le	lf gf st co di lt tx ep he
5	0.92878	ep	lf gf st co di lt tx le he
5	0.577031	tx	lf gf st co di lt ep le he
5	1	* lt	lf gf st co di tx ep le he
5	1	di, co, st, gf, lf	
6	0.970103	he	lf gf st co di tx ep le
6	0.938669	le	lf gf st co di tx ep he
6	0.92878	ep	lf gf st co di tx le he
6	0.577025	tx	lf gf st co di ep le he
6	1	* di	lf gf st co tx ep le he
6	1	co, st, gf, lf	

Table 4.3.3.5. Continued.

Level	$\rho(\mathbf{D}_i;\mathbf{D}_{i-1})$	Character out	Defining character subset
7	0.9701	he	lf gf st co tx ep le
7	0.938684	le	lf gf st co tx ep he
7	0.928778	ep	lf gf st co tx le he
7	0.577007	tx	lf gf st co ep le he
7	1	* co	lf gf st tx ep le he
7	0.998727	gf	lf st co tx ep le he
7	1	st, lf	
8	0.970091	he	lf gf st tx ep le
8	0.928089	le	lf gf st tx ep he
8	0.928774	ep	lf gf st tx le he
8	0.576952	tx	lf gf st ep le he
8	1	* st	lf gf tx ep le he
8	0.99874	gf	lf st tx ep le he
8	1	lf	gf st tx ep le he
9	0.97005	he	lf gf tx ep le
9	0.928285	le	lf gf tx ep he
9	0.928753	ep	lf gf tx le he
9	0.576703	tx	lf gf ep le he
9	0.998784	gf	lf tx ep le he
9	0.999997	* lf	gf tx ep le he
10	0.96982	he	gf tx ep le
10	0.929343	le	gf tx ep he
10	0.928632	ep	gf tx le he
10	0.575361	tx	gf ep le he
10	0.98884	* gf	tx ep le he
11	0.930294	he	tx ep le
11	0.746523	le	tx ep he
11	0.941766	* ep	tx le he
11	0.636738	tx	ep le he
12	0.939803	* he	tx le
12	0.656544	le	tx he
12	0.506252	tx	le he
13	0.681073	* le	tx
13	0.408273	tx	le

As described, the weighting can overcome the problem of obtaining different results for the same data set if a different initial character order is used (Table 4.3.3.4). However, the ranking outcome will not reflect only redundancy, but a balance between minimal redundancy and maximal environmental congruence in one case (Eq. 4.8) or maximal structural convergence in another (Eq. 4.9).

4.4. Character order and fuzzy adjustments

The fuzzy approach to character-based analysis is described in Chapter 3. It is pointed out that the use of fuzzy CSTs can reduce indeterminacy and modify the vegetation data structure in such a way that it can improve the utility of the analysis in revealing structural connections with environmental variables. It is also pointed out that a fuzzy CST analysis is affected by varying the character order, but not as much as a crisp CST analysis. How does this come about? Answers are suggested by an inspection of the foregoing results:

1. Firstly, we believe, in a fuzzy CST analysis the character ordering is not so critical than it is in a crisp CST analysis. The reason is that by using fuzzy CSTs, the resemblances at lower levels reflect resemblances that in a crisp CST analysis would only appear on higher levels under optimal character order. This trend is in some extent apparent in Fig. 4.4.1, in contrast with results from a crisp CST analysis. However, it is important to note that in fuzzy CST analysis, the character set should contain characters that are deterministic of the structures to such a degree that their influence counterbalances other characters that tend to obscure the structures. Since these influences are not *a priori* known, it is well advised to rank the characters according to criteria depending on the structure sought.
2. If maximum divergence is of interest, and character order is optimized accordingly, fuzzy CST analysis will have the resemblances on the lower levels much more affected by the convergent characters in the set than in the crisp CST analysis. In crisp CSTs, if the higher hierarchical level characters are divergent, the whole hierarchy tends to be divergent. In fuzzy CSTs, this effect is not so striking; the higher level characters do not have so dramatic influence upon resemblances on the lower levels (Fig. 4.4.1c).
3. The same is not true when the analysis is aimed at maximum convergence. In this case, an analysis of crisp CSTs, with convergent characters present on the top levels of the hierarchy, may not reveal community convergence on the lower

levels. This will depend on how strongly convergent are the characters and, more importantly, how consistent is the convergence through the adjacent levels. An analysis with the same data, but with fuzzy CSTs, imposes on the lower levels a convergence, depending upon how strongly convergent are the characters on and above that level, regardless whether or not this trend is consistent through adjacent levels. This phenomenon can be observed in Fig. 4.4.1 (a, b).

4. The trends observed in the relevé trajectories depend on the optimality criterion adopted in character ordering. This given, divergence is more clearly revealed in the fuzzy CST analysis (Fig. 4.4.2).

5. Similar considerations apply when the analysis is aimed at maximum congruence with the environmental structure. A crisp CST analysis with optimal character order for this criterion reveals stronger structural congruence on the higher hierarchical levels. A fuzzy CST analysis can do the same also on lower levels. If the character order is suboptimal, the congruence revealed by a crisp CST analysis on the lower levels is much dependent, on (1) how strong is the characters correlation with the environmental variables, and (2) how consistent is between adjacent levels, from the top down through the hierarchy, the correlation. In the case of a fuzzy CST analysis, congruence on the lower levels depends only on condition (1) (Fig. 4.4.3).

4.5. Optimization and bias

There is an *a priori* decision of what kind of structures are sought. If, based on this, the characters that best fit the purpose are selected, the analysis would be "loaded". Regarding this, Sneath and Sokal (1973:95) observes: "when we use only a set of characters known to show resemblance between certain groups, the similarity coefficients that will result from this study will reflect that choice". Furthermore, there would be a circularity in the approach of optimization, paraphrasing Dale (1968), since the results support the character order and the character order produces the results.

The nature of the choice is different in the approach which we describe. Functions are specified by which structures are sought; the structures themselves are not specified *a priori*. This in no ways should be seen as a practice different from specifying the covariance as the measuring function in a MANOVA, or a generalized distance in discriminant analysis. Selection of a measuring function is the same as selecting a scale and a reference framework within which structures and relationships can be described. As for selecting characters and optimizing order is no different from deciding what we wish to measure and in what way we wish to view a relationship within a range of natural possibilities. These are no different from selecting variables and optimizing the axes (canonical axes) to maximize a distance function in the light of which we make a statistical decision. These choices cannot be avoided, and on making these choices we must come to terms with the fact that the results are context dependent.

Table 4.4.1. Partitions of the total variation among the eigenvalues on each hierarchical level in the ordination of the Caatinga and Chaco data (Table 4.3.2.1). The scattergrams for the 1st and 2nd components are shown in Fig. 4.4.2. The analyses use optimal character order for maximum convergence or divergence between the Caatinga and Chaco. On level 1 the analyses present the same eigenvalues.

Level	% of total variation Component				
	1	2	3	4	5
	Analysis with character order optimal for maximum convergence (Fig.4.4.2a)				
1	88.93	7.46	3.35	0.19	0.05
2	77.19	13.67	8.52	0.41	0.20
3	74.51	15.93	8.87	0.64	0.05
4	69.83	17.97	12.02	0.18	
5	78.24	20.31	1.44	0.01	
6	95.26	4.73	0.01		
7	99.98	0.02			
	Analysis with character order optimal for maximum divergence (Fig.4.4.2b)				
1	88.93	7.46	3.35	0.19	0.05
2	92.31	6.45	0.99	0.19	0.05
3	93.11	5.69	0.98	0.17	0.04
4	95.16	4.42	0.34	0.06	0.03
5	96.13	3.47	0.36	0.02	0.02
6	96.10	3.64	0.21	0.05	
7	99.59	0.39	0.02		

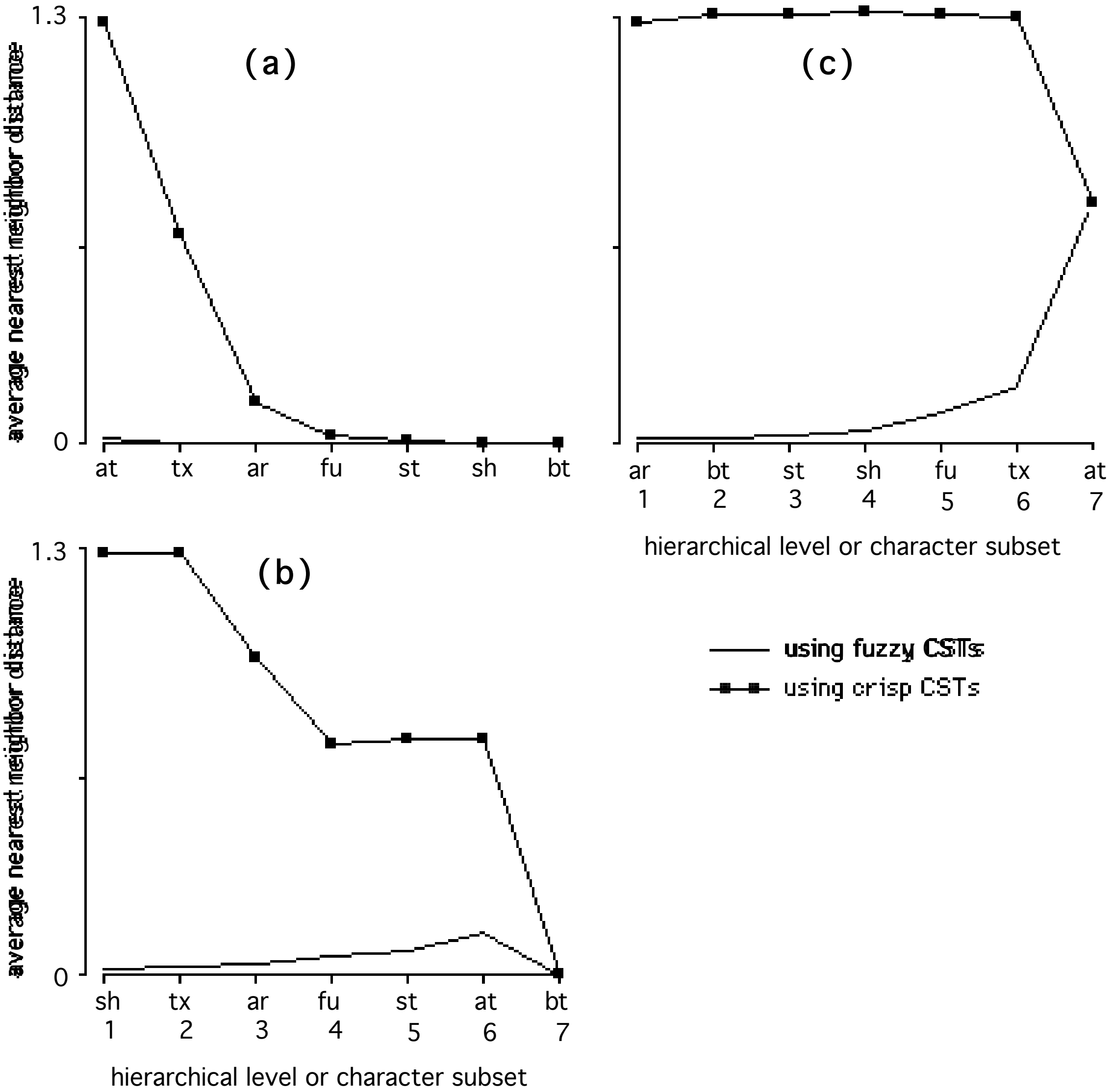


Figure 4.4.1. Profiles of average nearest neighbor distance between the Chaco and Caatinga relevé groups (Table 4.3.2.1). The character order is optimal in (a) and suboptimal in (b), for maximum convergence (Table 4.3.2.2). The order in (a) and (b) differ by the position of characters sh and at. In (c) the character order is optimal for maximum divergence (Table 4.3.2.3). The analysis uses crisp and fuzzy CSTs (global adjustment, Chapter 3). The dissimilarity function is the nominal squared chord distance. Note that for a suboptimal character order (b) the fuzzy CST analysis reveals more convergence (smaller distances) on the lower levels than the crisp CST analysis does. The vertical scale is the same in all profiles. Note the relative small distances in the fuzzy CST analysis with optimal character order for convergence (a). These distances are trended (see Fig. 4.4.2).

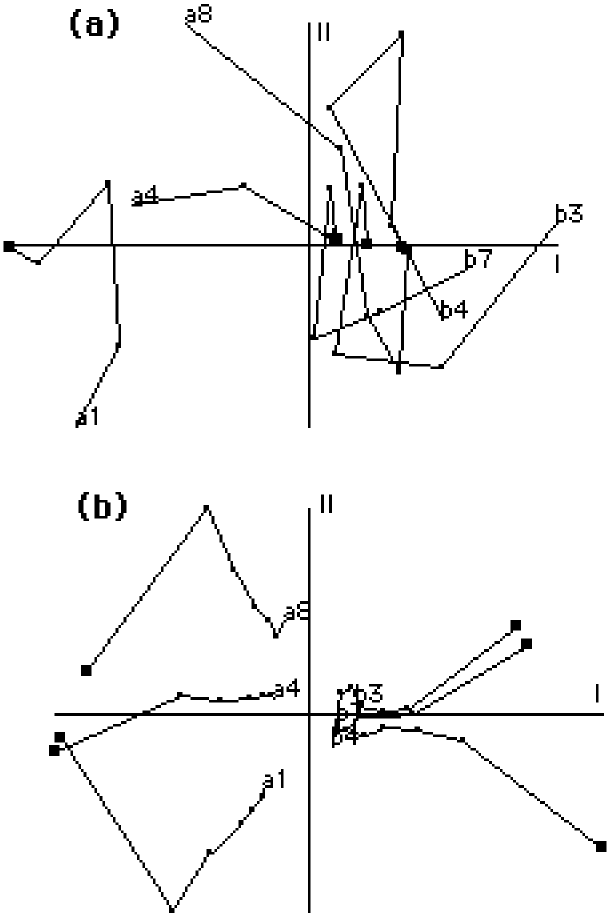


Figure 4.4.2. Comparison of the Caatinga and Chaco samples (Table 4.3.2.1) based on relevé trajectories through hierarchical levels in a fuzzy CST (global adjustment) analysis. Each relevé (a1 a4 a8 for the Chaco, and b3 b4 b7 for the Caatinga) is represented by points on a trajectory through hierarchical levels 1, 2, 5, 6, 7 in (a) and all 7 levels in (b). Points on hierarchical level 1 are labelled. The larger squares indicate the end of the trajectory. The character order maximizes convergence (Table 4.3.2.2) in (a), and divergence (Table 4.3.2.3) in (b). The dissimilarity function is the nominal squared chord distance. The ordination method is eigenanalysis. Components 1 and 2 are plotted in the graph. The total variation accounted for by the individual axes is given in Table 4.4.1. The ordinations on the different hierarchical levels are made comparable by using a Procrustes method (Chapter 5). Note the trend of increasing convergence in (a) with increasing hierarchical level, despite the relatively small distances revealed by the fuzzy CST analysis (Fig.4.4.1a). The opposite is observed in (b), which shows a stronger appearance of increasing divergence with increasing hierarchical level than in crisp CST analysis (Fig. 4.3.2.1d).

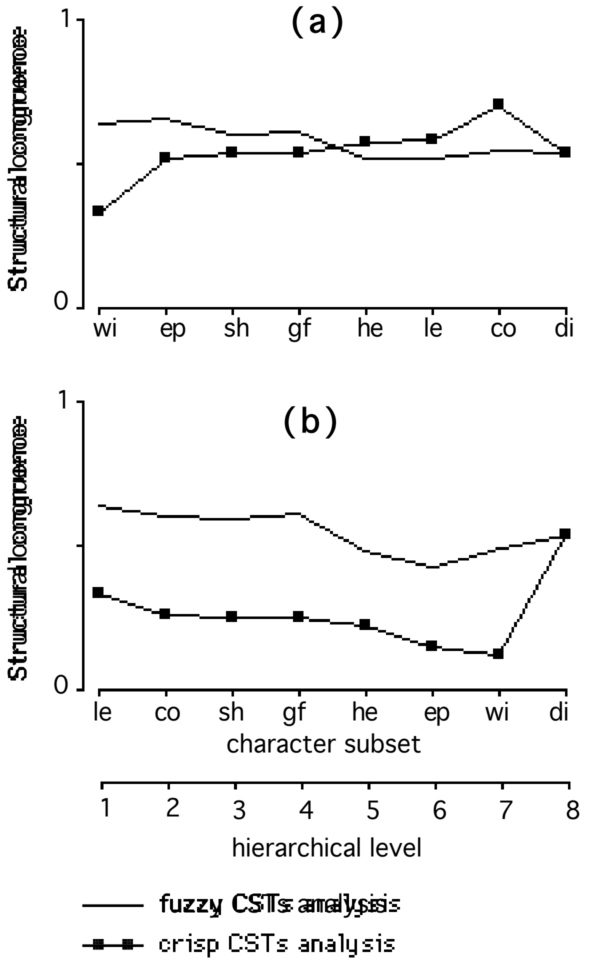


Figure 4.4.3. Comparison of fuzzy (pairwise adjustment) and crisp CST analysis, regarding the congruence $\rho(\mathbf{D}_i;\Delta)$ between vegetational and environmental structures for the Elk Lake data, using optimal (a) and suboptimal (b) character order. Vegetation and environmental structures are defined as in Table 4.3.1.1. The character order in (a) is the optimal character order according to the ranking shown in Table 4.3.1.1 (di co le he gf sh ep wi). The defining character subset on each level is cumulative from the top down the hierarchy. The character labels are identical as used in Table 4.3.1.1. Note in (b) the stronger congruence revealed by the fuzzy CST analysis on the lower hierarchical levels, despite the fact that characters with a low rank (ep and wi) are placed on levels 6 and 7.

Chapter 5

DATA EXPLORATION AND THE HIERARCHICAL NESTED MODEL

The linkage of the hierarchical nested scheme to exploratory techniques, particularly ordination, is addressed. In this scheme, the analysis is peculiar in having to deal with more than one hierarchical level simultaneously. As for ordination, a Procrustes method is described, which renders the spatial configuration on different hierarchical levels comparable. Examples are shown. Computations are performed by the application program SYNCSA.

5.1. The problem

Ordination and clustering techniques are well known tools of vegetation data exploration (cf. Orlóci 1978, Maarel 1980, Greig-Smith 1983, Orlóci and Kenkel 1985). The different methods require specific resemblance functions to generate the input. In the hierarchical nested model, ordination and cluster analysis can be applied on each hierarchical level. The community structures and structural connections revealed may vary according to the defining character subset on the hierarchical level considered. If CSTs are the objects in ordination or clustering, there is no point in comparing the different levels, since the CSTs are not the same. If, instead, relevés are the objects, it is relevant to compare the results on the different hierarchical levels. In this respect, the evolution of group structure in relevés can be traced through the hierarchy by comparing the classifications obtained on the different levels. In the latter case, the methods that compare partitions and dendrograms described in Podani and Dickinson (1984) and Podani (1986, 1989) are relevant. When ordination is the case, the study of trajectories of relevés through the different hierarchical levels in the ordination space is of interest. This requires commensurability between ordination scores on different hierarchical levels. As we see it, the problem amounts to obtaining a suitable method for the adjustment of the ordination scores without scrambling the relevé configuration.

5.2. A general eigenordination method

The method here described is a generalization of the Q-type and D-type PCA (Principal Components Analysis) techniques (Orlóci 1966, 1967b, Wildi and Orlóci 1990:66). It is applicable to almost any type of resemblance matrix, but its results coincide with the results of PCA only if the appropriate resemblance function is used. It is essentially the same method described as Principal Coordinate Analysis (Gower 1966, Sneath and Sokal 1973:248). Here we call it simply *eigenordination.* Eigenanalysis on each hierarchical level i is applied to a matrix $\mathbf{Q}_i$ of relevé similarities. To be consistent with a Q-PCA, it is required that the origin of the multidimensional space represented by $\mathbf{Q}_i$ be the centroid of the point cluster of relevés (Orlóci 1966). That is, data should be centered within attributes. When a distance matrix $\mathbf{D}_i$ is available, the centering condition is indirectly attained by applying the relation

$$q_{i\alpha\beta} = -\frac{1}{2}\left(d^2_{i\alpha\beta} - \frac{1}{\nu}\sum_{j=1}^{\nu} d^2_{ijk} - \frac{1}{\nu}\sum_{k=1}^{\nu} d^2_{ijk} + \frac{1}{\nu^2}\sum_{k=1}^{\nu}\sum_{j=1}^{\nu} d^2_{ijk}\right) \qquad (5.1)$$

where α, $\beta = 1, ..., \nu$ relevés, $i=1, ..., m$ levels, $q_{i\alpha\beta}$ is an element of matrix $\mathbf{Q}_i$, and $d^2_{i\alpha\beta}$ is a squared element of matrix $\mathbf{D}_i$. The dissimilarity measures based on information or probabilities (Sections 2.6 and 2.8) are not metric, but, as they are symmetric, they can be used in Eq. 5.1 as well.

As for some similarity functions, the matrices of centered cross products (Eq. 2.11) and of correlation coefficients (Eq. 2.13) are not suitable for a Q-eigenanalysis, since centering is within relevés. A centering within attributes, which is based on CST marginal totals, would affect hierarchical partition (Section 2.3). However, distances can be derived from these kind of products and coefficients by using

$$d^2_{i\alpha\beta} = 2(1 - r_{i\alpha\beta}) \qquad (5.2)$$

In this, $r_{i\alpha\beta}$ is a correlation coefficient (Eq. 2.13) The analysis proceed as for the $\mathbf{D}_i$ matrix. This procedure is used by Orlóci and Stofella (1986) in character-based analysis.

Matrix $\mathbf{Q}_i$ is subjected to eigenanalysis to find its eigenvalues λ_{ik}, ..., λ_{iv} and the corresponding eigenvectors $\boldsymbol{\beta}_{i1}$, ..., $\boldsymbol{\beta}_{iv}$. The eigenvectors are adjusted to satisfy the condition

$$\beta_{i1k}^{2} + \dots + \beta_{ivk}^{2} = \lambda_{ik} \tag{5.3}$$

Table 5.2.1. Nominal correlation coefficients between relevés described by crisp CSTs, and the derived **D** and **Q** matrices, for the Elk Lake data set (Tables 3.4.1 and 3.4.2). The defining characters are listed in Table 4.3.1.1, arranged in optimal order. Only hierarchical level 1 is shown.

	Relevés								
	1	2	3	4	5	6	7	8	9
Nominal correlations:									
1	1	0.820	0.644	0.469	0.605	0.135	0.317	0.398	0.364
2		1	0.708	0.347	0.471	0.0912	0.246	0.182	0.321
3			1	0.424	0.506	0.410	0.419	0.251	0.0244
4				1	0.720	0.529	0.684	0.723	0.0446
5					1	0.289	0.434	0.458	0.104
6						1	0.808	0.337	0.0200
7							1	0.337	0.0309
8								1	0.0432
9									1
Matrix $\mathbf{D}_1$:									
1	0	0.361	0.712	1.063	0.791	1.729	1.366	1.204	1.271
2		0	0.583	1.305	1.057	1.818	1.508	1.636	1.358
3			0	1.152	0.988	1.181	1.162	1.498	1.951
4				0	0.560	0.943	0.631	0.555	1.911
5					0	1.421	1.132	1.083	1.792
6						0	0.384	1.326	1.960
7							0	1.326	1.938
8								0	1.914
9									0
Matrix $\mathbf{Q}_1$:									
1	0.394	0.276	0.0784	-0.158	0.0166	-0.345	-0.236	-0.0947	0.0694
2		0.519	0.206	-0.217	-0.0537	-0.326	-0.245	-0.248	0.0886
3			0.475	-0.162	-0.0412	-0.0303	-0.0940	-0.201	-0.230
4				0.352	0.111	0.0272	0.110	0.209	-0.271
5					0.430	-0.173	-0.101	-0.0160	-0.173
6						0.645	0.380	-0.0295	-0.149
7							0.500	-0.103	-0.211
8								0.621	-0.138
9									1.016

The adjustment accords with

$$\beta_{ijk} = \beta_{ijk} \left(\frac{\lambda_{ik}}{\sum_{h=1}^{\nu} \beta_{ihk}^{2}} \right)^{\frac{1}{2}} \qquad (5.4)$$

for j, k = 1, ...ν relevés on each of i= 1, ... m levels. The adjusted eigenvectors are the ordination coordinates, which can be represented graphically in one, two or three dimensions.

The method is illustrated by example in Tables 5.2.1 and 5.2.2. The interpretation of the results follows standards conventionally used in quantitative ecology.

Table 5.2.2. Partition of the total variation among the eigenvalues for the ordination of the Elk Lake data set. The matrix **Q** subjected to eigenanalysis, level 1, is shown in Table 5.2.1. Only hierarchical levels 1, 2 and 3 are shown.

Level	Components							
	1	2	3	4	5	6	7	8
				Eigenvalues				
1	1.7119	1.1039	0.9702	0.4787	0.3003	0.1792	0.1266	0.08126
2	1.4196	1.2078	0.5434	0.2827	0.1691	0.122	0.09621	0.02758
3	1.4196	1.2078	0.5434	0.2827	0.1691	0.1220	0.09621	0.02758
				% of total variation				
1	34.57	22.29	19.59	9.67	6.06	3.62	2.56	1.64
2	36.70	31.22	14.05	7.31	4.37	3.15	2.49	0.71
3	40.35	24.93	14.76	8.32	4.34	3.46	3.01	0.83

5.3. Comparing different hierarchical levels

In this, the problem amounts to rendering the component scores of different hierarchical levels commensurable. The problem is not the scale of the scores, since the resemblance function and the ordination method is the same through the character hierarchy. If these were not the same, the Procrustes methods available would be appropriate (Schöneman and Carroll 1970, Gower 1975, Podani 1991, with applications in Bradfield and Kenkel 1987, Minchin 1987, Podani 1989). These involve a uniform expansion or contraction along component axes, a shift of the origin, and a rotation. The problem here, however, is related to eigenanalysis which yields an eigenvector $\boldsymbol{\beta}$ or its mirror image $(-1)\boldsymbol{\beta}$ arbitrarily. Both represent virtually the same relevé configuration, but the problem hinders the comparison of ordinations on different hierarchical levels. Accordingly, the Procrustes procedure to use involves only a 180° rotation, *i.e.*, the component axis on level i is taken as $\boldsymbol{\beta}_{ik}$ or $(-1)\boldsymbol{\beta}_{ik}$ in order to minimize

$$e_{i(i-1)} = \sum_{h=1}^{\nu} (\beta_{ihk} - \beta_{(i-1)hk})^2 \qquad (5.5)$$

for i=2, ..., m hierarchical levels, k=1, ..., ν ordination components. Note that if the component axis on level i is rotated, it remains rotated when the function is minimized on the next level i+1. An example is given in Table 5.3.1 and Figure 5.3.1.

Table 5.3.1. Example showing the application of the Procrustes method to ordination scores, as explained in the main text. The analysis uses the Elk Lake data set (Tables 3.4.1 and 3.4.2), with the defining characters listed in Table 4.3.1.1, arranged in optimal order. The resemblance function is the nominal correlation coefficient, using crisp CSTs (Table 5.2.1). The eigenvalues are given in Table 5.2.2. Only components 1 and 2 on hierarchical levels 1, 2 and 3 (column i) are shown. The asterisk indicates that the component axis was rotated ($e_{i(i-1)}$ is minimum in this case). The scattergrams are shown in Fig. 5.3.1.

	Relevés									$e_{i(i-1)}$		
i	1	2	3	4	5	6	7	8	9	β_{ik}	$-\beta_{ik}$	
					Component 1							
1	0.47	0.57	0.17	-0.37	-0.0060	-0.59	-0.51	-0.28	0.55	–	–	
2	0.39	0.34	-0.016	-0.051	0.14	-0.74	-0.64	0.29	0.29	0.65	5.6	
3	0.099	-0.094	-0.44	0.13	-0.018	-0.46	-0.44	0.61	0.62	4.8	0.84	*
					Component 2							
1	-0.20	-0.22	-0.37	-0.13	-0.29	0.27	0.13	-0.025	0.83	–	–	
2	-0.18	-0.43	-0.62	0.22	-0.12	0.078	0.0051	0.51	0.53	3.92	0.70	*
3	-0.27	-0.45	-0.28	-0.0067	-0.22	0.53	0.31	0.19	0.20	0.69	3.4	

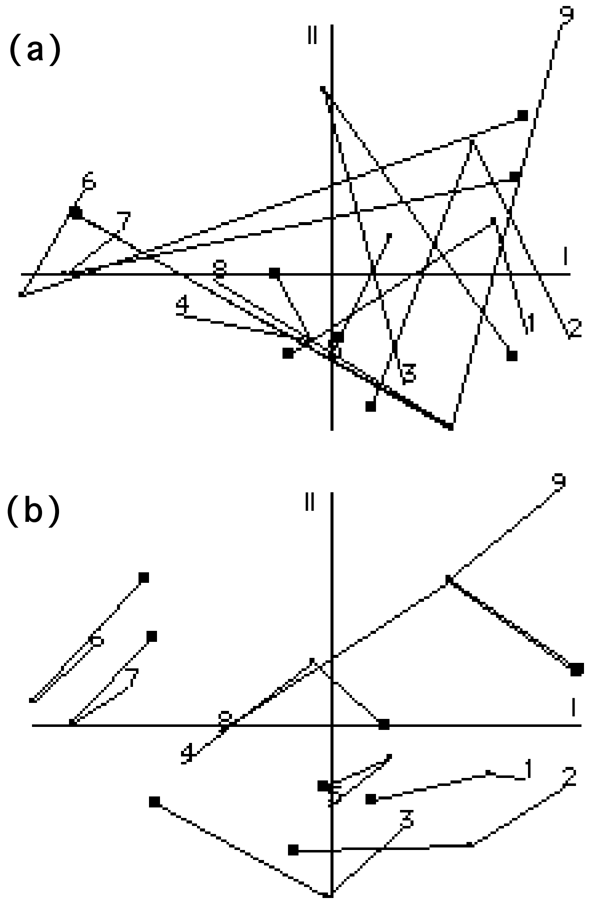


Figure 5.3.1. Eigenordination scattergrams for the analysis shown in Table 5.3.1 of the Elk Lake data set. The component scores in graph (a) are as generated by the eigenordination. In graph (b) the component scores are subjected to the Procrustes analysis (see the process in Table 5.3.1). Each relevé is represented by points on a trajectory through hierarchical levels 1, 2, and 3. Points on hierarchical level 1 are labelled. The larger squares indicate the end points of the trajectory. Eigenvectors 1 and 2 are plotted in the graph. Both graphs convey the same information, but graph (b) is more interpretable than graph (a).

Chapter 6

APPLICATION PROGRAM SYNCSA

6.1. General information

The application program SYNCSA[5] implements the analytical methods described in the previous chapters and also other methods not specific to the character-based approach. It is capable of performing an integrated analysis of community, population and environmental data structures. The code is written in C, most of it following the ANSI standard (Kernighan and Ritchie 1988). The current version is compiled and linked to run in Macintosh computers. The executable program file is less than 200 Kb in size, and is offered for users on a 3.5 inch disk along with sample data. The application is launched upon double-clicking SYNCSA's icon. A floating point coprocessor is strongly recommended, for which a specific version is available. Versions for other operational systems are planned.

Memory allocation is dynamic, thus there is no a priori upper limit for the number of characters, CSTs, environmental variables and relevés that can be handled. The default memory allocation size is 1000 Kb, which should be enough for medium size data sets, such as the examples in this book. If SYNCSA runs out of memory it will give a message and exit. To increase the memory allocated to the application, the Macintosh user should quit SYNCSA, select its icon in the Finder, choose Get Info in the File menu, and type a larger, compatible memory size in the memory box. In the Macintosh, SYNCSA can run in the background when under MultiFinder, that is, the user can operate on other applications while computations proceed in SYNCSA, which is specially

---

[5] Acronym for "Integrated Character Set Analysis"

useful for lengthy analyses, such as those involving randomization and character ranking.

The user interface is interactive, menu driven. Numbers and letters (upper or lowercase accepted) specify menu options. Major data entry is from text files. Keyboard input is processed upon pressing the Return key. Some questions may require multiple keyboard input, in which case it is enough to separate the entries by a blank space, a tab or a carriage return.

A session is the unit SYNCSA uses to group and connect information generated by a series of analytical steps performed on a data set. For this purpose several text files are created (see Table 6.1.1). These files can be opened and printed with any text editor, but not all of them are intended to show results, since specific text files hold printable outputs, such as file *Session*.prinda. In this regard, the text editor EDIT is automatically launched upon double-clicking any of the text files created by SYNCSA. The files can also be accessed by other applications, provided formats are compatible. Graphs generated by some procedures can be stored as picture files in the Macintosh hard disk by pressing the keys Shift Command 3 altogether while the graph is on the screen. With graphs on the screen, the user must close the graph's window in order to continue running the program.

6.2. Main menu

Upon starting SYNCSA and pressing the Return key, the main menu is shown on the screen (see Run 1). In the main menu the user can start a new session, continue analysis in an old session, create and run a macro, and set preferences if the default settings are not adequate.

Run 1:

```
CHARACTER-BASED COMMUNITY ANALYSIS                            SYNCSA v.1.0
--------------------------------------------------------------------------
Main menu:
   N  start new session
   O  attach to old session
   M  set macro mode
   S  set preferences
   X  quit
--------------------------------------------------------------------------
Type option:
```

Table 6.1.1. Files created by SYNCSA to store information associated to a session. The file name is the session's name followed by the extension. "Source" is the menu option that outputs to the file. Files of resemblances hold only the upper half of the matrices, including the diagonal. The appropriate sections in the main text give more explanations.

File extension	Information held	Source	Condition
formda	Formatted input data	New session	If raw data is formatted
track	Options the session has performed, and group partitions	New session	Any case
Files created in the analysis of community structures:			
prinda	Printable output	Several	Any case
ChRank	Optimal character order and the ranking process	Ranking characters	Any case
Hier	CST performances, by level, CST and relevé	Set data	If CSTs are crisp or fuzzy global
NRese	Matrices of nominal relevé resemblances, by level	Resemblance of relevés	Any case
PRese	Matrices of partial relevé resemblances, by level	Resemblance of relevés	If hierarchical partition is applicable
CstRese	Matrix of CST resemblances, one level	Resemblance of CSTs	Any case
Ro	Structure evaluation function (SEF), all levels	Evaluation	Not created for SEF opt. 5
ProbN	Matrices of probabilities of nominal relevé resemblances, by level	Randomization	Not created under the random partition null hypothesis
ProbRo	Probabilities of SEF, all levels	Randomization	If Ro is available, except for SEF opt. 2
Scores	Number of non-zero eigenvalues by level and relevé scores by level, eigenvalue and relevé	Ordination	Any case
Struct	Matrix of environmental relevé resemblances, matrices of relevé resemblances using CSTs or ordination axes, by level	Evaluation	For SEF opt. 1 or 2, and if selected in Preferences Menu
RndData	Random vegetation tables, relevé resemblance matrices and SEF, or random group partition and SEF, by iteration and level	Randomization	If selected in Preferences Menu
Files created in the analysis of environmental structures:			
ePrinda	Printable output	Several	Any case
eRese	Matrix of relevé resemblances	Resemblance	Any case
eRank	Environmental variables ordered by rank and the ranking process	Ranking variables	Any case
eScores	Number of non-zero eigenvalues and relevé scores, by eigenvalue and relevé	Ordination	Any case

6.3. Start a new session

By entering "n" in the main menu a name and pathway for the new session is asked. The new session in the example (Run 2, line 2) is named "ufrgs15re12ch". This means the session's files (Table 6.1.1) will have the prefix "ufrgs15re12ch." in their standard names. The pathway (line 3) specifies the volume and existing hierarchy of folders where the session's files are to be found or created. Enter ":" if the session is to be in the same folder with SYNCSA. In the example the pathway entered is "hd:csa", that is, the folder "csa" in the volume "hd". Volume and folder names are separated by colons and may contain any characters, except colons. Session names handled by SYNCSA must also not contain blank spaces.

A session requires a specifically formatted text file named *Session*.formda, which holds input data needed for the analyses. This file is created by SYNCSA using input from keyboard and from text files containing character descriptions (score matrices), CST performances in the relevés, and, optionally, species composition and environmental information, different formats accepted. The new session started in Run 2 reads raw data from file "ufrgs15re12ch.dat" (line 17), which is shown in Appendix A. The raw data file must be in the same session's folder only for this procedure. The program then asks how the raw data file is arranged. In the example, a relevé is a score matrix, with performance values following each CST (line 24); the score matrices are entered by CSTs, *i.e.*, CSTs are rows and characters are columns (line 28); the performances are one-character Braun-Blanquet codes (line 32), in which case code replacement values are informed (line 35). Entries in the raw data file may be separated by blank spaces, tabs, carriage returns, or a combination of these, except species names, which must be separated by carriage returns. Note the input for code replacement values is multiple. It is asked whether the multiple input for the last question is correct (line 36); if incorrect, enter "N" and you can type the values again. Instead of cover-abundance codes, CST performances may be quantities (option 2 line 32) expressed by any real number.

If option 2 is chosen in line 24, the raw data file will have to contain two matrices. The first matrix is a common score matrix for all relevés; the second is a performance matrix. The user then will be asked to specify what is in rows and what is in columns (CSTs or characters, CSTs or relevés). If option 3 is chosen in line 24 instead, the raw data file will have to contain a CSTs by relevés (or relevés by CSTs) performance matrix without the score matrix at the beginning, since in this case it is assumed the CSTs are defined by only one character. This option is specially useful for species-based descriptions. The data file "ufrgs15reSp" in Appendix C is an example.

Next, the program asks the number of characters (line 38) and their labels (line 42). In the example the user specifies labels through the keyboard; otherwise the program will attach ordered numbers as labels. The labels entered may contain up to two of any alphanumeric symbol (line 45). Hereafter these labels will identify the characters. The characters are then presented and the user specifies the character type (line 49). Dichotomous characters have only two states, such as presence/absence. Nominal characters are qualitative, with any number of unordered states, such as shape. The number of states and their labels, which may contain up to two of any alphanumeric symbol, will be searched by the program while reading the raw data file. Ordered characters are quantitative, with states limited to integers in the 0-99 range.

The user then specifies the number of relevés (line 52) and their labels (line 56). Labels of relevés may contain up to three alphanumeric symbols (line 59). The number of CSTs in each relevé (line 63) is needed only when option 1 in line 24 is chosen; if option 2 or 3 is selected instead, only the total number of CSTs in the table will be asked.

Species names corresponding to the CSTs may optionally be included in file *Session*.formda (line 66). Species names may be listed at the end of the file already informed, as in the example (Run 2 line 70), or listed in a separated file placed in the same folder. In the latter case the user will have to inform the file name. Species names in either files must be separated by carriage returns and listed in the same order in which the corresponding CSTs are listed in the score matrix or matrices. See sample data in the Appendix. Species labels may be the first 4-character word of the line containing the species name, or be created as in the example (line 74). Species affiliation may be added automatically as an additional character to the character set (line 75). If this option is checked "y" in the example, the number of characters will become 13.

Environmental variables may also be included in file *Session*.formda (line 77). It is assumed the variables are quantitative. The number of variables (line 78), labels (lines 82 and 85), file name containing the environmental information (line 87), and how it is arranged (line 88) are interrogated. The file must be in the current session's folder during this procedure (see Appendix B). Data are entered by relevés when relevés are the rows and environmental variables are the columns, and by variables when reversed. Appendix D shows the formda file created by Run 2.

Run 2:

```
Type option: N
Enter session name: ufrgs15re12ch
Enter pathway:       hd:csa


DATA FOR SESSION                                          ufrgs15re12ch
-----------------------------------------------------------------------
Options:
   1  format raw data
   2  use formatted data from another session
   3  return to main menu
Type option no.: 1


GETTING RAW DATA                                          ufrgs15re12ch
-----------------------------------------------------------------------
Enter data file name: ufrgs15re12ch.dat
Data format:
   1  A score matrix per relevé, each CST followed
        by a performance value
   2  A common score matrix for all relevés with
        performances matrix attached
   3  A performance matrix (no score matrix)
Type option no.: 1
Score matrix format:
   1  CSTs as columns / characters as rows
   2  CSTs as rows / characters as columns
Type option no.: 2
Type of CST performance data:
   1  One-character codes
   2  Quantities
Type option no.: 1
Code replacement:
   Code:  R       +     1     2     3     4     5
   Value: 1       2     3     5     7     8     9
Are these correct? y/n y
-----------------------------------------------------------------------
Enter number of characters: 12
Labels for characters:
   1  given by default
   2  specify new labels
Type option no.: 2
Enter labels (max. 2 letters, digits or combination):
 1  2  3  4  5  6  7  8  9 10 11 12
lf g1 g2 g3 st cr tx ed ev ve wi he
Are these correct? y/n y
Enter character type (1 for dichotomous, 2 for nominal, 3 for ordered):
lf g1 g2 g3 st cr tx ed ev ve wi he
2  2  2  2  2  2  3  2  2  2  3  3
Are these correct? y/n y
-----------------------------------------------------------------------
Enter number of relevés:     15
Labels for relevés:
   1  given by default
   2  specify labels
Type option no.: 2
Enter labels (max. 3 letters, digits or combination):
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
```

```
59    a   b   c   d   e   f   g   h   i   j   k   l   m   n   o
60   Are these correct? y/n y
61   Number of CST in each relevé:
62   a   b   c   d   e   f   g   h   i   j   k   l   m   n   o
63   8   11  14  5   10  12  15  14  7   12  6   12  10  7   5
64   Are these correct? y/n y
65   ------------------------------------------------------------
66   Include species names? y/n y
67   Species names are :
68      1  at the end of file ufrgs15re12ch.dat
69      2  in a separate file
70   Type option no.: 1
71   Species labels are:
72      1  the first 4 characters preceding each species name
73      2  to be created by the program
74   Type option no.: 2
75   Include species as an additional character? y/n n
76   ------------------------------------------------------------
77   Include environmental variables? y/n y
78   Number of environmental variables: 16
79   Labels for environmental variables:
80      1  given by default
81      2  specify labels
82   Type option no.: 2
83   Enter labels (max. 3 letters, digits or combination):
84     1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
85    pH   P   K   c  Al  Ca  Mg   S  Zn  Cu   B  Mn  Fe  hu   e  gr
86   Are these correct? y/n y
87   Environmental variables filename: ufrgs.env.dat
88   Data entered by relevés(N) or by variables(T)?: t
```

Alternatively, the new session can use another session's formatted data file (Run 3, line 12). The program asks the name of the session from which to get the formatted data (line 13). If the imported *Session*.formda file is taken unchanged (option 1, line 19), the program will jump to the Session Menu and the file will have to be kept in the folder in order to be accessed by other procedures in the current session. Otherwise, as in the example (option 2, line 19), the imported file remains unchanged and can be moved out afterwards, and a new file is created based on information the user enters specifying whether the set of characters (line 27), CSTs (line 36), relevés (line 42) and environmental variables (line 51) are to be modified. If so, as for the characters and relevés in the example, the user will tell how many (lines 28 and 43) and which elements (lines 30 and 45) are to be included in the new set. Identical CSTs are pooled, a situation which may arise when the new set contains less characters than the original one. CSTs do not have labels; in case you whish to select a subset of them, identify them by integers reflecting their order in the source formatted file. Note that the newly formatted data set will be automatically trimmed of CSTs and species names that are absent in the chosen subset of relevés. Also, a warning message will appear if the data contain relevés that are empty for the chosen subset of CSTs. Relevés may be pooled according to relevé groups (option 3, line 42), in which case the group partition will be asked (see Section 6.6.3); a relevé in the new file *Session*.formda will be tagged with the respective group's label and have the group's pooled CST composition.

Run 3:

```
1   Type option: N
2   Enter session name: ufrgs5re5ch
3   Enter pathway:      hd:csa
4
5
6   DATA FOR SESSION                                                  ufrgs5re5ch
7   -----------------------------------------------------------------------------
8   Options:
9      1  format raw data
10     2  use formatted data from another session
11     3  return to main menu
12  Type option no.: 2
13  Enter previous session name: ufrgs15re12ch
14
15  Use file ufrgs15re12ch.formda:
16     1  as it is
17     2  modified
18     3  return to main menu
19  Type option no.: 2
20
21
22  GETTING DATA FROM ANOTHER SESSION                                 ufrgs5re5ch
23  -----------------------------------------------------------------------------
24  Characters:
25     1  keep the same
26     2  modify
27  Type option no.: 2
28  Enter new number of characters: 5
29  Identify the characters by their labels:
30  g3 cr g1 st wi
31  Are these correct? y/n y
32  -----------------------------------------------------------------------------
33  Character set types (CSTs):
34     1  keep the same
35     2  modify
36  Type option no.: 1
37  -----------------------------------------------------------------------------
38  Relevés:
39     1  keep the same
40     2  modify
41     3  pool according to groups
42  Type option no.: 2
43  Enter new number of relevés:    5
44  Identify the relevés by their labels:
45    a   b   c   m   k
46  Are these correct? y/n y
47  -----------------------------------------------------------------------------
```

```
Environmental  variables:
   1  keep the same
   2  modify
Type  option  no.:  1
```

6.4. Attach to an old session

The user can attach to an old session and continue analysis from the step reached previously in the session. On selecting option O (letter O) in the main menu the user is asked to enter the session's name and pathway. If file *Session*.track is found in the informed folder the attachment is successful, but other session files (see Table 6.1.1) may also have to be found later depending on the procedures selected.

6.5. The session menu

After the user creates a new session or attaches to an old one the Session Menu is presented (Run 4). Analysis of community, population and, if pertinent, environmental data structures can be performed. Community, or vegetation structures are sought in each level's compositional matrix of CSTs by relevés. This option is the most relevant one in the context of character-based analysis. The other two options are accessory: population data structures, which are sought in the matrix of CSTs by characters, and environmental structures, which are sought in the matrix of environmental variables by relevés. The latter matrix is only defined when the environmental information is part of file *Session*.formda. Some options (evaluation, ranking), however, may take information from more than one of these matrices. The user can also access the Preferences Menu (see Section 6.9) or return to the main menu. The name of the current session is always informed at the upper right corner of the menus. The options available in the Session Menu are explained in the sequel.

Run 4:

```
SESSION MENU                                                   ufrgs5re5ch
--------------------------------------------------------------------------
Options:
   P  analysis of populational structures
   V  analysis of community structures
```

```
6      E  analysis of environmental structures
7      S  set preferences
8      C  return to main menu
9   Type  option:  V
```

6.6. Analysis of community structures

The menu offers the options that are active at the current stage of the analysis in the session, as shown in the flow chart (Figure 6.6.1). For instance, resemblance computations requires information produced by data settings; randomization, evaluation, ordination and cluster analysis are based on resemblance matrices; scattergrams depend on ordination scores. In the example (Run 5), no analysis has yet been performed in session "ufrgs5re5ch", thus only the options "set data", "ranking characters" and "specify relevé group partition" are offered (lines 4-6).

Depending on the procedures that are followed, text files are created for use by the program, such as those containing resemblances, ordination scores, probabilities, and structure evaluation function values (see Table 6.1.1). File *Session*.prinda, or file *Session*.chRank when option K is picked, holds the printable output. After a procedure is complete, the program informs where the print-

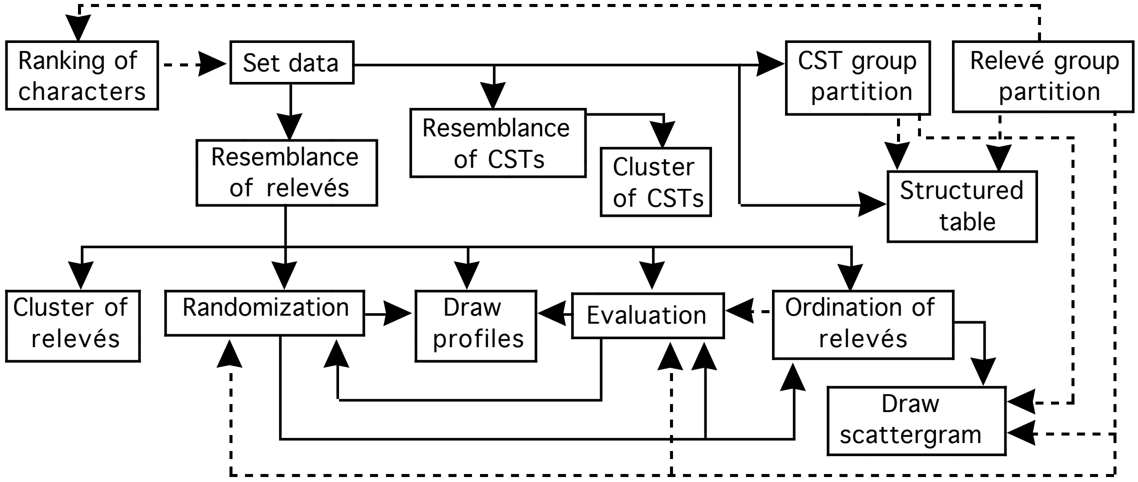


Figure 6.6.1. Flow of information in the analysis of community structures performed by SYNCSA. Full lines indicate the information produced by the procedure of origin is a sufficient condition for the pointed procedure be active in the menu. For instance, "Draw profiles" depends on information from "Resemblance of relevés", "Randomization" or "Evaluation". Dashed lines indicate the information produced by the procedure of origin may determine addition possibilities within the pointed option. For instance, a relevé group partition allows character ranking for maximum group convergence or divergence.

out is placed. Typing the Return key will bring the updated menu.

Because a session can hold only one file of a kind, results of old runs in the same session and for the same option in community structures menu are overwritten, except file *Session*.prinda, which is always appended when analysis is beyond the Set Data procedure. For instance, if option R is selected when it has been selected before in the same session, the old file containing resemblances is lost and subsequent analyses that depend on resemblances will use the new information. This limitation may also affect other options in the session. That is, once new information is produced at a given point in the flow chart (Figure 6.6.1), this point is set as the level reached by the analysis, and information produced by old runs of options beyond that point will be disregarded. If you wish to retain a session intact for future use, while trying alternative analytical pathways, start a new session with the same formatted data (see Run 3, line 19).

6.6.1. Set data

On choosing this option, the program prepares vegetation tables (CSTs by relevés) corresponding to the levels of the character hierarchy, which will base subsequent analyses. It starts asking the data partition type (see Section 2.7). The example (Run 5, line 17) uses mixed data. Next, it asks whether CSTs are to be taken as crisp or fuzzy (see Chapter 3). In the example (Run 5, line 23) the option for no adjustment (crisp CSTs) is followed. If an option for fuzzy CSTs (global or pairwise) is selected instead, the degree of fuzziness will also be asked.

The character order can be specified. By convention the first character in the list is at the highest level of the character hierarchy. That is, at the highest level the CSTs are solely defined by the first character in the list; at the lowest level they are defined by the complete character set. The characters may be taken as ordered in the raw data (opt. 1), as entered by the user through the keyboard (opt. 2), or as ordered in a file generated by character ranking procedures (opt. 3). In Run 5 the character order is specified (line 32). If it is taken from a ChRank file (see Run 9, line 33), the user will be asked to specify the session from which the file ChRank is to be taken, which can be the current or other session (in any case the file must be in the same folder with the current session).

The user can select (see Run 5, lines 35-37) whether the printout file *Session*.prinda is to contain vegetation tables, species names corresponding to the CSTs, and indeterminacy index matrices (Eq. 3.1). When the CSTs are fuzzy, the fuzzy sets (CST similarities) and adjusted vegetation tables may also be in the printout. Keep in mind that pairwise fuzzy adjustments yield a vegetation table for each relevé pair. The printout file of Run 5 is shown in Output 1. Note the species/CST correspondence. First, the species are listed by code and name. Second, at each hierarchical level, the CSTs are listed in the same sequence as in the vegetation tables, with the respective species codes to which the CST may belong attached. For instance, CST #3 level 5 may belong to species *Axonopus affinis* (Axaf) or *Centella biflora* (Cebi). Note also that a species may correspond to different CSTs, *e.g.*, CSTs #1 and #18, level 1, belong to *Andropogon lateralis* (Anla).

Run 5:

```
ANALYSIS OF COMMUNITY STRUCTURES                                  ufrgs5re5ch
------------------------------------------------------------------------------
Options:
   H  set data
   K  ranking characters
   G  specify relevé group partition
   C  return to session menu
Type option: H


DATA SETTINGS                                                     ufrgs5re5ch
------------------------------------------------------------------------------
Data type:
   1  mixed
   2  quantitative
   3  qualitative
Type option no.: 1

Fuzzy set adjustment:
   0   no adjustment (crisp CSTs)
   1   fuzzy CSTs global adjustment
   2  fuzzy CSTs pairwise adjustment
Type option no.: 0

Character order:
   1  as in data
   2  user specified
   3  from file ChRank
Type option no.: 2

Enter characters, identified by labels, in the order desired:
g3 cr g1 st wi

Printout:
      - vegetation tables        y/n   y
      - species names            y/n   y
      - indeterminacy index      y/n   y
```

38 `Results on file hd:csa:ufrgs5re5ch.prinda`

Output 1:

```
CHARACTER-BASED COMMUNITY ANALYSIS                            SYNCSA v.1.0
--------------------------------------------------------------------------
EXPLORATION OF VEGETATIONAL DATA STRUCTURES
--------------------------------------------------------------------------
Tue Jul 21 11:53:14 1992
Session:             ufrgs5re5ch
Formatted data:      ufrgs5re5ch.formda
Unformatted data:   ufrgs15re12ch.dat

Character set:
Order:              g3 cr g1 st wi
Number of states: 3  3  3  3  4
Character type:   2  2  2  2  3

Number of relevés: 5
Relevé labels: a   b   c   m   k
Fuzzy transformation: none (crisp)
Data type: mixed

Contingency table level 1:
CST   g3    cr    g1    st    wi     a     b     c     m     k
---   --    --    --    --    --    ---   ---   ---   ---   ---
1      3     2     3     3     2     5     3     2     5     0
2      2     1     1     3     2     5     0     7     0     0
3      3     3     3     3     1     5     2     3     2     0
4      3     1     1     1     3     2     3     3     2     0
5      3     1     1     3     2     3     2     6     5     0
6      1     2     1     3     3     5     5     0     0     5
7      3     1     1     1     1     2     1     2     3     0
8      2     1     1     3     3     0     8     0     2     0
9      1     1     1     1     4     0     3     0     0     0
10     2     1     2     3     2     0     1     0     0     0
11     3     1     1     3     3     0     1     2     0     0
12     3     1     2     3     3     0     2     1     0     5
13     3     2     3     3     1     0     0     5     0     0
14     2     1     1     1     3     0     0     1     0     0
15     3     1     1     2     4     0     0     3     3     1
16     1     1     2     1     4     0     0     0     0     5
17     3     1     3     3     1     0     0     0     0     5
18     3     2     1     3     2     0     0     0     0     5
19     2     1     2     3     3     0     0     0     3     0
20     3     1     3     3     2     0     0     0     1     0

Contingency table level 2:
CST   g3    cr    g1    st     a     b     c     m     k
---   --    --    --    --    ---   ---   ---   ---   ---
1      3     2     3     3     5     3     7     5     0
2      2     1     1     3     5     8     7     2     0
3      3     3     3     3     5     2     3     2     0
4      3     1     1     1     4     4     5     5     0
5      3     1     1     3     3     3     8     5     0
6      1     2     1     3     5     5     0     0     5
7      1     1     1     1     0     3     0     0     0
8      2     1     2     3     0     1     0     3     0
```

```
9      3      1      2      3       0      2      1      0      5
10     2      1      1      1       0      0      1      0      0
11     3      1      1      2       0      0      3      3      1
12     1      1      2      1       0      0      0      0      5
13     3      1      3      3       0      0      0      1      5
14     3      2      1      3       0      0      0      0      5

Contingency  table  level  3:
CST    g3     cr     g1       a      b      c      m      k
---    --     --     --     ---    ---    ---    ---    ---
1      3      2      3        5      3      7      5      0
2      2      1      1        5      8      8      2      0
3      3      3      3        5      2      3      2      0
4      3      1      1        7      7     16     13      1
5      1      2      1        5      5      0      0      5
6      1      1      1        0      3      0      0      0
7      2      1      2        0      1      0      3      0
8      3      1      2        0      2      1      0      5
9      1      1      2        0      0      0      0      5
10     3      1      3        0      0      0      1      5
11     3      2      1        0      0      0      0      5

Contingency  table  level  4:
CST    g3     cr       a      b      c      m      k
---    --     --     ---    ---    ---    ---    ---
1      3      2        5      3      7      5      5
2      2      1        5      9      8      5      0
3      3      3        5      2      3      2      0
4      3      1        7      9     17     14     11
5      1      2        5      5      0      0      5
6      1      1        0      3      0      0      5

Contingency  table  level  5:
CST    g3       a      b      c      m      k
---    --     ---    ---    ---    ---    ---
1      3       17     14     27     21     16
2      2        5      9      8      5      0
3      1        5      8      0      0     10


 #     Species  name
---    ---------------------------------------------
1      Anla  Andropogon  lateralis
2      Arfi  Aristida  filifolia
3      Axaf  Axonopus  affinis
4      Batr  Baccharis  trimera
5      Cebi  Centella  biflora
6      Cose  Coelorachis  selloana
7      Dein  Desmodium  incanum
8      Elgl  Eleocharis  glauco-virens
9      Erho  Eryngium  horridum
10     Erne  Eragrostis  neesii
11     Fare  Facelis  retusa
12     Hepu  Herbertia  pulchella
13     Pano  Paspalum  notatum
14     Papu  Paspalum  pumilum
15     Pimo  Piptochaetium  montevidense
16     Rehi  Relbunium  hirtum
17     Rihu  Richardia  humistrata
18     Rusp  Ruellia  sp.
19     Sege  Setaria  geniculata
```

```
20     Sopt  Soliva  pterosperma
21     Spin  Sporobolus  indicus
22     Sysp  Sysyrinchium  sp.

CST/Species  correspondence,  level  1:
CST#   Species
----   ----------------------------------------------------
1      Anla
2      Pano
3      Pimo  Arfi
4      Rusp
5      Sege  Cose  Erne
6      Axaf
7      Rehi  Fare  Sopt
8      Pano
9      Cebi
10     Erho
11     Hepu
12     Pano  Sysp  Papu
13     Spin
14     Rihu
15     Dein  Batr
16     Cebi
17     Elgl
18     Anla
19     Erho
20     Sege

CST/Species  correspondence,  level  2:
CST#   Species
----   ----------------------------------------------------
1      Anla  Spin
2      Pano
3      Pimo  Arfi
4      Rusp  Rehi  Fare  Sopt
5      Sege  Hepu  Cose  Erne
6      Axaf
7      Cebi
8      Erho
9      Pano  Sysp  Papu
10     Rihu
11     Dein  Batr
12     Cebi
13     Sege  Elgl
14     Anla

CST/Species  correspondence,  level  3:
CST#   Species
----   ----------------------------------------------------
1      Anla  Spin
2      Pano  Rihu
3      Pimo  Arfi
4      Rusp  Sege  Rehi  Fare  Hepu  Cose  Sopt  Erne  Dein  Batr
5      Axaf
6      Cebi
7      Erho
8      Pano  Sysp  Papu
9      Cebi
10     Sege  Elgl
11     Anla
```

```
CST/Species  correspondence,  level  4:
CST#   Species
----   -------------------------------------------
1      Anla   Spin
2      Pano   Erho   Rihu
3      Pimo   Arfi
4      Pano   Rusp   Sege   Rehi   Fare   Hepu   Cose   Sysp   Sopt   Erne   Dein   Papu
       Elgl   Batr
5      Axaf
6      Cebi

CST/Species  correspondence,  level  5:
CST#   Species
----   -------------------------------------------
1      Anla   Pano   Pimo   Rusp   Arfi   Sege   Rehi   Fare   Hepu   Cose   Sysp   Spin
       Sopt   Erne   Dein   Papu   Elgl   Batr
2      Pano   Erho   Rihu
3      Axaf   Cebi

Level  1:  Indeterminacy:
     0          0.5          0.5       0.5455       0.9167
                  0       0.5333       0.5714       0.8667
                               0       0.5714       0.8667
                                            0       0.9286
                                                         0

Level  2:  Indeterminacy:
     0       0.3333       0.4444       0.4444       0.9091
                  0       0.4545       0.4545       0.8462
                               0          0.4       0.8333
                                            0       0.8333
                                                         0

Level  3:  Indeterminacy:
     0        0.375       0.3333       0.4286       0.7778
                  0        0.375       0.4444       0.7273
                               0       0.4286       0.7778
                                            0          0.8
                                                         0

Level  4:  Indeterminacy:
     0       0.1667          0.2          0.2          0.5
                  0       0.3333       0.3333       0.3333
                               0            0       0.6667
                                            0       0.6667
                                                         0

Level  5:  Indeterminacy:
     0            0       0.3333       0.3333       0.3333
                  0       0.3333       0.3333       0.3333
                               0            0       0.6667
                                            0       0.6667
                                                         0
```

6.6.2. Resemblance

SYNCSA offers resemblance functions to compare relevés and CSTs. These functions are described in Chapter 2 in the context of relevé comparisons, but they are also defined for CST comparisons at a specific level of the character hierarchy. The example in Run 6 shows the resemblance function options in lines 18-27. Information divergence within CSTs (opt. 1) and mutual (opt. 2) are based respectively on Eqs. 2.21 and 2.22; these are available only when data is mixed. Cross product calculations may use data without centering (opt. 3, Eq. 2.10) or execute implicit centering adjustment (opt. 4, Eq. 2.12). The correlation coefficient (opt. 5, Eq. 2.13) is based on centered cross products. The computation of squared euclidean distances (opt. 6) follows Eq. 2.15; the relative function (opt. 7) uses the same equation, but includes adjustment of the relevé vectors to unit sum (Eq. 2.7). The absolute value function (opt. 8) and its relative form (opt. 9) are based respectively on Eqs. 2.19 and 2.20. The computation of squared chord distances (opt. 10) follows Eq. 2.18.

Next, the user can select whether nominal and partial resemblance matrices are to be shown in the printout (Run 6, lines 31-32). The printout file is listed in Output 2.

If resemblances of CSTs are selected instead, the user will enter the resemblance function option and the level of the character hierarchy at which the CSTs are defined. The resemblance matrix will be appended to file *Session*.prinda. Resemblance of CSTs is not applicable when the CSTs are fuzzy under the pairwise adjustment. Also, partial resemblances are undefined.

Run 6:

```
1   ANALYSIS OF COMMUNITY STRUCTURES                        session: ufrgs5re5ch
2   ----------------------------------------------------------------------------
3   Options:
4      H  set data
5      K  ranking characters
6      R  resemblance
7      G  specify relevé group partition
8      M  specify CST group partition
9      C  return to session menu
10  Type  option:  R
11
12
13  RESEMBLANCE SETTINGS                                    session: ufrgs5re5ch
14  ----------------------------------------------------------------------------
```

```
Resemblance of relevés(R) or CSTs(C)? r

Resemblance function option:
      1    information divergence (within CSTs)
      2    information divergence (mutual)
      3    cross product of uncentered data
      4    cross product of centered data
      5    correlation coefficient
      6    squared euclidean distance
      7    relative squared euclidean distance
      8    absolute value function
      9    relative absolute value function
      10   squared chord distance
Type option no.: 10

Printout:
   - nominal resemblances y/n  y
   - partial resemblances y/n  y
Results appended to file hd:csa:ufrgs5re5ch.prinda
```

Output 2:

... (see Output 1 for data settings and vegetation tables)

```
-------------------------------------------------------------
RESEMBLANCE OF RELEVÉS
-------------------------------------------------------------
Tue Jul 21 11:53:17 1992
Resemblance function: 10 (squared chord distance)
Fuzzy transformation: none (crisp)

Level 1: Nominal resemblances (Option 10):
    0      0.9661      0.6759      0.8306       1.588
                0       1.445       1.005       1.455
                            0      0.8505       1.884
                                        0       1.944
                                                    0

Level 2: Nominal resemblances (Option 10):
    0      0.2675      0.3961       0.583       1.602
                0      0.4898      0.7826       1.475
                            0      0.2244       1.901
                                        0       1.859
                                                    0

Level 3: Nominal resemblances (Option 10):
    0       0.227      0.2999      0.4696       1.533
                0      0.3605      0.6205       1.417
                            0      0.1302       1.808
                                        0        1.78
                                                    0

Level 4: Nominal resemblances (Option 10):
      0      0.2093      0.3109      0.3627      0.5137
                  0      0.2804      0.3376      0.4782
                              0    0.009649      0.4356
                                          0      0.3827
                                                      0
```

```
Level  5:  Nominal  resemblances  (Option  10):
       0      0.09999      0.07516      0.07779       0.1462
                    0       0.2693       0.2992        0.255
                                 0     0.002949       0.3739
                                              0       0.3501
                                                           0

Level  1:  Partial  resemblances  (Option  10):
     0        1.222       0.7914       0.9243        1.583
                  0        1.862         1.09        1.448
                               0        1.131        1.876
                                            0        1.974
                                                         0

Level  2:  Partial  resemblances  (Option  10):
     0       0.2938       0.4179       0.5729         1.64
                  0       0.5433       0.8442        1.506
                               0       0.3822        1.976
                                            0        1.922
                                                         0

Level  3:  Partial  resemblances  (Option  10):
     0       0.2257        0.293       0.5258        2.266
                  0       0.4067       0.8142        2.245
                               0       0.2016        2.875
                                            0        2.943
                                                         0

Level  4:  Partial  resemblances  (Option  10):
      0       0.2816       0.8256       0.9494        1.255
                   0       0.2748       0.3789       0.7732
                                0      0.01978       0.5352
                                             0       0.4339
                                                          0

Level  5:  Partial  resemblances  (Option  10):
       0      0.09999      0.07516      0.07779       0.1462
                    0       0.2693       0.2992        0.255
                                 0     0.002949       0.3739
                                              0       0.3501
                                                           0
Global  resemblances  (Option  10):
     0       0.9661       0.6759       0.8306        1.588
                  0        1.445        1.005        1.455
                               0       0.8505        1.884
                                            0        1.944
                                                         0
```

6.6.3. Specify group partitions

Relevé and CST group partitions are entered manually. The infomation is stored in file *Session*.track and is lost when a new partition of the same kind is given. The example (Run 7, lines 1-27) specifies a relevé group partition (opt. G). The user informs the number of groups (line 12), labels (lines 14-22), and whether objects are ordered by groups or not (line 23). If so, the number of objects in each group will be asked. Otherwise, as in the example, the group (its label) to which each relevé or CST belongs is informed (line 27). Labels for groups may be entered by the user and must not exceed 2 alphanumeric symbols, or assigned by the program as ordered integers (see Run 11, line 23). If a CST group partition (opt. M) is being specified, the user will also have to inform the level of the character hierarchy at which the CSTs are defined. Note that information generated by the Set Data procedure is needed for option M be available.

6.6.4. Character ranking

On choosing option K (Run 7, line 37), which is available when the character set contains more than one character, the menu offers the character ranking criteria that are applicable (Run 7, lines 43-48). Chapter 4 describes several structure evaluation functions. Ranking for maximum congruence of community resemblance / environmental resemblance (opt. 1, Eq. 4.1) is only feasible if file *Session*.formda contains environmental variables. Environmental resemblances in this case are relevé squared euclidean distances after data adjustments involving centering and normalization within variables. Ranking for maximum convergence or divergence of relevé groups (opt. 2 or 3, Eqs. 4.2 or 4.3) is only offered when a group partition is already specified in the current session. Ranking based on character redundancy in community data structure (opt. 5, Eqs. 4.5 or 4.6) always applies. The combined criteria involve a weighting of the redundancy measure (Eq. 4.7) by the environmental congruence (opt. 6, Eq. 4.8) or by the structural divergence of relevé groups (opt. 7, Eq. 4.9). The examples given below follow option 3 (Run 7) and option 1 (Run 8) in session "ufrgs15re12ch". When group convergence or divergence is selected as ranking criterion the user is asked to specify the relevé groups and the group dissimilarity method for computations. Regarding the latter, average dissimilarity (Eq. 4.2) or nearest neighbor dissimilarity (Eq. 4.3) may be selected. The example in Run 7, uses groups 1 and 2 (line 58) of the current relevé partition (lines 1-27) to compute an average dissimilarity (line 63).

Next the user is asked to specify the data type, fuzzy adjustment options and resemblance function as already explained (Sections 6.6.1 and 6.6.2), except that cross products cannot be selected. If community resemblances are defined in correlation coefficient terms, they will be transformed into distances (Eq. 5.2). The examples (Runs 7 and 8) use mixed data type, crisp CSTs and squared chord distance. While SYNCSA performs computations, the current ranking step is informed on the screen. Computations may be slow for large character sets. The output is stored on file *Session*.ChRank, the first line being the optimal character order, which is followed by an explanation of the ranking process (see Outputs 3 and 4 with results of Runs 7 and 8). In the Data Settings Menu a file *Session*.ChRank may be used to specify the character order.

Run 7:

```
ANALYSIS OF COMMUNITY STRUCTURES                            ufrgs15re12ch
--------------------------------------------------------------------------
Options:
   H  set data
   K  ranking characters
   G  specify relevé group partition
   C  return to session menu
Type option: g

SET GROUP PARTITION                                         ufrgs15re12ch
--------------------------------------------------------------------------
Specify the number of groups of relevés: 3

Labels for groups:
   1  given by default
   2  specify new labels
Type option no.: 2

Enter labels (max. 2 alphanumeric characters):
 1  2  3
AA BB CC
Are these correct? y/n y
Are the relevés ordered by groups? y/n n

Enter group membership (label) for relevés:
relevés:   a   b   c   d   e   f   g   h   i   j   k   l   m   n   o
Group:    AA  BB  AA  BB  AA  AA  AA  AA  BB  BB  CC  AA  BB  AA  CC


ANALYSIS OF COMMUNITY STRUCTURES                            ufrgs15re12ch
--------------------------------------------------------------------------
Options:
   H  set data
   K  ranking characters
   G  specify relevé group partition
   C  return to session menu
Type option: k


CHARACTER RANKING SETTINGS                                  ufrgs15re12ch
--------------------------------------------------------------------------
Character ranking criterion maximized:
    1  congruence of community resemblance/environmental resemblance
   2  convergence of community groups
   3  divergence of community groups
    5  character redundancy in community data structure
   6  criteria 1 and 5 combined
```

```
    7  criteria 2 and 5 combined
Type option no.: 3

Groups to be considered in computations:
   1  all
   2  some
Type option no.: 2

Select groups (0 for no, 1 for yes):
AA BB CC
1  1  0

Group dissimilarity computation method:
    1  average dissimilarity
    2  nearest neighbor dissimilarity
Type option no.: 1


DATA SETTINGS                                               ufrgs15re12ch
-------------------------------------------------------------------------
Data type:
   1  mixed
   2  quantitative
   3  qualitative
Type option no.: 1

Fuzzy set adjustment:
    0  no adjustment (crisp CSTs)
    1  fuzzy CSTs global adjustment
    2  fuzzy CSTs pairwise adjustment
Type option no.: 0

Resemblance function option:
      1    information divergence (within CSTs)
      2    information divergence (mutual)
      3    cross product of uncentered data
      4    cross product of centered data
      5    correlation coefficient
      6    squared euclidean distance
      7    relative squared euclidean distance
      8    absolute value function
      9    relative absolute value function
      10   squared chord distance
Type option no.: 10
Results on file hd:csa:ufrgs15re12ch.ChRank
```

Output 3:

```
he wi tx g2 ev ed st g3 g1 ve lf cr
The characters listed above are ordered by decreasing rank.

CHARACTER-BASED COMMUNITY ANALYSIS                          SYNCSA v.1.0
-------------------------------------------------------------------------
CHARACTER  RANKING
-------------------------------------------------------------------------
Mon Jun 29 23:38:47 1992
Session:          ufrgs15re12ch
Formatted data:  ufrgs15re12ch.formda
```

```
Character ranking criterion: 3 (divergence community groups)
Group dissimilarity computation method: 1 (average dissimilarity)
Resemblance option: 10 (squared chord distance)
Fuzzy transformation: none (crisp)
Data component: mixed
Relevés:
   a   b   c   d   e   f   g   h   i   j   k   l   m   n   o
Group membership:
  AA  BB  AA  BB  AA  AA  AA  AA  BB  BB  CC  AA  BB  AA  CC
Relevés used in computations:
   1   1   1   1   1   1   1   1   1   1   0   1   1   1   0

SEF          Character subset used:
0.08821      lf
0.2165       g1
0.4137       g2
0.1135       g3
0.02986      st
0.07993      cr
0.4345       tx
0.1307       ed
0.3048       ev
0.4117       ve
0.625        wi
1.157        he
1.208        he   lf
```

... (not shown)

```
1.908        he   wi   tx   g2   ev   ed   st   g3   g1   ve   lf
1.908        he   wi   tx   g2   ev   ed   st   g3   g1   ve   cr
1.908        he   wi   tx   g2   ev   ed   st   g3   g1   ve   lf   cr
Elapsed time: 55.3333 seconds
```

Run 8:

```
Type option: K

CHARACTER RANKING SETTINGS                           session: ufrgs15re12ch
---------------------------------------------------------------------------
Character ranking criterion maximized:
   1  congruence of community resemblance/environmental resemblance
   2  convergence of community groups
   3  divergence of community groups
   5  character redundancy in community data structure
   6  criteria 1 and 5 combined
   7  criteria 2 and 5 combined
Type option no.: 1


DATA SETTINGS                                        session: ufrgs15re12ch
---------------------------------------------------------------------------
```

... (the same as in Run 7)

```
Results on file hd:csa:ufrgs15re12ch.ChRank
```

Output 4:

```
g3 cr g1 st wi lf tx ve ed ev he g2
The characters listed above are ordered by decreasing rank.

CHARACTER-BASED COMMUNITY ANALYSIS                        SYNCSA v.1.0
----------------------------------------------------------------------
CHARACTER  RANKING
----------------------------------------------------------------------
Tue Jun 30 09:06:15 1992
Session:          ufrgs15re12ch
Formatted  data:  ufrgs15re12ch.formda
Character ranking criterion: 1 (congruence community resemblance
/environmental  resemblance)
Resemblance option: 10 (squared chord distance)
Fuzzy transformation: none (crisp)
Data component: mixed

SEF          Character  subset  used:
-0.1705      lf
0.2323       g1
-0.09846     g2
0.4494       g3
-0.1241      st
0.08424      cr
0.05899      tx
-0.03895     ed
0.08995      ev
-0.09045     ve
0.1006       wi
-0.05266     he
0.2884       g3   lf
0.3789       g3   g1
0.02147      g3   g2
0.3117       g3   st
0.4653       g3   cr
0.3666       g3   tx
0.162        g3   ed
0.2381       g3   ev
0.008302     g3   ve
0.2947       g3   wi
0.03126      g3   he
0.2462       g3   cr   lf
0.4776       g3   cr   g1
0.07424      g3   cr   g2
0.4397       g3   cr   st
0.3161       g3   cr   tx
0.2095       g3   cr   ed
0.2384       g3   cr   ev
0.07147      g3   cr   ve
0.2742       g3   cr   wi
0.09105      g3   cr   he
0.431        g3   cr   g1   lf
0.3077       g3   cr   g1   g2
0.4476       g3   cr   g1   st
0.395        g3   cr   g1   tx
0.4209       g3   cr   g1   ed
0.3867       g3   cr   g1   ev
0.3667       g3   cr   g1   ve
0.4457       g3   cr   g1   wi
0.2808       g3   cr   g1   he
```

```
0.3991      g3  cr  g1  st  lf
0.2607      g3  cr  g1  st  g2
0.3722      g3  cr  g1  st  tx
0.3987      g3  cr  g1  st  ed
0.3592      g3  cr  g1  st  ev
0.3419      g3  cr  g1  st  ve
0.4265      g3  cr  g1  st  wi
0.2639      g3  cr  g1  st  he
0.4101      g3  cr  g1  st  wi  lf
0.278       g3  cr  g1  st  wi  g2
0.3849      g3  cr  g1  st  wi  tx
0.3752      g3  cr  g1  st  wi  ed
0.4047      g3  cr  g1  st  wi  ev
0.3884      g3  cr  g1  st  wi  ve
0.2687      g3  cr  g1  st  wi  he
0.249       g3  cr  g1  st  wi  lf  g2
0.3724      g3  cr  g1  st  wi  lf  tx
0.3683      g3  cr  g1  st  wi  lf  ed
0.3721      g3  cr  g1  st  wi  lf  ev
0.3646      g3  cr  g1  st  wi  lf  ve
0.2656      g3  cr  g1  st  wi  lf  he
0.2157      g3  cr  g1  st  wi  lf  tx  g2
0.3182      g3  cr  g1  st  wi  lf  tx  ed
0.3092      g3  cr  g1  st  wi  lf  tx  ev
0.3516      g3  cr  g1  st  wi  lf  tx  ve
0.2845      g3  cr  g1  st  wi  lf  tx  he
0.2533      g3  cr  g1  st  wi  lf  tx  ve  g2
0.2979      g3  cr  g1  st  wi  lf  tx  ve  ed
0.2852      g3  cr  g1  st  wi  lf  tx  ve  ev
0.2765      g3  cr  g1  st  wi  lf  tx  ve  he
0.2141      g3  cr  g1  st  wi  lf  tx  ve  ed  g2
0.2724      g3  cr  g1  st  wi  lf  tx  ve  ed  ev
0.2294      g3  cr  g1  st  wi  lf  tx  ve  ed  he
0.1958      g3  cr  g1  st  wi  lf  tx  ve  ed  ev  g2
0.1983      g3  cr  g1  st  wi  lf  tx  ve  ed  ev  he
0.1143      g3  cr  g1  st  wi  lf  tx  ve  ed  ev  he  g2
Elapsed time: 43.6333 seconds
```

6.6.5. Evaluation of structures

Evaluation of structures is mainly based on functions described for character ranking. The character order and other set of conditions that define data structures are the ones given by procedures prior to Evaluation (see Figure 6.6.1). The example (Run 9) depicts a complete analysis from data settings to evaluation of community structures. It applies the character order given in Output 4, computes relevé resemblances and at the end (lines 76-106) evaluates the community data structures so defined (nominal resemblances, line 105) with regard to the congruence with environmental resemblance (line 99). If probabilities of nominal resemblances are available, they will also be offered to define community data structure (opt. 3, lines 102-105).

If ordination scores and environmental variables are available, options 2 and 5 will also be offered to be selected in line 99. Under option 2 (congruence of community ordination scores / environmental resemblance, Section 3.4) the user will be asked to specify the number of ordination components that are to define community structure at each level of the character hierarchy. Under option 5 (correlation of community ordination axes / environmental variables) the printout will present for each hierarchical level and major ordination component a table with the correlation coefficients. This table is similar to the one shown in Output 8, but here the correlation involves environmental variables instead of CSTs. By default 6 axes are shown, but the user can change it in Preferences Menu.

The results are sent to file *Session*.prinda (Output 5). Following options 1 or 2 in line 99 and if the appropriate item in Preferences Menu is set on, a file named *Session*.Struct will be created holding the distance matrices involved in computations.

Run 9:

```
 1  ANALYSIS OF COMMUNITY STRUCTURES                             ufrgs15re12ch
 2  ------------------------------------------------------------------------
 3  Options:
 4     H  set data
 5     K  ranking characters
 6     G  specify relevé group partition
 7     C  return to session menu
 8  Type option: h
 9
10  DATA SETTINGS                                                ufrgs15re12ch
11  ------------------------------------------------------------------------
12  Data type:
13     1  mixed
14     2  quantitative
15     3  qualitative
16  Type option no.: 1
17
18  Fuzzy set adjustment:
19     0  no adjustment (crisp CSTs)
20     1  fuzzy CSTs global adjustment
21     2  fuzzy CSTs pairwise adjustment
22  Type option no.: 0
23
24  Character order:
25     1  as in data
26     2  user specified
27     3  from file ChRank
28  Type option no.: 3
29
30  ChRank file to use:
31     1  from this session
```

```
     2  from another session
Type option no.: 1

Printout:
        - vegetation tables        y/n  n
        - species names            y/n  n
        - indeterminacy index      y/n  n
Results on file hd:csa:ufrgs15re12ch.prinda


ANALYSIS OF COMMUNITY STRUCTURES                             ufrgs15re12ch
--------------------------------------------------------------------------
Options:
    H  set data
    K  ranking characters
    R  resemblance
    G  specify relevé group partition
    M  specify CST group partition
    C  return to session menu
Type option: r

RESEMBLANCE SETTINGS                                         ufrgs15re12ch
--------------------------------------------------------------------------
Resemblance of relevés(R) or CSTs(C)? r

Resemblance function option:
        1    information divergence (within CSTs)
        2    information divergence (mutual)
        3    cross product of uncentered data
        4    cross product of centered data
        5    correlation coefficient
        6    squared euclidean distance
        7    relative squared euclidean distance
        8    absolute value function
        9    relative absolute value function
        10   squared chord distance
Type option no.: 10

Printout:
    - nominal resemblances y/n  n
    - partial resemblances y/n  n
Results appended to file hd:csa:ufrgs15re12ch.prinda


ANALYSIS OF COMMUNITY STRUCTURES                             ufrgs15re12ch
--------------------------------------------------------------------------
Options:
    H  set data
    K  ranking characters
    R  resemblance
    U  cluster
    D  randomization
    O  ordination
    P  draw profile
    E  evaluation of data structures
    G  specify relevé group partition
    M  specify CST group partition
    T  create structured table
    C  return to session menu
Type option: e

```

```
EVALUATION SETTINGS                                         ufrgs15re12ch
-------------------------------------------------------------------------
Structure evaluation function option:
    1  congruence of community resemblance/environmental resemblance
    3  divergence of community groups
    4  character redundancy in community data structure
Type option no.: 1
Resemblance function available: option 10

Use:
    1  nominal resemblances
    2  partial resemblances
Type option no.: 1
Results appended to file hd:csa:ufrgs15re12ch.prinda
```

Output 5:

```
CHARACTER-BASED COMMUNITY ANALYSIS                          SYNCSA v.1.0
-------------------------------------------------------------------------
EXPLORATION OF VEGETATIONAL DATA STRUCTURES
-------------------------------------------------------------------------
Tue Jun 30 10:05:37 1992
Session:             ufrgs15re12ch
Formatted data:      ufrgs15re12ch.formda
Unformatted data:   ufrgs15re12ch.dat

Character set:
Order:               g3 cr g1 st wi lf tx ve ed ev he g2
Number of states: 3  3  3  3  4  4  4  3  3  3  5  3
Character type:   2  2  2  2  3  2  3  2  2  2  3  2

Number of relevés: 15
Relevé labels: a   b   c   d   e   f   g   h   i   j   k   l   m   n   o
Fuzzy transformation: none (crisp)
Data type: mixed

-------------------------------------------------------------------------
RESEMBLANCE OF RELEVÉS
-------------------------------------------------------------------------
Tue Jun 30 10:05:45 1992
Resemblance function: 10 (squared chord distance)
Fuzzy transformation: none (crisp)

-------------------------------------------------------------------------
EVALUATION OF STRUCTURES
-------------------------------------------------------------------------
Tue Jun 30 10:05:53 1992
Based on nominal resemblances.
Structure evaluation function: 1 (congruence community/environmental resem-
blance)
Level 1: 0.1143
Level 2: 0.1983
Level 3: 0.2724
Level 4: 0.2979
Level 5: 0.3516
Level 6: 0.3724
Level 7: 0.4101
Level 8: 0.4265
```

```
Level 9: 0.4476
Level 10: 0.4776
Level 11: 0.4653
Level 12: 0.4494
```

6.6.6. Randomization

SYNCSA implements the randomization procedures described in Sections 2.8 and 3.4 to generate probabilities of relevé resemblances and structure evaluation functions. The example in Run 10 performs randomization to obtain probabilities of the resemblances computed in Run 6. On choosing option D, the user is asked to select the null hypothesis to define the reference set (Run 10, line 22). Under random composition (opt. 1) or random taxon (opt. 2) null hypotheses (see Section 2.8.2), the program will compute probabilities for relevé resemblances and, if available, structure evaluation functions. Next, the user specifies the number of iterations (line 23) and whether the pseudorandom number generator is to take the clock time or a specified number as seed (line 28). Note that an identical seed number will yield identical probabilities in repeated runs with the same data. The output is appended to file *Session*.prinda (Output 6). Note that in the output the relevé resemblance probabilities are printed below the diagonal of the resemblance matrix. For instance, the probability corresponding to the resemblance of relevés 1 and 2 at level 1 (0.9661) is found in line 2 column 1 (0.03). The probabilities given are $\alpha = P(r_{rnd} \geq r)$ when r is a similarity, and $1-\alpha = P(r_{rnd} < r)$ when r is a dissimilarity.

In addition to the null hypotheses described above, SYNCSA implements the randomization method under the random partition null hypothesis (Run 11, line 102). This option needs information produced by evaluation of community groups divergence (SEF opt. 3). The null hypothesis defines a reference set in which only the relevé group membership is randomized. That is, the complete reference set is formed by the $\nu!$ permutations of the group membership list containing ν relevés. The relevé resemblance matrices remain constant for the data sets members of the reference set. Therefore, only probabilities for the group divergence measure are meaningful. The example in Run 11 shows how this option is made available in session "ufrgs5ch5re". First, a group partition is specified (lines 1-28). Second, option E is chosen to compute the average resemblance between groups 1 and 2 (lines 31-76). Third, randomization under the random partition null hypothesis is selected (lines 79-109). The output is appended to file *Session*.prinda (Output 7). The probabilities given are $1-\alpha =$

$P(\sigma_{rnd} < \sigma)$ where σ may be an average dissimilarity (Eq. 4.2) or average nearest neighbor dissimilarity (Eq. 4.3) of the groups selected. For instance, as shown in Output 7, at hierarchical level 1, the probability of finding an average dissimilarity of relevé groups 1 and 2 smaller than 1.139, were the null hypothesis of random group partition true, is 0.888. That is, the group dissimilarity is not so uncommonly high as to indicate a non-random partition, since the probability of a Type I error is 0.112.

The user can choose in the Preferences Menu to have the reference set, resemblance matrices and SEF values stored in file *Session*.RndData. Keep in mind that this option may produce a huge file if it is not balanced by a small number of iterations and data set size.

Run 10:

```
ANALYSIS OF COMMUNITY STRUCTURES                                   ufrgs5re5ch
----------------------------------------------------------------------------
Options:
   H  set data
   K  ranking characters
   R  resemblance
   U  cluster
   D  randomization
   O  ordination
   P  draw profile
   E  evaluation of data structures
   G  specify relevé group partition
   M  specify CST group partition
   C  return to session menu
Type option: d

RANDOMIZATION SETTINGS                                             ufrgs5re5ch
----------------------------------------------------------------------------
Reference set based on:
   1  random composition hypothesis
   2  random taxon hypothesis
Type option no.: 1
Enter number of iterations: 100

Initialization of the random number generator:
   1  taken from the clock
   2  specified by the user
Type option no.: 1
Results appended to file hd:csa:ufrgs5re5ch.prinda
```

Output 6:

... (see Output 2 for results of relevé resemblances)

```
-------------------------------------------------------------------
RANDOMIZATION
-------------------------------------------------------------------
Tue Jul 21 11:53:37 1992
Elapsed time: 18.6 seconds
Reference set option: 1 (random composition hypothesis)
Number of iterations: 100
Random generator seed: 5633

Probabilities (1 - alpha = P(Xrnd < X) are given below the diagonal.
Nominal resemblances (option 10) are given in the upper half matrix.

Level 1: Nominal resemblances (upper half), probabilities (below diagonal):
    0      0.9661      0.6759      0.8306       1.588
 0.03           0       1.445       1.005       1.455
    0        0.43           0      0.8505       1.884
 0.07        0.15        0.06           0       1.944
 0.55        0.35        0.81        0.86           0

Level 2: Nominal resemblances (upper half), probabilities (below diagonal):
    0      0.2675      0.3961       0.583       1.602
    0           0      0.4898      0.7826       1.475
 0.01        0.01           0      0.2244       1.901
 0.07        0.19           0           0       1.859
 0.82        0.65        0.96        0.92           0

Level 3: Nominal resemblances (upper half), probabilities (below diagonal):
    0       0.227      0.2999      0.4696       1.533
 0.05           0      0.3605      0.6205       1.417
 0.11        0.13           0      0.1302       1.808
 0.29        0.48        0.03           0        1.78
 0.95        0.92        0.99        0.98           0

Level 4: Nominal resemblances (upper half), probabilities (below diagonal):
      0      0.2093      0.3109      0.3627      0.5137
   0.27           0      0.2804      0.3376      0.4782
   0.37        0.33           0    0.009649      0.4356
   0.53        0.52           0           0      0.3827
   0.59        0.47        0.41        0.42           0

Level 5: Nominal resemblances (upper half), probabilities (below diagonal):
      0     0.09999     0.07516     0.07779      0.1462
   0.53           0      0.2693      0.2992       0.255
   0.35        0.72           0    0.002949      0.3739
   0.36        0.86        0.03           0      0.3501
    0.6        0.79        0.88         0.8           0
```

Run 11:

```
ANALYSIS OF COMMUNITY STRUCTURES                         ufrgs5re5ch
-------------------------------------------------------------------
Options:
   H  set data
    K  ranking characters
    R  resemblance
    U  cluster
    D  randomization
    O  ordination
```

```
    P  draw profile
    E  evaluation of data structures
    G  specify relevé group partition
    M  specify CST group partition
    C  return to session menu
Type option: g

SET GROUP PARTITION                                         ufrgs5re5ch
-----------------------------------------------------------------------
Specify the number of groups of relevés: 3
Labels for groups:
    1  given by default
    2  specify new labels
Type option no.: 1
Are the relevés ordered by groups? y/n n

Enter group membership (label) for each relevés:
relevés:   a   b   c   m   k
Group:     1   2   1   1   3


ANALYSIS OF COMMUNITY STRUCTURES                            ufrgs5re5ch
-----------------------------------------------------------------------
Options:
    H  set data
    K  ranking characters
    R  resemblance
    U  cluster
    D  randomization
    O  ordination
    P  draw profile
    E  evaluation of data structures
    G  specify relevé group partition
    M  specify CST group partition
    T  create structured table
    C  return to session menu
Type option: e

EVALUATION SETTINGS                                         ufrgs5re5ch
-----------------------------------------------------------------------
Structure evaluation function option:
     1  congruence of community resemblance/environmental resemblance
     3  divergence of community groups
     4  character redundancy in community data structure
Type option no.: 3

Groups to be considered in computations:
    1  all
    2  some
Type option no.: 2

Select groups (0 for no, 1 for yes):
1  2  3
1  1  0

Group dissimilarity computation method:
     1  average dissimilarity
     2  nearest neighbor
Type option no.: 1
Resemblance function available: option 10 and its probabilities
```

```
Partition available. Use:
    1  nominal resemblances
    2  partial resemblances
    3  probabilities of nominal resemblances
Type option no.: 1
Results appended to file hd:csa:ufrgs5re5ch.prinda


ANALYSIS OF COMMUNITY STRUCTURES                            ufrgs5re5ch
-----------------------------------------------------------------------
Options:
    H  set data
    K  ranking characters
    R  resemblance
    U  cluster
    D  randomization
    O  ordination
    P  draw profile
    E  evaluation of data structures
    G  specify relevé group partition
    M  specify CST group partition
    T  create structured table
    C  return to session menu
Type option: d

RANDOMIZATION SETTINGS                                      ufrgs5re5ch
-----------------------------------------------------------------------
Reference set based on:
    1  random composition hypothesis
    2  random taxon hypothesis
    3  random partition hypothesis
Type option no.: 3
Enter number of iterations: 100

Initialization of the random number generator:
    1  taken from the clock
    2  specified by the user
Type option no.: 1
Results appended to file hd:csa:ufrgs5re5ch.prinda
```

Output 7:

```
-----------------------------------------------------------------------
EVALUATION OF STRUCTURES
-----------------------------------------------------------------------
Tue Jul 21 11:53:38 1992
Based on nominal resemblances.
Structure evaluation function: 3 (divergence community groups)
     Relevés:          a   b   c   m   k
     Group membership: 1   2   1   1   3
Group dissimilarity computation method: 1 (average dissimilarity)
Level 1: 1.139
Level 2: 0.5133
Level 3: 0.4027
Level 4: 0.2758
```

```
Level 5: 0.2228

------------------------------------------------------------
RANDOMIZATION
------------------------------------------------------------
Tue Jul 21 11:53:42 1992
Elapsed time: 18.8833 seconds
Reference set option: 3 (random partition hypothesis)
Number of iterations: 1000
Random generator seed: 51581

Probabilities for structure evaluation:
Structure evaluation function: 3 (divergence community groups)
Based on nominal resemblances.
Level 1: 0.888
Level 2: 0.796
Level 3: 0.796
Level 4: 0.687
Level 5: 0.888
```

6.6.7. Draw profiles

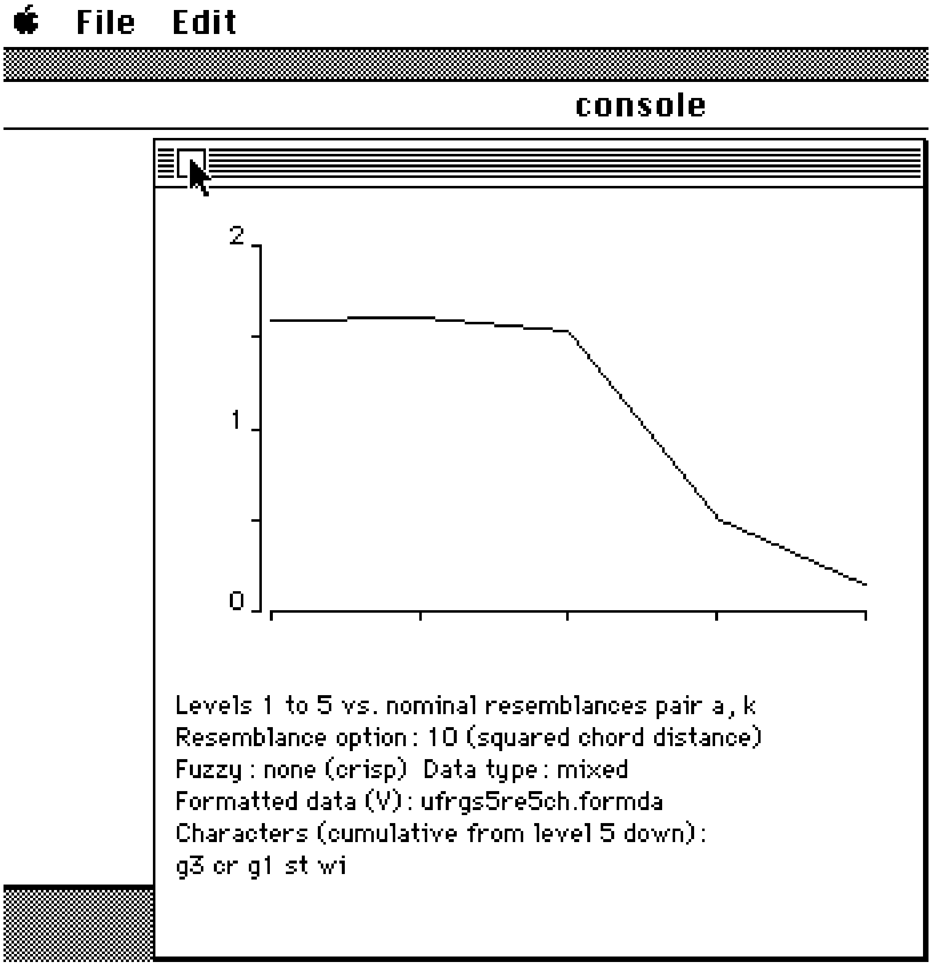
The profiles are drawn on the screen and depict the variation of functions, such as resemblances of a relevé pair, structural evaluations and probabilities of these quantities, along the character hierarchy. The hierarchical levels are on the x axis and the function of interest on the y axis. The first example (Run 12) plots the nominal resemblances of the relevé pair a, k (obtained in Run 6). The vertical scale can be redefined (lines 28-33), which is useful to obtain comparable profiles. The screen's contents can be stored as a picture file as already explained in 6.1. The user must click in the close box by the arrow (see Screen 1) to proceed running the program.

The second example (Run 13, Screen 2) generates a profile of the structure evaluation function (congruence of community resemblance / environmental resemblance obtained in Run 9). The user can specify transformations to be performed in the function (line 13) if it is a correlation, such as plot its squared value, or the derived stress (the one-complement of the squared function). The size and appearance of profiles can be changed in the Preferences Menu.

Run 12:

```
ANALYSIS OF COMMUNITY STRUCTURES                                    ufrgs5re5ch
-------------------------------------------------------------------------------
Options:
   H  set data
   K  ranking characters
   R  resemblance
   U  cluster
   D  randomization
   O  ordination
   P  draw profile
   E  evaluation of data structures
   G  specify relevé group partition
   M  specify CST group partition
   T  create structured table
   C  return to session menu
Type option: p


PROFILE SETTINGS                                                    ufrgs5re5ch
-------------------------------------------------------------------------------
Draw profile using:
   3  nominal resemblances
   4  partial resemblances
Type option no.: 3

Enter labels of the pair of relevés to plot: a    k

Profile's vertical scale:
   1  based on actual data
   2  user specified
Type option no.: 2
Specify minimum value: 0
Specify maximum value: 2
```

Screen 1:

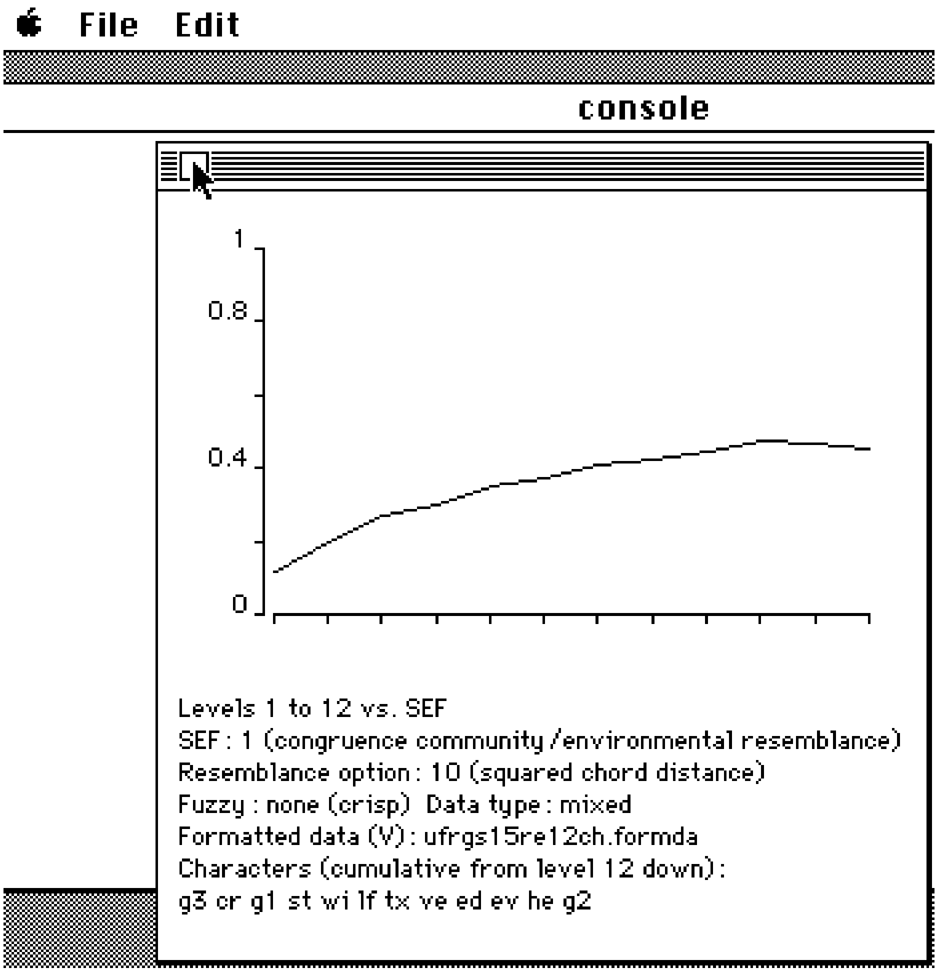


```
Levels 1 to 5 vs. nominal resemblances pair a, k
Resemblance option : 10 (squared chord distance)
Fuzzy : none (crisp)  Data type : mixed
Formatted data (V) : ufrgs5re5ch.formda
Characters (cumulative from level 5 down) :
g3 cr g1 st wi
```

Run 13:

```
PROFILE SETTINGS                                         ufrgs15re12ch
----------------------------------------------------------------------
Draw profile using:
    1  structural evaluation function (SEF)
    3  nominal resemblances
    4  partial resemblances
Type option no.: 1

Scalar transformation:
    0  no transformation
    1  plot the squared correlation
    2  plot the stress function
Type option no.: 0

Profile's vertical scale:
    1  based on actual data
    2  user specified
Type option no.: 1
```

Screen 2:



```
Levels 1 to 12 vs. SEF
SEF : 1 (congruence community/environmental resemblance)
Resemblance option : 10 (squared chord distance)
Fuzzy : none (crisp)  Data type : mixed
Formatted data (V) : ufrgs15re12ch.formda
Characters (cumulative from level 12 down) :
g3 cr g1 st wi lf tx ve ed ev he g2
```

6.6.8. Ordination

The method implemented by SYNCSA is eigenordination (Section 5.2). On selecting ordination (letter O), the user is asked to inform which available resemblance matrices are to be subjected to ordination. The example (Run 14, line 27) uses nominal resemblances. Note that if probabilities of nominal resemblances are selected, they will be interpreted as a dissimilarity with values in the 0-1 range. For each level, the printout presents all the non-zero eigenvalues and their percentages with respect to the total (see Output 8). The printout may also include relevé scores (line 30), which are previously adjusted by the Procrustes method (Section 5.3), and correlation coefficients between CSTs and ordination components (line 31). The correlation coefficients are akin to the eigenvector elements in a Q-PCA, *i.e.*, they indicate which CSTs are better represented by the ordination component. For instance, in Output 8, the first ordination component, hierarchical level 1, reflects a decrease in CSTs #5 and #7, and an increase in CSTs #12, #16, #17 and #18. By default the printout only shows scores and correlation coefficients corresponding to the 6 largest eigenvalues, but the user can change this in the Preferences Menu.

Run 14:

```
ANALYSIS OF COMMUNITY STRUCTURES                             ufrgs5re5ch
------------------------------------------------------------------------
Options:
   H  set data
    K   ranking characters
    R   resemblance
    U   cluster
    D   randomization
    O   ordination
    P   draw profile
    E   evaluation of data structures
    G   specify relevé group partition
    M   specify CST group partition
    T   create structured table
    C   return to session menu
Type option: o


ORDINATION SETTINGS                                          ufrgs5re5ch
------------------------------------------------------------------------
Resemblance function available: option 10 and its probabilities

Use:
    1   nominal resemblances
    2   partial resemblances
    3   probabilities of nominal resemblances
Type option no.: 1

Printout:
   - ordination scores y/n              y
    - correlation of scores and CSTs y/n   y
Results appended to file hd:csa:ufrgs5re5ch.prinda
```

Output 8:

```
------------------------------------------------------------------------
ORDINATION
------------------------------------------------------------------------
Tue Jul 21 11:53:54 1992
Based on nominal resemblances.

Level: 1
Eigenvalues:        1.182      0.6733       0.399      0.2744
```

```
    Percent:          46.7         26.6         15.8         10.8

Scores  of  relevés  on  the  first  4  components:
                    a            b            c            m            k
Axis  1:      -0.2246       0.1366      -0.4146       -0.387       0.8896
Axis  2:      -0.1014       0.6167      -0.4372       0.1714      -0.2495
Axis  3:      -0.3667       -0.174     -0.04386       0.4675       0.1171
Axis  4:       0.3334      -0.1985      -0.3169       0.1498      0.03213

Level:  2
Eigenvalues:          1.266       0.4428       0.1289      0.07846
    Percent:           66.1         23.1         6.73          4.1

Scores  of  relevés  on  the  first  4  components:
                    a            b            c            m            k
Axis  1:      -0.1906     -0.09083      -0.3747      -0.3298       0.9859
Axis  2:       0.2695       0.4045      -0.1523      -0.4145      -0.1072
Axis  3:      -0.2762       0.1934       0.1185     -0.03388     -0.00187
Axis  4:     -0.03479       0.1024      -0.2035       0.1577     -0.02187

Level:  3
Eigenvalues:          1.233       0.3503       0.1091      0.03703
    Percent:           71.3         20.3         6.31         2.14

Scores  of  relevés  on  the  first  4  components:
                     a            b            c            m            k
Axis  1:       -0.1903      -0.1086      -0.3578      -0.3201       0.9768
Axis  2:        0.2403       0.3559      -0.1361      -0.3741     -0.08604
Axis  3:       -0.2571       0.1941      0.07294    -0.006135    -0.003796
Axis  4:      -0.00248      0.05565      -0.1484       0.1083     -0.01313

Level:  4
Eigenvalues:          0.3002       0.2587        0.104     0.001311
    Percent:            45.2         38.9         15.7        0.197

Scores  of  relevés  on  the  first  4  components:
                     a            b            c            m            k
Axis  1:       -0.2666      -0.2242    -0.003485      0.07885       0.4155
Axis  2:        0.1684        0.146      -0.2716      -0.2805       0.2378
Axis  3:       -0.2169       0.2381     0.000515     -0.01365    -0.008104
Axis  4:      0.001271     0.002549     -0.02598      0.02491    -0.002755

Level:  5
Eigenvalues:          0.2643       0.1235     0.002024
    Percent:            67.8         31.7        0.519

Scores  of  relevés  on  the  first  3  components:
                     a            b            c            m            k
Axis  1:        0.0118       0.1773      -0.2571      -0.2527       0.3206
Axis  2:       0.00967       0.2741    -0.009753     -0.06388      -0.2101
Axis  3:      -0.04021      0.01151     0.009791     0.008918     0.009993


Correlation  of  ordination  axes  and  CSTs:
Level  of  the  character  hierarchy: 1
CSTs    Axis 1       Axis 2       Axis 3       Axis 4
   1   -0.7538       0.3808     -0.03949        0.534
   2   -0.5507      -0.6467      -0.5041      -0.1565
   3   -0.7259     -0.08137      -0.6006       0.3252
   4   -0.7724       0.3375      -0.2922      -0.4517
   5    -0.929      -0.2136       0.2076      -0.2194
```

```
   6     0.673        0.2956      -0.6123        0.2911
   7   -0.9288       0.02872       0.2828        0.2378
   8   0.04233        0.9281      -0.1045       -0.3549
   9    0.1405        0.8402       -0.308       -0.4236
  10    0.1405        0.8402       -0.308       -0.4236
  11   -0.3561       -0.1756      -0.2317       -0.8881
  12    0.9551       -0.1326      0.07389       -0.2546
  13   -0.4263       -0.5957     -0.07763       -0.6763
  14   -0.4263       -0.5957     -0.07763       -0.6763
  15   -0.4595       -0.4205       0.7245       -0.2952
  16    0.9148       -0.3399       0.2073       0.06858
  17    0.9148       -0.3399       0.2073       0.06858
  18    0.9148       -0.3399       0.2073       0.06858
  19    -0.398        0.2336       0.8275        0.3198
  20    -0.398        0.2336       0.8275        0.3198
```

(... also for other levels up to level 5)

```
Level  of  the  character  hierarchy:  5
CSTs     Axis 1       Axis 2       Axis 3
   1   -0.8377      -0.2667       0.1923
   2   -0.4617        0.859      0.08059
   3    0.9988      0.04361     -0.02207
```

6.6.9. Draw scattergrams

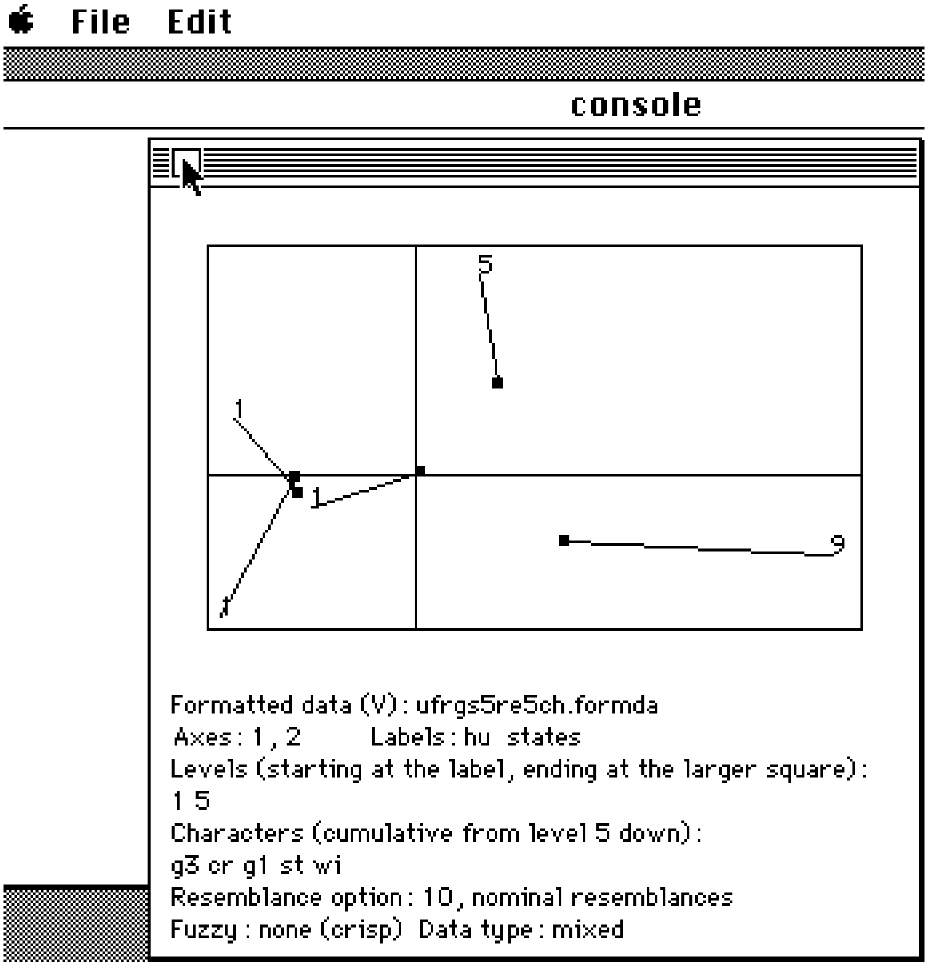
SYNCSA can produce two-dimension ordination scattergrams on the screen. The user is asked to specify the ordination axes (Run 15, line 22), the hierarchical levels to plot if there is more than one character (line 32), and the relevé labels (line 38). Since the scores are comparable, scattergrams defined at different hierarchical levels are superimposed. In this case the positions assumed by a relevé in the different ordinations are connected by lines, showing its trajectory through the community structures defined at different levels of the character hierarchy (Screen 3).

The labels identifying relevés on the scatter may be, provided the information is available in the session: (opt. 1) the relevé labels themselves; (opt. 2) the relevé group labels; (opt. 3) the state of a given environmental variable in the relevé, in which case the environmental variable is specified (as in Run 15, line 40); (opt. 4) the performance total in the relevé of CSTs presenting a given character state, in which case the character (its label) and the character states (all states or a specific one) will have to be appointed; (opt. 5) the perfomance total in the relevé of a CST group, in which case the user will be asked to specify the CST group (all groups or one of them). In options 3, 4 or 5 the states or quantities identifying the relevés may be shown as they are or rescaled in a 1-9 range (line 45). On choosing options 4 or 5, and if scattergrams for more than one character state or CST group is wanted, the user must click in the close box of the scattergram window to see the complete series of scattergrams. The size and appearance of scattergrams can be changed in the Preferences Menu.

Run 15:

```
 1  ANALYSIS OF COMMUNITY STRUCTURES                               ufrgs5re5ch
 2  ----------------------------------------------------------------------------
 3  Options:
 4     H  set data
 5     K  ranking characters
 6     R  resemblance
 7     U  cluster
 8     D  randomization
 9     O  ordination
10     S  draw scattergram
11     P  draw profile
12     E  evaluation of data structures
13     G  specify relevé group partition
14     M  specify CST group partition
15     T  create structured table
16     C  return to session menu
17  Type option: s
18
19  SCATTERGRAM SETTINGS                                           ufrgs5re5ch
20  ----------------------------------------------------------------------------
21  Specify 2 ordination components (up to component 3) to be plotted:
22  1 2
23
24  Hierarchical levels:
25     1  all levels
26     2  some levels
27  Type option no.: 2
28
29  Select levels:
30  (enter 0 for exclusion or 1 for inclusion)
31  1  2  3  4  5
32  1  0  0  0  1
33
34  Labels to use in the scatter:
35     1  relevé labels
36     3  environmental state
37     4  weight of character state in the relevé
38  Type option no.: 3
39
40  Enter environmental variable (label):   hu
41
42  Print hu state of the relevé:
43     1  as it is
44     2  as a class
45  Type option no.: 2
```

Screen 3:



```
Formatted data (V): ufrgs5re5ch.formda
Axes: 1, 2       Labels: hu  states
Levels (starting at the label, ending at the larger square):
1 5
Characters (cumulative from level 5 down):
g3 cr g1 st wi
Resemblance option: 10, nominal resemblances
Fuzzy: none (crisp)  Data type: mixed
```

6.6.10. Cluster analysis and dendrogram

Cluster analysis (option U) is offered if resemblances of relevés or CSTs are available in the session. When relevés are the objects to be classified, the user will specify the hierarchical level (if more than one) and the resemblance matrix among the ones that are available (nominal, partial or probabilities). When CSTs are the objects the hierarchical level is already determined when the resemblance is computed. If resemblances are defined in correlation coefficient terms, they will be transformed into distances (Eq. 5.2).

SYNCSA implements three clustering methods: (opt.1) single linkage, (opt. 2) complete linkage and (opt. 3) sum of squares (see Orlóci and Kenkel 1985, Wildi and Orlóci 1991). The algorithms are agglomerative. Initially each object has its own group; as the clustering process advances the groups increase in size. At each clustering step, the program finds the group pair that minimizes the clustering criterion and fuses it in a new group. The criterion for single linkage is the nearest neighbor dissimilarity, that is

$$\sigma = \mathrm{INF}\,[d_{hi}, \text{ for } h=1, ..., n-1, i=h+1, ..., n, \text{ and } g_h \neq g_i] \tag{6.1}$$

where d_{hi} is the dissimilarity of objects h and i, n is the total number of objects in the two groups, and g_h and g_i identify the group to which objects h or i belong. Single linkage fuses the group pair with minimum σ. Note that this criterion is not the average nearest neighbor dissimilarity defined by Eq. 4.3. Complete linkage is similar to single linkage, but the criterion minimized is the dissimilarity of the most dissimilar elements, that is

$$\sigma = \mathrm{SUP}\,[d_{hi}, \text{ for } h=1, ..., n-1, i=h+1, ..., n, \text{ and } g_h \neq g_i] \tag{6.2}$$

The sum of squares clustering algorithm (Orlóci 1967a) minimizes the increase in the average squared dissimilarity in the fused group j+k, which is defined by the relation

$$Q_{jk} = Q_{j+k} - Q_j - Q_k = \frac{n_j n_k}{n_j+n_k}\, d_{jk}^2 \tag{6.3}$$

In these, Q_{j+k}, Q_j and Q_k are within group sum of squared dissimilarities weighted by group sizes, n_j and n_k are the number of objects in groups j or k, and d_{jk}^2 is the squared centroid distance between groups j and k. The algorithm presumes d_{jk} is euclidean (see Section 2.4), and treats as such the dissimilarity functions defined in other terms (absolute value, information divergence and probability measures).

In the example (Run 16), relevé nominal resemblances on hierarchical level 1 are subjected to sum of squares clustering. The clustering process is shown on file *Session*.prinda (see Output 9), and the dendrogram is automatically drawn on the screen (Screen 4). The size of dendrograms can be changed in the Preferences Menu.

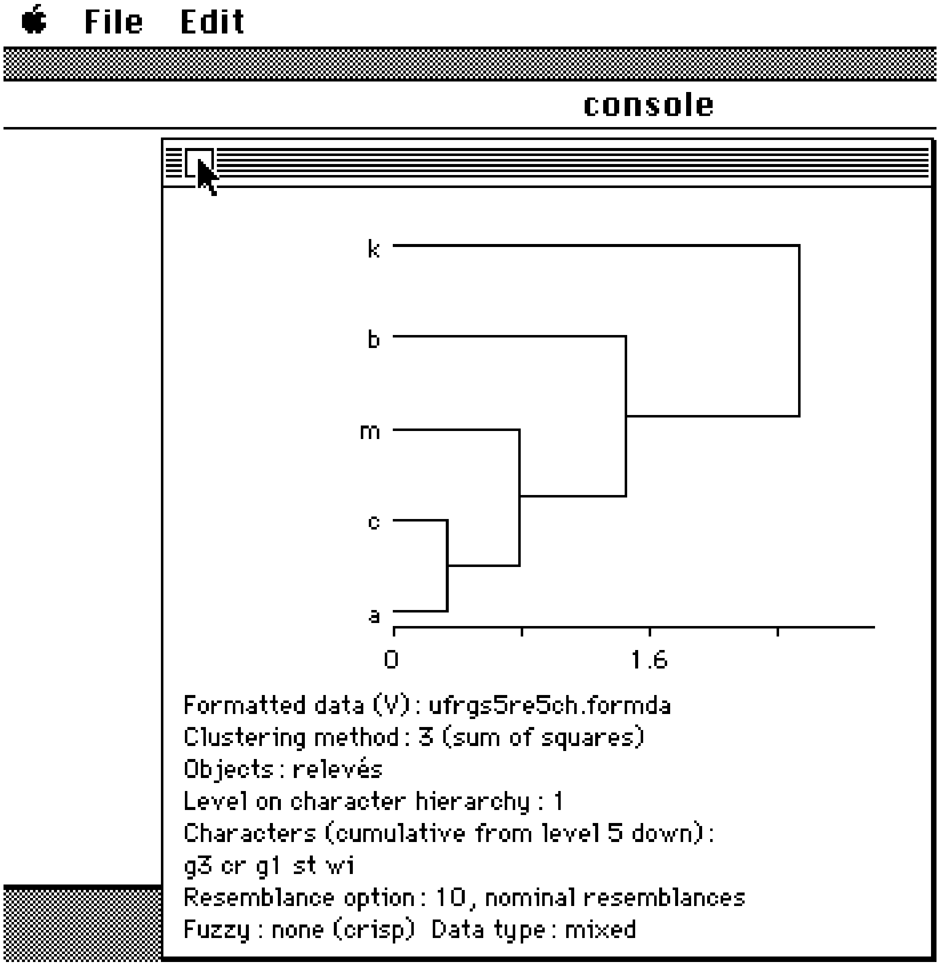
Run 16:

```
ANALYSIS OF COMMUNITY STRUCTURES                                  ufrgs5re5ch
-----------------------------------------------------------------------------
( ... )
Type option: u

CLUSTER ANALYSIS SETTINGS                                         ufrgs5re5ch
-----------------------------------------------------------------------------
Clustering of relevés(R) or CSTs(C)? r
Specify hierarchical level: 1

Resemblance function available: option 10
Use:
   1  nominal resemblances
   2  partial resemblances
Type option no.: 1

Clustering method:
   1  single linkage (nearest neighbor)
   2  complete linkage
   3  sum of squares
Type option no.: 3
Results appended to file hd:csa:ufrgs5re5ch.prinda
```

Output 9:

```
-----------------------------------------------------------------------------
CLUSTER ANALYSIS
-----------------------------------------------------------------------------
Tue Jul 21 11:54:00 1992
Objects: relevés
Based on nominal resemblances.
Resemblance option: 10 (squared chord distance)
Clustering method: 3 (sum of squares)
Level on character hierarchy: 1
Fusion directory:
Step  SSwithin     SSbetween          Objects in group
1     0.3379       0.3379             a     c
2     0.7857       0.4477             a     c     m
3     1.443        0.6578             a     c     m     b
4     2.529        1.086              a     c     m     b     k

The dendrogram places the objects from the bottom up as ordered in the last
clustering step.
```

Screen 4:



```
Formatted data (V): ufrgs5re5ch.formda
Clustering method: 3 (sum of squares)
Objects: relevés
Level on character hierarchy: 1
Characters (cumulative from level 5 down):
g3 cr g1 st wi
Resemblance option: 10, nominal resemblances
Fuzzy: none (crisp)  Data type: mixed
```

6.6.11. Structured table

This option is offered when at least a partition (relevés or CSTs) is available in the session. It rearranges a vegetation table according to relevé and CST groups. In the example (Run 17), a group partition of CSTs is specified (lines 7-21), option T is picked (line 40) and the table is created. The hierarchical level in this case is the one associated with the CST group partition. If no partition of CSTs is available, but there is one of relevés, the hierarchical level will be asked upon selecting option T. The table is appended to file *Session*.prinda (see Output 10).

Run 17:

```
ANALYSIS OF COMMUNITY STRUCTURES                                 ufrgs5re5ch
----------------------------------------------------------------------------
( ... )
Type option: m
Specify hierarchical level: 1

SET GROUP PARTITION                                              ufrgs5re5ch
----------------------------------------------------------------------------
Specify the number of groups of CSTs: 4

Labels for groups:
   1  given by default
   2  specify new labels
Type option no.: 1
Are the CSTs ordered by groups? y/n n

Enter group membership (label) for CSTs:
CSTs:   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17
18  19  20
Group:     1    2    1    1    1    3    1    4    4    4    2    3    2    2    1    3    3
3   1   1


ANALYSIS OF COMMUNITY STRUCTURES                                 ufrgs5re5ch
----------------------------------------------------------------------------
Options:
   H  set data
   K  ranking characters
   R  resemblance
   U  cluster
   D  randomization
   O  ordination
   S  draw scattergram
   P  draw profile
   E  evaluation of data structures
   G  specify relevé group partition
   M  specify CST group partition
   T  create structured table
   C  return to session menu
Type option: t
Table appended to file ufrgs5re5ch.prinda
```

Output 10:

```
-----------------------------------------------------------------
STRUCTURED  TABLE
-----------------------------------------------------------------
Level  on  character  hierarchy:  1
Relevés  and  CSTs  are  arranged  according  to  groups.
Relevé  group  label  below  each  relevé  label.

CST    Gr     g3     cr     g1     st     wi      a      c      m      b      k
                                                  1      1      1      2      3
---    --     --     --     --     --     --     ---    ---    ---    ---    ---
1      1       3      2      3      3      2      5      2      5      3      0
3      1       3      3      3      3      1      5      3      2      2      0
4      1       3      1      1      1      3      2      3      2      3      0
5      1       3      1      1      3      2      3      6      5      2      0
7      1       3      1      1      1      1      2      2      3      1      0
15     1       3      1      1      2      4      0      3      3      0      1
19     1       2      1      2      3      3      0      0      3      0      0
20     1       3      1      3      3      2      0      0      1      0      0
2      2       2      1      1      3      2      5      7      0      0      0
11     2       3      1      1      3      3      0      2      0      1      0
13     2       3      2      3      3      1      0      5      0      0      0
14     2       2      1      1      1      3      0      1      0      0      0
6      3       1      2      1      3      3      5      0      0      5      5
12     3       3      1      2      3      3      0      1      0      2      5
16     3       1      1      2      1      4      0      0      0      0      5
17     3       3      1      3      3      1      0      0      0      0      5
18     3       3      2      1      3      2      0      0      0      0      5
8      4       2      1      1      3      3      0      0      2      8      0
9      4       1      1      1      1      4      0      0      0      3      0
10     4       2      1      2      3      2      0      0      0      1      0


CST/Species  correspondence,  level  1:
CST#   Species
----   ----------------------------------------
1      Anla
3      Pimo   Arfi
4      Rusp
5      Sege   Cose   Erne
7      Rehi   Fare   Sopt
15     Dein   Batr
19     Erho
20     Sege
2      Pano
11     Hepu
13     Spin
14     Rihu
6      Axaf
12     Pano   Sysp   Papu
16     Cebi
17     Elgl
18     Anla
8      Pano
9      Cebi
10     Erho
```

6.7. Analysis of environmental structures

The hierarchy of information that determines the options that will be available at a given step in the analysis is similar to the one guiding the analysis of community structures (Fig. 6.6.1), but less complex. Ranking variables (opt. K) is available independently from other options. Ordination (opt. O) and cluster (opt. U) use the matrix generated by resemblance of relevés (opt. R). Scattergrams (opt. S) are drawn using the ordination scores. Dendrograms are automatically generated after cluster analysis. These options are offered provided there is more than one environmental variable in file *Session*.formda. The user can also specify a group partition of relevés (opt. G). The partition is shared within the session, so a relevé partition set here will erase another one that may have been set elsewhere in the session. The hierarchy of options can be followed in the examples (Runs 18-19).

6.7.1. Ranking variables and drawing profiles

A similar approach as in character ranking is adopted. Environmental structure is defined by a matrix of relevé euclidean distances ($\mathbf{\Delta}$), after centering and normalization within variables (Eqs. 6.4 and 6.5). The algorithm for minimum structural redundancy (opt. 1) is equivalent to the one described in Section 4.3.3.2, but the function optimized is $\rho(\mathbf{\Delta}_k;\mathbf{\Delta}_{k-1})$, where the relevé distance matrix $\mathbf{\Delta}_k$ is based on k variables and $\mathbf{\Delta}_{k-1}$ is based on k-1 variables. The algorithm for convergence of community groups (opt. 3) ranks the variables so to minimize the average group dissimilarity (Eq. 4.2, but dissimilarities are environmental). A relevé group partition must be available for this option be offered.

The options for congruence of environmental resemblance / community resemblance (opt. 2) and environmental resemblance / community ordination axes (opt. 4) use information generated in the analysis of community structures. The algorithm is akin to the one described in Section 4.3.1 and maximizes the same function $\rho(\mathbf{D}_i;\mathbf{\Delta})$, but here the community structure $\mathbf{D}_i$ is given and the set of variables defining $\mathbf{\Delta}$ increases in size at each ranking step (see part of the ranking process in Output 12). Under option 2, the community structure $\mathbf{D}_i$ is defined by the relevé resemblances (nominal, partial, or probabilities), hierarchical level i, as produced in the analysis of community structures. Under option 4 it is defined by relevé euclidean distances computed on the basis of a number of community ordination components the user is asked to select (Run 18, lines 31-35).

Option 4 is specially useful to interpret in environmental terms nonlinear configurations frequently revealed by community ordination, since the algorithm finds the subset of variables that defines a parallel environmental structure maximally associated to the point configuration. The state of these variables in the relevés can then be used for labeling relevés in the scattergram as illustrated in Screen 3. This is and alternative to the traditional correlation analysis involving eigenaxes and environmental variables (Section 6.6.5).

The example follows option 4 (Run 18, line 28). The results are stored on file *Session*.eRank (Output 12). The first part of this file, which contains SEF values and the ranked variables, for the different hierarchical levels, is accessed by the draw profile procedure (opt. P). The second part is the printable output showing the ranking process.

The profile is drawn with the structure evaluation function (SEF) in the vertical axis and the sets of variables in the horizontal axis, starting with the set containing one variable, ending with the set containing all variables (see Screen 5). When there are several levels in the character hierarchy, the user should specify one of them (Run 18, line 57).

Run 18:

```
 1  SESSION MENU                                                 ufrgs5re5ch
 2  ------------------------------------------------------------------------
 3  Options:
 4     P  analysis of population structures
 5     V  analysis of community structures
 6     E  analysis of environmental structures
 7     S  set preferences
 8     C  return to main menu
 9  Type option: e
10
11  ANALYSIS OF ENVIRONMENTAL STRUCTURES                         ufrgs5re5ch
12  ------------------------------------------------------------------------
13  Options:
14      R  resemblance of relevés
15      K  ranking variables
16      G  specify relevé group partition
17      C  return to session menu
18  Type option: k
19
20
21  RANKING OF ENVIRONMENTAL VARIABLES                           ufrgs5re5ch
22  ------------------------------------------------------------------------
23  Environmental variables ranking criterion:
24      1  structural redundancy
```

```
     2   congruence of environmental resemblance/community resemblance
    3   convergence of community groups
     4   congruence of environmental resemblance/community ordination axes
Type option no.: 4

Choose number of ordination components to use in distances:
Level 1, rank=4, dimensions:2
Level 2, rank=4, dimensions:2
Level 3, rank=4, dimensions:2
Level 4, rank=4, dimensions:2
Level 5, rank=3, dimensions:2
Results on file hd:csa:ufrgs5re5ch.eRank


ANALYSIS OF ENVIRONMENTAL STRUCTURES                          ufrgs5re5ch
-----------------------------------------------------------------------------
Options:
   R  resemblance of relevés
   K  ranking variables
   P  draw profile
   G  specify relevé group partition
   C  return to session menu
Type option: p

PROFILE SETTINGS                                              ufrgs5re5ch
-----------------------------------------------------------------------------
Profile's vertical scale:
   1  based on actual data
   2  user specified
Type option no.: 1
-----------------------------------------------------------------------------
Hierarchical level to use (among 5 levels):
1
```

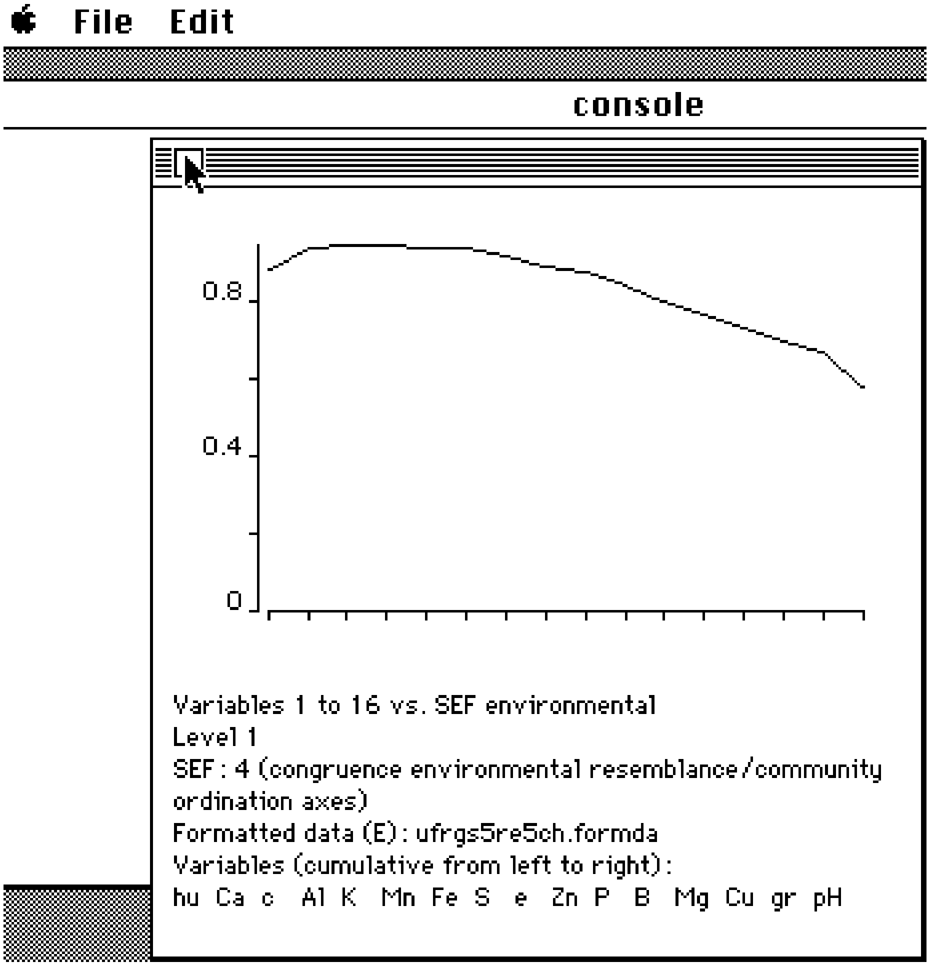
Output 12:

```
    5
0.8825       0.9404       0.9452       0.9465       0.9429       0.9434
      0.9201       0.8926       0.8775       0.8442       0.8029       0.7703
      0.7316       0.6972       0.6708       0.5814
0.7675       0.8475       0.8922       0.853        0.8279       0.8149
      0.8466       0.8383       0.8026       0.7748       0.7202       0.6847
      0.6154       0.4788       0.3566       0.2159
0.777        0.8625       0.9073       0.8708       0.85         0.8322
      0.8612       0.8451       0.8152       0.7899       0.7371       0.7047
      0.6287       0.4957       0.3724       0.232
0.6966       0.7097       0.7942       0.7774       0.7595       0.7264
      0.7096       0.6777       0.6724       0.6584       0.5824       0.5026
      0.4043       0.272        0.1359       0
0.7481       0.7833       0.7487       0.7628       0.7322       0.735
      0.6872       0.6667       0.6427       0.5723       0.513        0.4727
      0.4115       0.362        0.2598       0.136
hu    Ca    c     Al    K     Mn    Fe    S     e     Zn    P     B     Mg
      Cu    gr    pH
hu    Al    e     K     Mn    P     c     gr    B     Ca    pH    Cu    S
      Zn    Fe    Mg
hu    Al    e     K     Mn    P     c     gr    B     Ca    pH    Cu    S
      Zn    Fe    Mg
```

```
Al     hu     P      e      c      gr     K      pH     Ca     Cu     S      Mn     B
       Fe     Zn     Mg
hu     S      Al     c      gr     K      Ca     P      Cu     B      e      Mg     Fe
       Mn     Zn     pH

The variables listed above are ordered by level and decreasing rank.

CHARACTER-BASED COMMUNITY ANALYSIS                           SYNCSA v.1.0
------------------------------------------------------------------------
ENVIRONMENTAL VARIABLES RANKING
------------------------------------------------------------------------
Wed Jul 22 15:08:43 1992
Session:         ufrgs5re5ch
Formatted data:  ufrgs5re5ch.formda
Ranking criterion: 4 (congruence environmental resemblance/community ordina-
tion axes)
Environmental data structure:
    Number of variables: 16
    Vector transformation: 6 (centering and normalization within columns)
    Resemblance option: Euclidean distances
    Number of vegetation ordination components considered (level 1 to 5):
   2  2  2  2  2
Vegetation data structure:
    Data component: mixed
    Fuzzy transformation: none (crisp)
    Resemblance option: nominal resemblances
    Based on nominal resemblances

SEF          variable subset used:
-0.6639      pH
-0.4523      P
0.3095       K
0.5271       c
0.4305       Al
0.3027       Ca
0.07938      Mg
0.2121       S
0.01148      Zn
0.08794      Cu
-0.2282      B
0.006516     Mn
0.1349       Fe
0.8825       hu
0.2437       e
-0.3454      gr
0.05163      hu     pH
```

... (the process continues for the whole set of variables and hierarchical levels)

Screen 5:



```
Variables 1 to 16 vs. SEF environmental
Level 1
SEF : 4 (congruence environmental resemblance/community
ordination axes)
Formatted data (E): ufrgs5re5ch.formda
Variables (cumulative from left to right):
hu Ca c  Al K  Mn Fe S  e  Zn P  B  Mg Cu gr pH
```

6.7.2. Resemblance of relevés

The resemblance function implemented is the euclidean distance. Vector transformations can be specified (Run 19, line 20), which may be centering, normalization or both, and within relevés or variables. Centering is defined by

$$x_{ij} = x_{ij} - \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{6.4}$$

and normalization by

$$x_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^{2}}} \tag{6.5}$$

When the transformation is within relevés, x_{ij} is the state of variable i in relevé j, and n is the number of variables. When the transformation is within variables, x_{ij} is the state of variable j in relevé i, and n is the number of relevés. The printable output is stored on file *Session*.ePrinda (see Output 13).

6.7.3. Ordination and scattergrams

For ordination (option letter O), SYNCSA applies the eigenordination method (Section 5.2) to the environmental relevé resemblance matrix available in the session. The user can select (Run 19, lines 40-41) whether ordination scores and the correlation coefficients between them and environmental variables are to be included in the printout (see results in Output 13).

Scattergrams of ordination scores are produced by option S. The same guidelines given in Section 6.6.9 apply here. Screen 6 is the result of the options selected in Run 19 (lines 58-69). Note that information from the analysis of community structures is needed when the user selects labels as the weight of a character state (opt. 4 in line 69) or the performance of a CST group (opt. 5) in the relevé.

6.7.4. Cluster analysis

The instructions given in Section 6.6.10 for cluster analysis of CSTs apply here. The example (Run 18, lines 78-85) follows sum of squares clustering. The results are in Output 13, and the dendrogram is automatically drawn on the screen (Screen 7).

Run 19:

```
1
2  ANALYSIS OF ENVIRONMENTAL STRUCTURES                         ufrgs5re5ch
3  ------------------------------------------------------------------------
4  Options:
5     R  resemblance of relevés
6     K  ranking variables
```

```
   P  draw profile
   G  specify relevé group partition
   C  return to session menu
Type  option:  r

Vector  transformation:
      0   none
      1    centering  within  relevés
      2    centering  within  variables
      3    normalization  within  relevés
      4    normalization  within  variables
      5    centering  and  normalization  within  relevés
      6    centering  and  normalization  within  variables
Type  option  no.:  6

Printout:
   - relevé resemblances y/n  y
Results  on  file  hd:csa:ufrgs5re5ch.ePrinda


ANALYSIS OF ENVIRONMENTAL STRUCTURES                           ufrgs5re5ch
----------------------------------------------------------------------------
Options:
   R  resemblance of relevés
   K  ranking variables
   O  ordination
   U  cluster
   P  draw profile
   G  specify relevé group partition
   C  return to session menu
Type  option:  o

Printout:
   - ordination scores y/n              y
   - correlation of scores and variables y/n  y
Results  on  file  hd:csa:ufrgs5re5ch.ePrinda


ANALYSIS OF ENVIRONMENTAL STRUCTURES                           ufrgs5re5ch
----------------------------------------------------------------------------
Options:
   R  resemblance of relevés
   K  ranking variables
   O  ordination
   U  cluster
   S  draw scattergram
   P  draw profile
   G  specify relevé group partition
   C  return to session menu
Type  option:  s

SCATTERGRAM SETTINGS                                           ufrgs5re5ch
----------------------------------------------------------------------------
Specify  2  ordination  components  (up  to  component  4)  to  be  plotted:
1 2

Labels  to  use  in  the  scatter:
   1  relevé labels
   2  relevé group labels
   3  environmental state
   4  weight of character state in the relevé
```

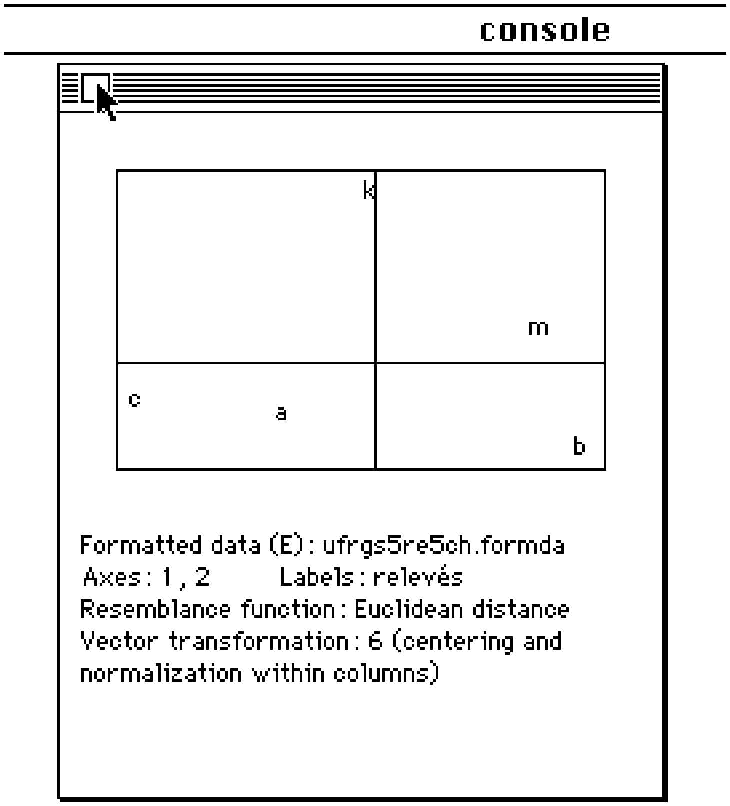
```
   5  performance of CST group in the relevé
Type  option  no.:  1


ANALYSIS  OF  ENVIRONMENTAL  STRUCTURES                        ufrgs5re5ch
--------------------------------------------------------------------------
Options:
(...)
Type  option:  u

CLUSTER  ANALYSIS  SETTINGS                                    ufrgs5re5ch
--------------------------------------------------------------------------
Clustering  method:
   1  single linkage (nearest neighbor)
   2  complete linkage
   3  sum of squares
Type  option  no.:  3
Results  on  file  hd:csa:ufrgs5re5ch.ePrinda
```

Output 13:

```
CHARACTER-BASED  COMMUNITY  ANALYSIS                          SYNCSA v.1.0
--------------------------------------------------------------------------
EXPLORATION  OF  ENVIRONMENTAL  DATA  STRUCTURE
--------------------------------------------------------------------------
Wed  Jul  22  02:12:51  1992
Session:          ufrgs5re5ch
Formatted  data:  ufrgs5re5ch.formda
Number  of  environmental  variables:  16

Vector  transformation:  6  (centering  and  normalization  within  columns)
Resemblance  option:  Euclidean  distances
Resemblance  matrix:
    0  2.682  2.264  2.633  2.871
           0  3.209  2.485  3.226
                  0  3.074  3.021
                         0  2.655
                                0

--------------------------------------------------------------------------
ORDINATION  OF  RELEVES  (environmental)
--------------------------------------------------------------------------
Wed  Jul  22  02:12:52  1992
Eigenvalues:        5.955  5.176  2.614  2.256
    Percent:         37.2   32.3   16.3   14.1

Scores  of  relevés  on  the  first  4  components:
                     a            b            c            m            k
Axis  1:      -0.6544        1.339       -1.637        1.025     -0.07267
Axis  2:      -0.6341       -1.034        -0.48       0.3093        1.838
Axis  3:       0.1274      -0.8284      0.07363        1.238      -0.6108
Axis  4:       -1.205       0.2602       0.8276       0.2056     -0.08797

Correlation  of  ordination  axes  and  environmental  variables:
Variable        Axis 1       Axis 2       Axis 3       Axis 4
      pH        0.4037       0.3205       0.1745        0.839
       P       -0.1083      -0.7215      -0.1994      -0.6541
       K       -0.1329      -0.9214      -0.1602       0.3281
```

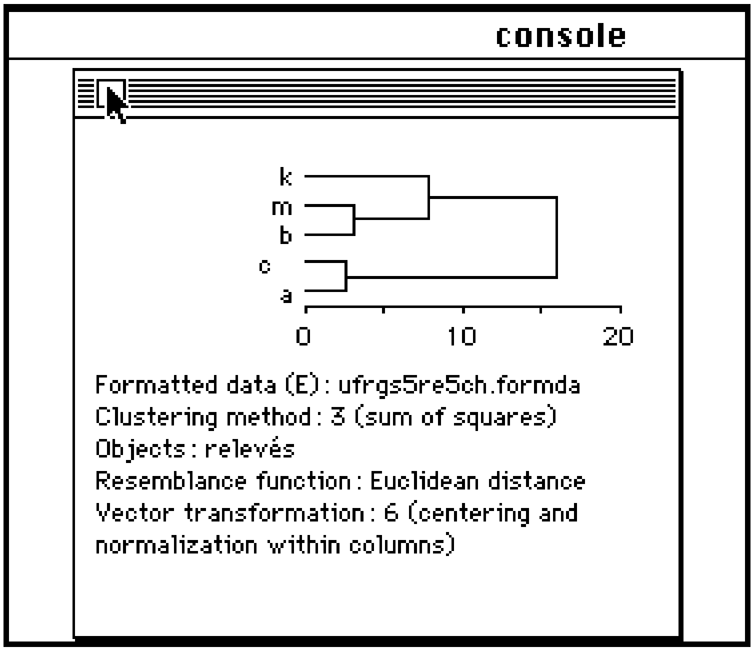
```
     c       0.724      0.2984     -0.6123     -0.1085
    Al     -0.1911      0.8536     -0.2099      0.4367
    Ca      0.8222     -0.5583   0.0007184      0.1104
    Mg      0.9301     -0.3249   -0.009323      0.1708
     S       0.602     -0.5687     -0.5257      0.1945
    Zn      0.9191     -0.2078      0.3108     -0.1243
    Cu      0.7685      0.3994     -0.1208      -0.485
     B     -0.1113     -0.6073     -0.4262      0.6612
    Mn      0.8349      -0.321      0.4461     0.03125
    Fe      0.8875      0.3417     -0.2746     -0.1417
    hu      0.2734      0.6495     -0.7088     0.03138
     e      0.2653      0.9532      0.1408     0.03466
    gr      0.4695       0.152      0.8562       0.153


-----------------------------------------------------------------
CLUSTER ANALYSIS (environmental)
-----------------------------------------------------------------
Wed Jul 22 02:13:04 1992
Objects: relevés
Clustering method: 3 (sum of squares)
Fusion directory:
Step  SSwithin    SSbetween        Objects in group
1     2.562       2.562            a     c
2     3.087       3.087            b     m
3     7.877       4.789            b     m     k
4     16          5.561            a     c     b     m     k

The dendrogram places the objects from the bottom up as ordered in the last
clustering step.
```

Screen 6:



Formatted data (E): ufrgs5re5ch.formda
Axes: 1, 2 Labels: relevés
Resemblance function: Euclidean distance
Vector transformation: 6 (centering and normalization within columns)

Screen 7:



Formatted data (E): ufrgs5re5ch.formda
Clustering method: 3 (sum of squares)
Objects: relevés
Resemblance function: Euclidean distance
Vector transformation: 6 (centering and normalization within columns)

6.8. Analysis of population structures

Within this option only the method for ranking characters based on redundancy at the populational level (Section 4.3.3.1) is available. A sample run is presented (Run 20). The output is stored on file *session.*ChRank (Output 14).

Run 20:

```
SESSION MENU                                                  ufrgs5re5ch
-------------------------------------------------------------------------
Options:
    P   analysis of population structures
    V   analysis of community structures
    E   analysis of environmental structures
    S   set preferences
    C   return to main menu
Type option: p

ANALYSIS OF POPULATION STRUCTURES                             ufrgs5re5ch
-------------------------------------------------------------------------
Options:
    K   ranking characters
    C   return to session menu
Type option: k
Results on file hd:csa:ufrgs5re5ch.ChRank
```

Output 14:

```
st g3 cr g1 wi
The characters listed above are ordered by decreasing rank.

CHARACTER-BASED COMMUNITY ANALYSIS                           SYNCSA v.1.0
-------------------------------------------------------------------------
CHARACTER RANKING
-------------------------------------------------------------------------
Wed Jul 22 17:21:08 1992
Session:          ufrgs5re5ch
Formatted data:  ufrgs5re5ch.formda
Character ranking criterion: redundancy in population data structure
SEF          Char. out    Character subset used:
0.9669       wi     g3     cr     g1     st
0.8809       st     g3     cr     g1     wi
0.8888       g1     g3     cr     st     wi
0.8846       cr     g3     g1     st     wi
0.884        g3     cr     g1     st     wi
0.8479       st     g3     cr     g1
0.8559       g1     g3     cr     st
0.8515       cr     g3     g1     st
0.8501       g3     cr     g1     st
0.7826       st     g3     cr
0.7938       cr     g3     st
0.7874       g3     cr     st
```

```
0.718        st    g3
0.7262       g3    st
Elapsed time: 1.21667 seconds
```

6.9. Preferences menu

Several options related to the appearance of the printable outputs and graphs can be selected in this menu when the default settings are not suitable (see Run 21). The size (in millimeters) of scattergrams, profiles and dendrograms can be changed provided you set them within the limits of the computer screen you are using. Enter zero to select the default sizes (as in Run 21, line 37). The size (in pixels) of the small squares representing objects in scattergrams can also be reset (line 27). Profiles may show lines with full squares (line 38) instead of smooth lines (as in Fig. 7.1.2.1), in which case the size (in pixels) of the squares will be inquired. Regarding the printable output files (lines 56-65), you can change the number of significant digits (the default is 6) to which figures are rounded (line 59), the maximum text width in the printout (line 61) and the maximum number of ordination components that are printed (line 63). You can also set the program to create output files, which by default are not created, containing the matrices used to evaluate structures (line 64), and the tables and resemblances generated by randomization (line 65). Keep in mind that the latter option may create a huge file on disk depending on the data size and number of iterations. The settings are stored on file SYNCSA.prefs in the application's folder, and are recalled each time you start the application.

Run 21:

```
 1  CHARACTER-BASED COMMUNITY ANALYSIS                           SYNCSA v.1.0
 2  ----------------------------------------------------------------------
 3  Main menu:
 4     N  start new session
 5     O  attach to old session
 6     M  set macro mode
 7     S  set preferences
 8     X  quit
 9  ----------------------------------------------------------------------
10  Type option: s
11
12  PREFERENCES
13  ----------------------------------------------------------------------
14  Set preferences of:
15     S  scatter diagram
16     P  profiles
```

```
    D  dendrograms
    F  output files
    C  return to main menu
Type option: s
----------------------------------------------------------------------
Dimensions of the scattergram in mm (zero for default 101 x 66 mm):
 width (max. 225):  80
 height (max. 121): 55
Size of quadrats representing points (pixels):
    (enter zero for default of 3 pixels)
0


PREFERENCES
----------------------------------------------------------------------
Set preferences of:
(...)
Type option: p
----------------------------------------------------------------------
Dimensions of the profiles in mm (zero for default 101 x 66 mm):
 width (max. 225):  0
Draw points on profiles (default is no)? y/n n


PREFERENCES
----------------------------------------------------------------------
Set preferences of:
(...)
Type option: d
----------------------------------------------------------------------
Dimensions of the dendrograms in mm (zero for default 101 x 66 mm):
 width (max. 225):  80
 height (max. 121): 33


PREFERENCES
----------------------------------------------------------------------
Set preferences of:
(...)
Type option: f
----------------------------------------------------------------------
Number of significant digits on printouts (enter zero for default):
4
Maximum line length (characters) on printout files (zero for default):
100
Maximum number of ordination components on printout (zero for default):
0
Save matrices used to evaluate structures? y/n               n
Save tables and resemblances generated by randomization? y/n n
```

6.10. Setting the macro mode

A macro is a text file containing the keyboard input for a run. It is specially useful for performing a series of lengthy analyses. To run a macro, select option M in the main menu, and R in the macro mode options menu. The name of the file containing the macro is asked (Run 23). From this point on the keyboard input is taken from the macro file and the screen output is redirected to a file named MacroRunReport, placed in the same folder where the macro is located, until the program reads a stop running macro instruction (option R in the macro mode menu) in the macro file. Other output files remain the same as already explained.

A macro can be created by using any text editor, or by the program itself by selecting option N (start creating new macro) in the macro mode options menu (Run 22). A file name is asked to store the macro. All the keyboard input, in any menu, is then stored in the macro file until option N (stop creating macro) in the macro options menu is selected. The macro file must contain at its end the option to stop running the macro. This is automatically added when the macro is created by the program. The macro file so created can be edited in any text editor.

Run 22:

```
SET MACRO MODE
--------------------------------------------------------------------
   N  start creating new macro
   R  run macro on file
   C  return to main menu
Type option: n

Macro file name:     ufrgs5re5ch.macroPrefs
Enter pathway:       hd:csa
```

... (the main menu appears and analysis may proceed to any options)

```
SET MACRO MODE
--------------------------------------------------------------------
   N  stop creating macro
   C  return to main menu
Type option: n
Macro saved on file hd:csa:ufrgs5re5ch.macro
```

Run 23:

```
SET MACRO MODE
-----------------------------------------------------------------------
    N   start creating new macro
    R   run macro on file
    C   return to main menu
Type option: r

Macro file name:     ufrgs5re5ch.macro
Enter pathway:       hd:csa
```

... (the program performs the options stored in the macro)

```
Macro:               hd:csa:ufrgs5re5ch.macroPrefs
Report on file:  hd:csa:MacroRunReport
```

Chapter 7

EXAMPLES OF ANALYTICAL STRATEGIES

This chapter describes two analyses using the application program SYNCSA. The first uses data from grassland communities of the Brazilian Campos and is concerned with an ecological comparison of species-based vs. character-based taxonomies. The second is a study in the measurement of convergence and ecological trends in community structures using data from the Brazilian Caatinga and the Argentine Chaco and Monte.

7.1. Species-based and character-based analysis of grassland communities

The question here addressed is if there is any advantage in using character-based instead of species-based taxonomies in small scale community studies. For the same set of relevés, and the same species-based primary taxonomy (Section 1.2), the analytical taxonomy is defined by the species name in one case, and by a character set in the other. Using the ranking method of maximum congruence with the environmental structure (Section 4.3.1), an optimal character subset is found to define the analytical, character-based taxonomy. The congruence revealed based on the optimal character-based taxonomy is then compared with the congruence revealed by the species-based taxonomy. A probabilistic assessment is also performed, and eigenordination applied. The community structure so represented can be accounted for by the numerical relationships of variation in environmental conditions and, which is unique to the character-based approach, the variation in CSTs and plant characters.

7.1.1. Data set

The example uses data containing 15 relevés taken in a 4 ha natural grassland situated at 30°05' S and 51°13' W, in the experimental station of the Federal University of Rio Grande do Sul (UFRGS), near Porto Alegre, Brazil. Physiognomically the vegetation belongs to the Campos (Rambo 1956, Cabrera 1971). The climate is Koeppen's Cfa, with 1322 mm mean annual precipitation and 19.3°C mean annual temperature. The soil is yellow-red latosol on convex slopes, and hydromorphic on low-lying land. The area has been exposed to controlled, medium grazing pressure by cattle. The sampling used quadrats, 0.5 m.sq. each, located along relief gradients. The sampling and community description followed the Braun-Blanquet (1964) method. Estimates of cover-abundance and character descriptions were made for 26 species (see Appendix A). The species were the ones having had a presence over 70% in at least one of the relevé groups defined in an early study (1986) on the same quadrats (Pillar 1988, Pillar, Jacques and Boldrini 1989). For practical reasons, other species encountered in the quadrats were not recorded. The character description was local to each relevé, using the defining characters given in Appendix A. The environmental information consists of relief position, soil moisture and 13 soil macro and micronutrients, and grazing intensity estimates (see Appendix B).

7.1.2. Results

Character-based analysis (Fig. 7.1.2.1) reveals a stronger congruence with the environmental data structure than the species-based analysis does (arrows' level in Fig. 7.1.2.1). This was expected, since the character order is optimal for maximum expression of the characters as indicators of specific environmental conditions, which the species in general do not have. Structural congruence increases with hierarchical levels. By choice, the analysis proceeds using hierarchical level 8, which presents a structural congruence of 0.427, with a very low α (type I error) probability (0.007).

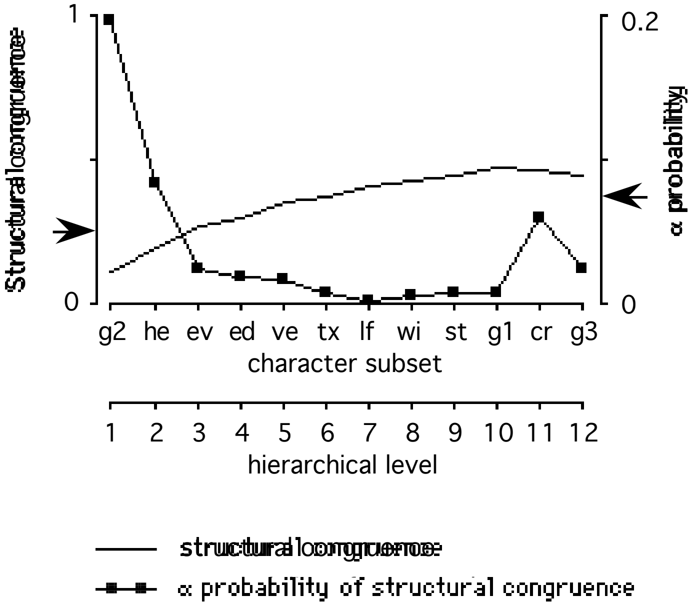


Figure 7.1.2.1. Congruence $\rho(\mathbf{D}_i;\Delta)$ between vegetation and environmental data structures for the EEA/UFRGS sample (smooth line), and the corresponding α probabilities (connected full squares). Character order is optimal (Section 4.3.1). Probabilities are generated by randomization carried to 1000 iterations under the random composition hypothesis (Sections 2.8 and 3.4). Vegetation data structure is defined by the squared chord distance matrix ($\mathbf{D}_i$) between relevés on each level i. Environmental data structure (Δ) is defined by distances (not chord) between relevés using 16 environmental variables, after centering within variables and normalization. The defining character subset on each level is cumulative from the top down in the hierarchy, *e.g.*, only character g3 on level 12 and the complete character set on level 1. The character labels are identified in Appendix A. For comparison, the arrows indicate the congruence and its α probability found in species-based analysis. The two profiles were generated separately by SYNCSA using the "draw profiles" option in the analysis of community structures (Section 6.6.7), and then superimposed and edited using a commercial graphs editor (the smooth line profile, for instance, is the same profile produced in Run 13).

The eigenordination results for both character-based and species-based analysis are displayed in Figure 7.1.2.2. Both analyses show horseshoe shaped, very similar relevé configurations on the first two eigenaxes. The interpretation of these in environmental terms is helped by the method described in Section 6.7.1, which finds a subset of environmental variables maximally related to the community structure defined by the CST performances and ordination axes (Figure 7.1.2.3). Upon examination of the results in Figure 7.1.2.3, it is also obvious that even on the plane of the first two eigenaxes, the differences in structural congruence between species and character-based are retained, though reduced.

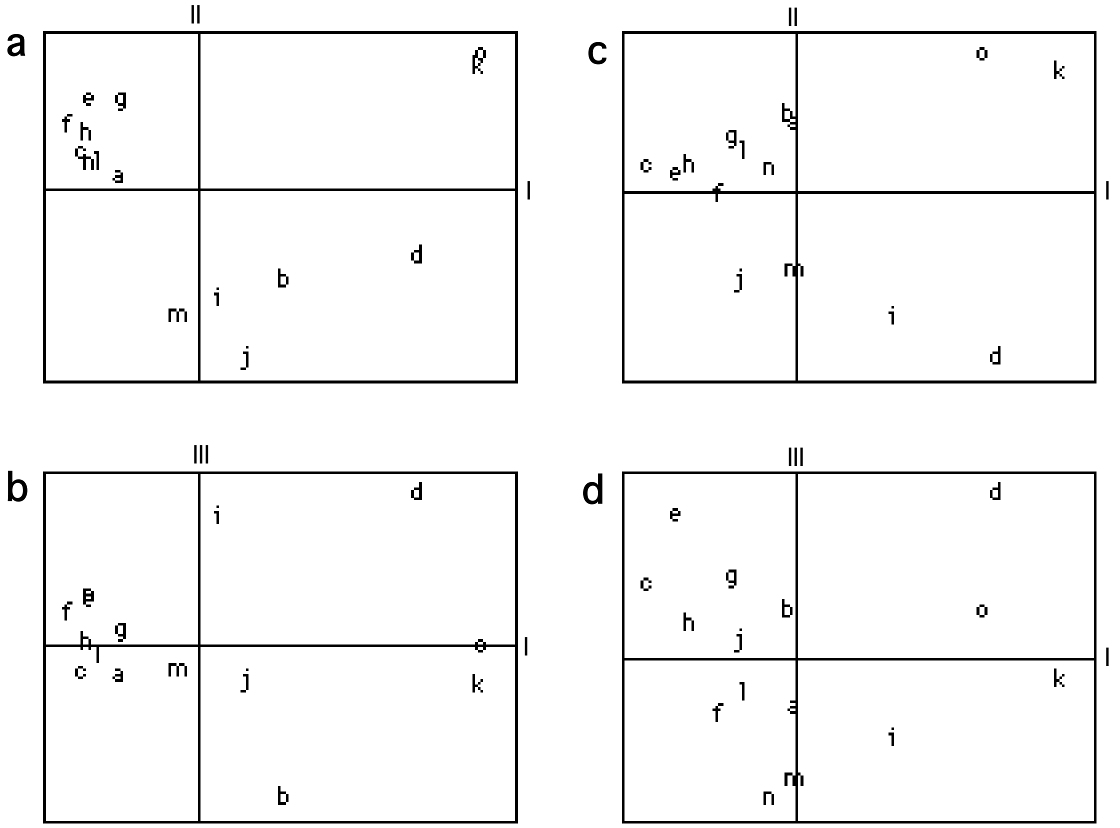


Figure 7.1.2.2. Eigenordination scattergrams of relevés for the EEA/UFRGS grassland data set. The analysis is character-based in (a, b) and species-based in (c, d). The character order in (a, b) is optimal for maximum congruence with the environmental data structure (Fig. 7.1.2.1); hierarchical level 8 is shown. The relevé labels (letters) are as in Appendix A. Eigenvectors 1, 2 and 3 are plotted. These axes account respectively for 31.81%, 20.09% and 12.73% of the trace (sum of all eigenvalues) in (a, b), and for 28.16%, 23.94% and 15.13% in (c, d). Scattergrams generated separately by SYNCSA (see Section 6.6.9).

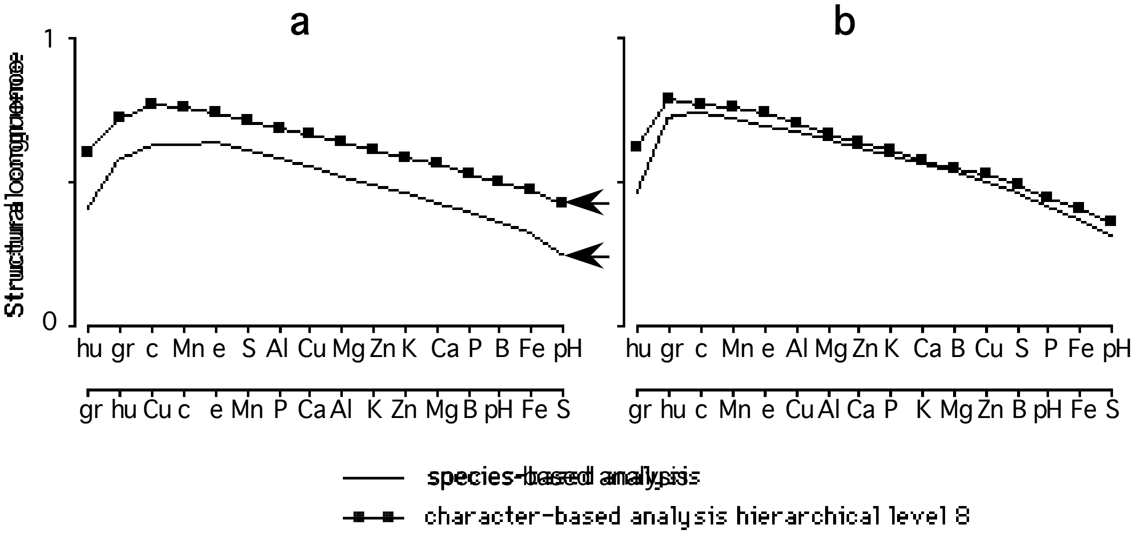


Figure 7.1.2.3. Graphs showing congruence between community data structure and environmental data structures. The community data structure is defined by the squared chord distances of the relevés, using CSTs on hierarchical level 8 or species performances in (a), and using the first two eigenordination axes (Fig.7.1.2.2 a, c) in (b). The environmental data structure is defined by distances (not chord) between relevés using the variables cumulatively from left to right, after centering within variables and normalization. Variable order is optimal (Section 6.7.1 explains the ranking process); the upper horizontal scale refers to the character-based analysis; the lower one refers to the species-based analysis. The arrows indicate the level of congruence obtained with the complete set of environmental variables. This is the same as in Figure 7.1.2.1 for the species-based and for the character-based on level 8 analyses. Environmental variables: soil moisture (hu), grazing intensity (gr), relief position (e), soil organic matter (c), soil pH, and soil available content of several elements (Cu, Mn, P, Ca, Al, K, Zn, Mg, B, Fe, S). In (a) congruence is maximal with variables soil moisture, grazing intensity and soil organic matter taken collectively. In (b) it is maximal with soil moisture and grazing intensity. Note the reduced gap between the species-based and the character-based analyses in (b). Profiles generated separately by SYNCSA (see Section 6.7.1)

The subset of the environmental variables that maximizes congruence (Figure 7.1.2.3) is mapped onto the community eigenordination scattergram (Figure 7.1.2.4), as in Gittins (1965), for visual evaluations.

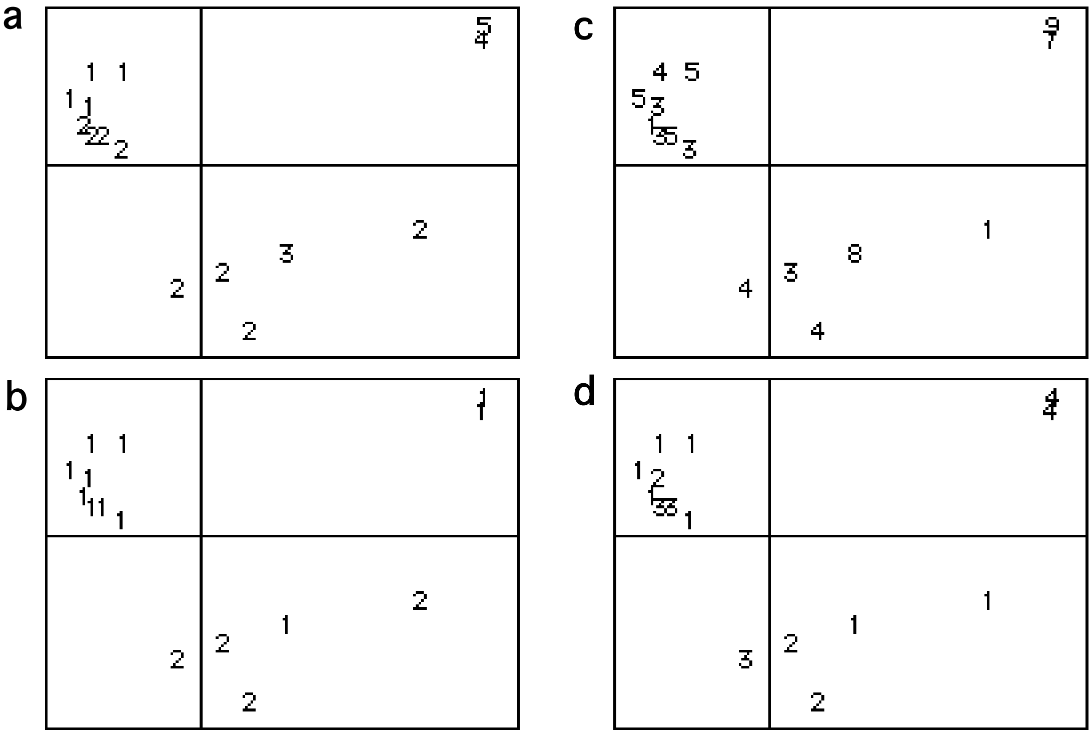


Figure 7.1.2.4. Variation of environmental variables within the eigenordination scattergram shown in Fig. 7.1.2.2 (a). The environmental variables include soil moisture (a), grazing intensity (b), soil organic matter (c), and relief position (d). These variables represent the ranks on the congruence scale (Fig. 7.1.2.3 b). In (a) the trend in soil moisture regime along the horseshoe-shaped relevé configuration is from very dry (1) to wet (5). In (b), the grazing effect structures the scattergram (1-grazed, 2-ungrazed). In (c), soil organic matter, on a scale from 1 to 9, exhibits a weak trend. Graph (d) shows the dispersion of the relief position of the sample in the rolling landscape (1-flat top, 2-convex slope, 3-concave slope, 4-lowland). Scattergrams produced separately by SYNCSA (Section 6.6.9).

The main advantage of the character-based approach is the possibility of explaining community patterns on the basis of characters and CSTs. Table 7.1.2.1 displays species, CSTs defined on hierarchical level 8, and their performances in the relevés. The rows and columns in the table are arranged according to relevé (Figure 7.1.2.5) and CST groups (not shown) obtained by sum of squares cluster analysis using chord distances based on the CST performances. Note that, since the character description is local to the relevé, the same species may assume different CST forms, reflecting phenotypic variation in the sample. The dispersion of CST groups in the relevé groups is displayed also in Figure 7.1.2.6. Upon relating these trends to the environmental trends as seen in Figure

7.1.2.4, it is obvious that CST group 1 occurs mainly in the very dry and dry sites, CST group 2 in ungrazed and dry to mesic sites, and CST group 4 in moist and wet sites.

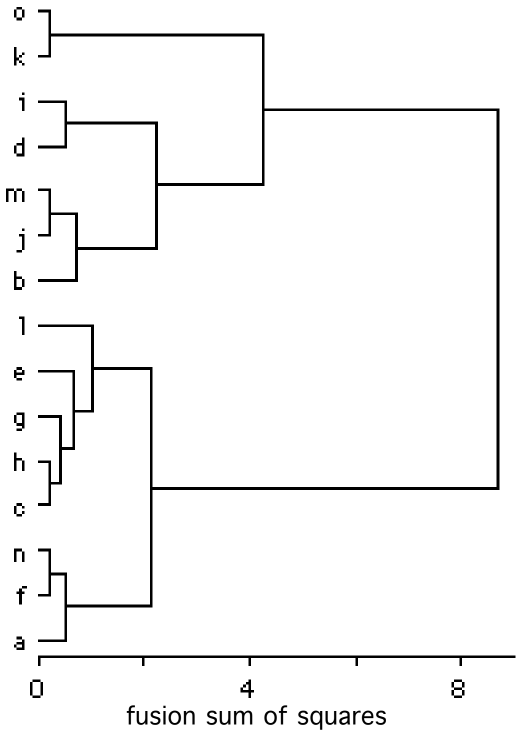


Figure 7.1.2.5. Cluster analysis of relevés for the EEA/UFRGS grassland data set, using the sum of squares method (see Section 6.6.10), based on chord distances computed from the CST cover-abundance values, hierarchical level 8. On this level the CSTs are defined by the character subset as shown in Figure 7.1.2.1. Analysis and dendrogram produced by SYNCSA.

The community structure is interpreted in plant characters terms in Figures 7.1.2.7 and 7.1.2.8. These are related to the environmental variation shown in Figure 7.1.2.4, which reveals, for instance, that the stoloniferous growth-form is characteristic (but not exclusive) in wet sites, the rhizomatous form in drier and mesic sites, and the rolled leaf cross section and caespitose growth-form in dry sites.

Table 7.1.2.1. Structured table on hierarchical level 8 for the EEA/UFRGS data set. On this level the CSTs are defined by the character subset shown. The groups of relevés and CSTs were obtained by cluster analysis. The species corresponding to the CSTs are indicated (abbreviations in Table 7.1.2.2). Note the concentration of CST group 1 in relevé group 1, CST group 2 in relevé group 2, and CST group 4 in relevé group 3. Table produced by SYNCSA (see Section 6.6.11).

							Relevés														
							a	c	e	f	g	h	l	n	b	d	i	j	m	k	o
gr	Species	CSTs					Relevé groups														
#		g3	cr	gl	st	wi	1	1	1	1	1	1	1	1	2	2	2	2	2	3	3
1	Anla	3	2	3	3	2	5	2	0	5	3	3	5	8	3	0	5	3	5	0	0
1	Pano	2	1	1	3	2	5	7	8	7	7	5	7	5	0	0	3	0	0	0	0
1	Pimo Arfi	3	3	3	3	1	5	3	3	3	3	6	0	5	2	0	1	2	2	0	0
1	Rusp	3	1	1	1	3	2	3	0	3	3	3	2	0	3	0	2	3	2	0	0
1	Sege Hepu Cose Erne	3	1	1	3	2	3	6	4	1	1	2	5	0	2	0	0	5	5	0	0
1	Rehi Fare Sopt Boer	3	1	1	1	1	2	2	2	1	3	0	2	2	1	1	2	1	3	0	0
1	Erho	2	1	2	3	2	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0
1	Hepu	3	1	1	3	3	0	2	0	0	1	2	0	0	1	0	0	1	0	0	0
1	Spin	3	2	3	3	1	0	5	2	2	3	3	0	0	0	0	0	3	0	0	0
1	Rihu	2	1	1	1	3	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0
1	Dein Batr	3	1	1	2	4	0	3	0	0	3	3	4	0	0	0	0	3	3	1	0
1	Rusp Asmo	3	1	1	1	4	0	0	3	0	2	2	0	2	0	5	0	0	0	0	0
1	Sege Cose Erne	3	1	3	3	2	0	0	3	6	2	3	0	3	0	0	0	0	1	0	0
2	Axaf	1	2	1	3	3	5	0	0	0	0	0	0	0	5	2	0	0	0	5	0
2	Pano	2	1	1	3	3	0	0	0	0	0	0	0	0	8	0	0	7	2	0	0
2	Cebi	1	1	1	1	4	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
2	Erho	2	1	2	3	3	0	0	0	0	0	0	0	0	0	7	7	7	3	0	0
2	Bame	3	1	1	2	1	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0
2	Arfi	3	3	1	3	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
3	Rihu	2	1	1	1	2	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
3	Spin	3	2	1	3	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	Cebi	3	1	2	1	4	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
4	Pano Sysp Papu	3	1	2	3	3	0	1	1	0	2	1	1	0	2	0	0	0	0	5	5
4	Trmo Elgl	3	1	3	3	1	0	0	0	0	0	0	0	0	0	5	0	0	0	5	7
4	Anla Axaf	3	2	1	3	2	0	0	0	1	5	0	0	0	0	0	0	0	0	5	3
4	Axaf	3	2	1	3	3	0	0	0	0	0	2	2	0	0	0	0	0	0	0	5
4	Cebi	1	1	2	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	5	7

Table 7.1.2.2. List of species described in the EEA/UFRGS data set. Abbreviations are indicated.

#	Species abbreviation and name	
1	Anla	Andropogon lateralis
2	Arfi	Aristida filifolia
3	Asmo	Aspilia montevidensis
4	Axaf	Axonopus affinis
5	Bame	Baccharis megapotamica
6	Batr	Baccharis trimera
7	Boer	Borreria eryngioides
8	Cebi	Centella biflora
9	Cose	Coelorachis selloana
10	Dein	Desmodium incanum
11	Elgl	Eleocharis glauco-virens
12	Erho	Eryngium horridum
13	Erne	Eragrostis neesii
14	Fare	Facelis retusa
15	Hepu	Herbertia pulchella
16	Pano	Paspalum notatum
17	Papu	Paspalum pumilum
18	Pimo	Piptochaetium montevidense
19	Rehi	Relbunium hirtum
20	Rihu	Richardia humistrata
21	Rusp	Ruellia sp.
22	Sege	Setaria geniculata
23	Sopt	Soliva pterosperma
24	Spin	Sporobolus indicus
25	Sysp	Sysyrinchium sp.
26	Trmo	Trachypogon montufari

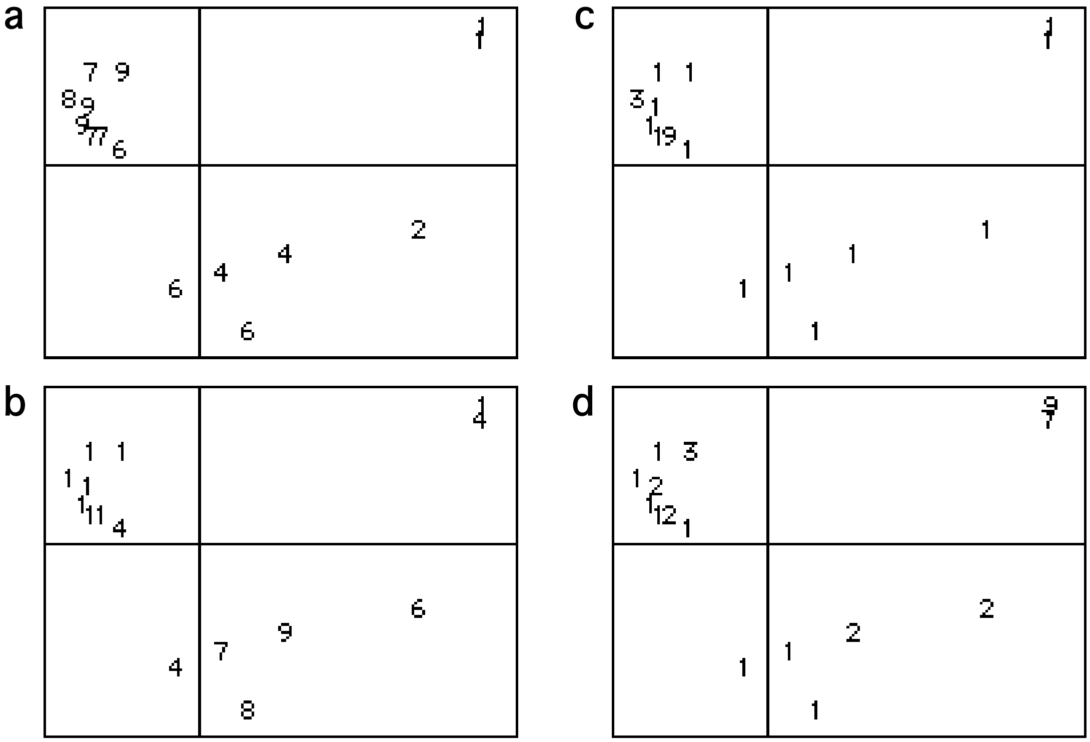


Figure 7.1.2.6. Dispersion of the CST groups (Table 7.1.2.1) within the eigenordination scattergram (Fig. 7.1.2.2 a). Relevé labels identify the performance total of CST groups 1 (a), 2 (b), 3 (c) and 4 (d) in the relevé, scaled from 1 (low) to 9 (high). Note the large cover-abundance of CST group 1 in the relevé group 1 (a), CST group 2 in relevé group 2 (b), CST group 3 in relevé group 1 (c), and CST group 4 in relevé group 3 (d). Scattergrams produced separately by SYNCSA (see Section 6.6.9)

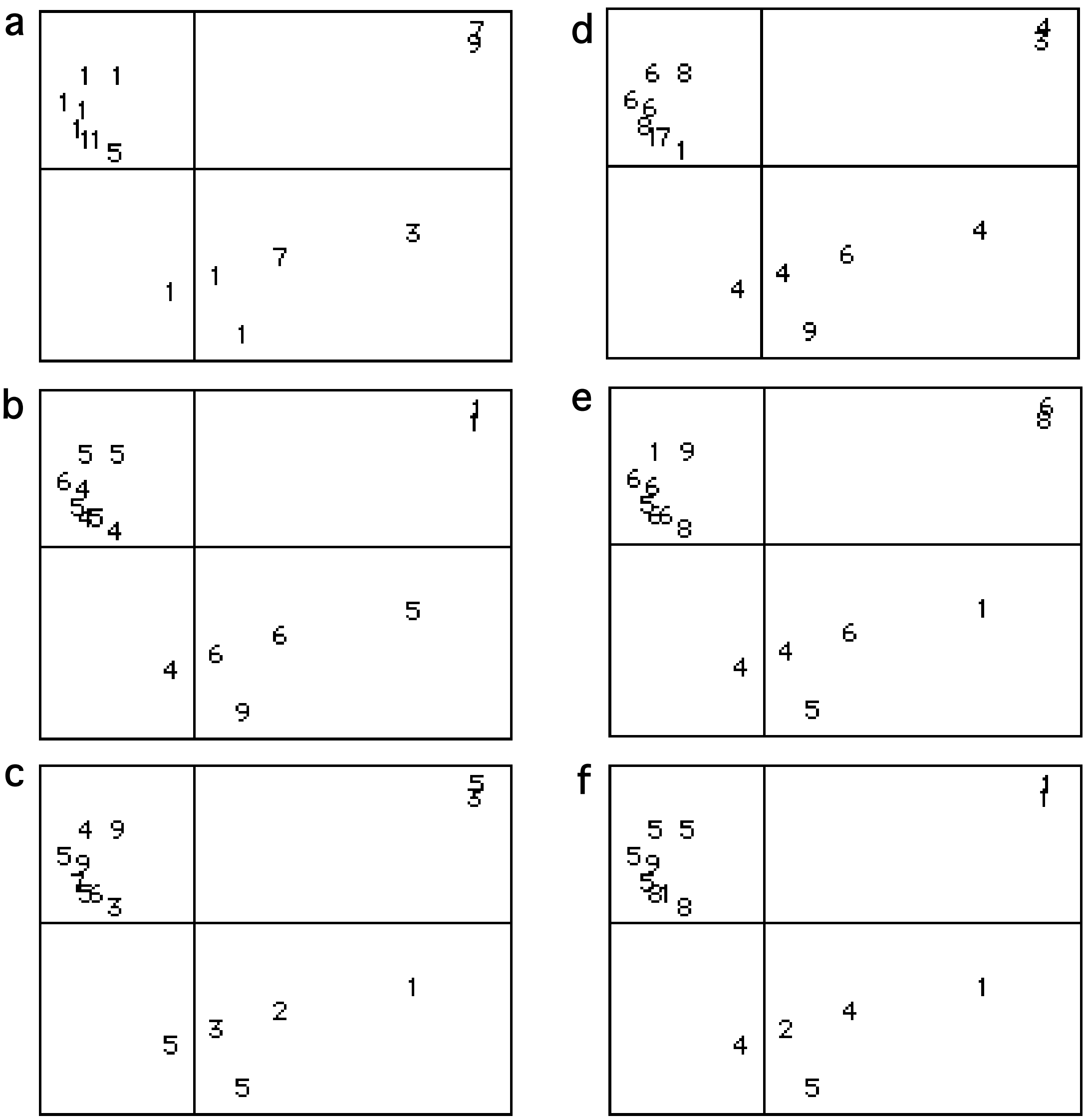


Figure 7.1.2.7. Dispersion of the states of growth-form (a:stoloniferous, b:rhizomatous, c:else; Appendix A) and leaf cross section (d:straight, e:folded, f:rolled) within the eigenordination scattergram of Fig. 7.1.2.2 (a). Relevé labels identify the performance total of the CSTs that present the character state in the relevé, scaled from 1-9. Note in graph (a) the larger cover-abundance of CSTs with stoloniferous growth-form in relevé group 3, which in turn, as shown in (b), has very low cover-abundance of CSTs with rhizomatous growth-form. Also, note in (f) the concentration of CSTs with rolled leaf cross section in relevé group 1. Each graph is produced separately by SYNCSA (see Section 6.6.9).

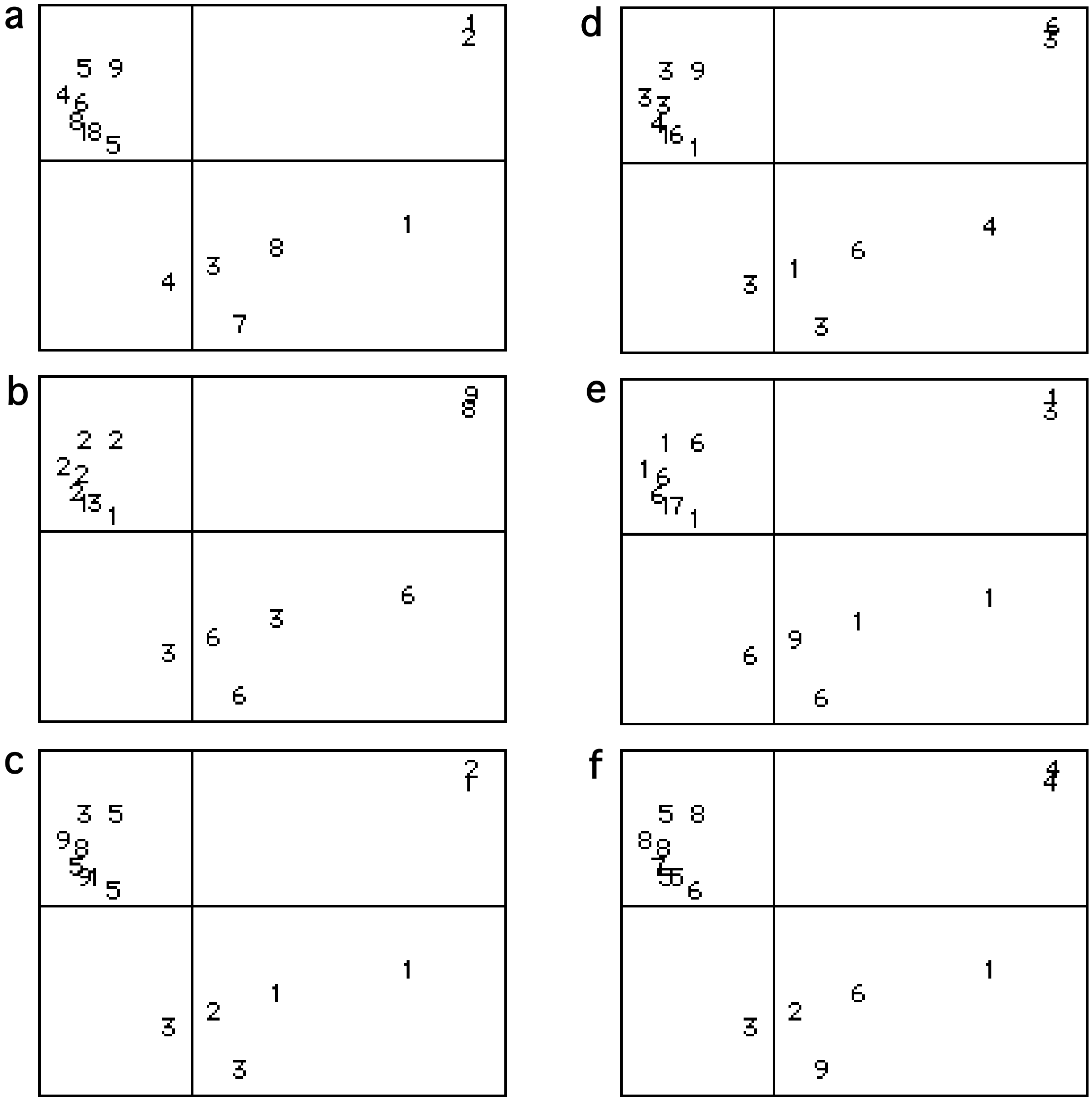


Figure 7.1.2.8. Dispersion of the states of growth-form (a:solitary, b:rosette, c:caespitose; Appendix A) and stem issue type (d:herbaceous, e:woody, f:no stem) within the eigenordination scattergram of Fig. 7.1.2.2 (a). The relevé labels are explained in Fig. 7.1.2.7. Note the concentration of CSTs with rosette growth-form in relevé groups 2 and 3 in (b), and CSTs with caespitose growth-form and no stem in relevé group 1 in (c, f).

7.2. Joint analysis of floristically disjunct communities: Caatinga and Chaco/Monte

The fact that climatically similar regions of the world also present similar vegetation structure has long been noted (Grisebach 1872, Schimper 1898, Raunkiaer 1908), particularly in Mediterranean type climates (Naveh 1967, Specht 1969). The convergent community evolution hypothesis can be assessed by structural comparison of communities. We examined several studies of this kind (see *e.g.*, Mooney and Dunn 1970, Mooney *et al.* 1970, Sarmiento 1972, Barbour and Diaz 1973, Parsons 1976, Orians and Solbrig 1977a, Bucher 1982), but, with the exception of Parsons', We found them lacking an adequate quantitative approach. By definition the communities being compared contain populations of different genetic stock, geographically distant, but existing under similar environmental conditions. However, the assumption that the present and, to some extent, past environmental conditions are similar is very difficult to fulfill, even when efforts are made to match locations as much as possible, such as in Orians and Solbrig (1977a). Also, even if sites presenting matching environmental conditions and communities are found, the degree of convergence per se cannot tell whether the convergent characters are adaptive or not (see Section 1.4), *i.e.*, whether they are evolutionary responses linked to the survival of the organisms under these specified conditions.

Having the character-based approach and the optimization algorithms (Chapter 4) as tools, the problem can be redefined quantitatively. Other than a simple measure of convergence between communities, the interest may also lie in finding environmental explanations of character-based community structures. If satisfactory explanations are found, the ecologist could for instance work with the hypothesis that these explanations may also be valid, in a global scale, for other community data sets, and in different floristic regions. Also, these explanations could guide the building of models to predict vegetation response, defined in CST composition terms, to spatial and temporal environmental changes. This analytical strategy requires several community descriptions from a broad environmental variation.

In the example, character-based relevés from the Caatinga, Chaco and Monte formations in South America are jointly analyzed using the application SYNCSA. The study reveals trended variation in community structures, which are explained in environmental terms. Also, the convergence between Caatinga and Chaco communities is evaluated.

7.2.1. Data sets

The Caatinga data set contains 9 relevés in the region of Sobral, state of Ceará, between latitudes 3° 20' and 4° S. The Chaco and Monte relevés are located between latitudes 27° and 28° S in the provinces of Santiago del Estero and Catamarca, Argentina, with 9 relevés in the driest part of the Chaco, and 10 relevés in the Monte. The area sampled in the Monte is in the Salar de Pipanaco, to the west from the Subandean ranges, in Catamarca, which was the same area studied by Barbour and Diaz (1973) and Orians and Solbrig (1977a). The sampling avoided stands in initial secondary successional stages, specially in the very disturbed Caatinga, and the unit boundaries were defined with a view at vegetation homogeneity. All woody species were recorded and described using the defining character set given in Table 7.2.1.1. Information on annual precipitation, altitude and soil texture was recorded. The complete data set is found in Pillar (1992).

Table 7.2.1.1. Character set used in the description of the Caatinga, Chaco and Monte vegetation.

Biological type

1. Type (bt) 1:bryoid, 2:lichen, 3:pteridophyte, 4:conifer, 5:graminoid, 6:cactoid, 7:other

Stem (stem-like structure)

2. Tissue type (st) 1:succulent, 2:herbaceous, 3:woody, 4:plant with no stem
3. Function (fu) 1:regular, 2:twin-purpose, 3:plant with no stem
4. Armature type (at) 1:thorn/spine, other vestures, 2:none, 3: plant with no stem
5. Growth form (gf) (states from key in Pillar 1992)

Leaf (leaflet, leaf-like structure)

6. Duration (du) 1:aseasonal deciduous, 2:seasonal deciduous, 3:withering, 4:persistent, 5:plant leafless
7. Tissue type (lt) 1:succulent, 0: else
8. Texture (tx) 1 to 5: scale from herbaceous to fibroid/leathery; 6:else
9. Shape (sh) 1:scale, 2:filiform/needle, 3:other, 4:plant leafless
10. Arrangement (ar) 1:simple, 2:compound, 3:plant leafless
11. Epidermal surface dorsal (ed) 1:glabrous, 2:glaucous, 3:trichomous sparse, 4:trichomous dense, 5:plant leafless
12. Epidermal surface ventral (ev) 1:glabrous, 2:glaucous, 3:trichomous sparse, 4:trichomous dense 5:plant leafless
13. Width (wi) 1:< 2.5 mm, 2:2.5-5, 3:5-10, 4:10-50, 5:50-100, 6:100<, 0:plant leafless
14. Length (le) 1:< 5 mm, 2:5-25, 3:25-75, 4:75-125, 5:125<, 0: plant leafless
15. Thickness (th) 1:< 1 mm, 2:1-3, 3:3-5, 4:5<, 0:plant leafless

Plant height

16. Height class (hc) 1: <5cm, 2:5-25, 3:25-75, 4:75-125, 5:125-250, 6: 250-500, 7:500-1000, 8:>1000

The Caatinga vegetation covers the semi-arid region in Northeastern Brazil between latitudes 3° and 15° S. Related descriptive studies are found in Egler (1951), Numata (1970), Hayashi and Numata (1976), Gomes (1979), Eiten (1982) and Queiroz (1985). Information on the Caatinga flora is given by Andrade-Lima (1954), Braga (1960), Kirmse *et al.* (1983) and Prado (1991). The climax vegetation is a thorn scrub, deciduous rain-green, open woodland composed of low and medium sized trees and thorny shrubs (Andrade-Lima 1954:11, Hueck 1966:276, Hueck and Seibert 1972). The climate is characterized by very distinct wet (south hemisphere summer) and dry (winter) seasons. The annual precipitation is less than 750 mm over most of the region. In addition, there is an extreme variability of precipitation between years, with cyclic floods and droughts. Because of orography, the climate changes in short distances along altitudinal gradients. Hills exposed to moisture bearing winds (mainly from the southeast) are covered by forest in contrast to the low, sparse vegetation of the surrounding Caatinga (Andrade-Lima 1982). Seasonal temperature variation is minimal. The monthly mean of maximum and minimum temperatures range from 31-37 °C and 20-24 °C respectively (Queiroz 1985:28). Since the region is relatively densely populated by small farmers, the vegetation is under frequent disturbance by grazing (goats, cattle) and wood cutting for fuel and fencing (Queiroz 1985:8).

The Chaco vegetation covers the Quaternary plain located between the Paraguay/Paraná Rivers in the east and the foothills of the Pampean and Subandean ranges in the west. Physiognomic and floristic descriptions are found in Cabrera (1971) and Prado (1991), and climate and soil descriptions in Burgos (1963) and Soriano and Prego (1963). The climate in the Chaco is continental, with mean annual temperature 20 to 23 °C, and annual precipitation, concentrated in the south hemisphere summer, decreasing from east (1200mm) to west (450 mm). In the west the dry season lasts 6-7 months, and the vegetation is xerophilous subtropical forest (Cabrera 1971, Sarmiento 1972). The climax community is a forest of *Schinopsis quebracho-colorado* and *Aspidosperma quebracho-blanco* (Cabrera 1971). The Monte is an arid formation that extends through western Argentina, between latitudes of 38° to 44° S (Cabrera 1971). The annual precipitation ranges from 80 to 250 mm. Floristically, the region is characterized by species of *Larrea* and *Prosopis*. The Chaco and Monte regions are sparsely populated.

Some studies have pointed out structural similarities between Caatinga and Chaco vegetation (Hueck 1966:278, Bucher 1982). Also, there are some common animal species (Bucher 1982). However, Prado (1991) reviewed the floras and found almost no floristic links.

7.2.2. The vegetation structure in the Chaco/Monte and its connections with the Caatinga

The analysis uses the optimal character order for environmental congruence in the Chaco/Monte data (Fig. 7.2.2.1) as a framework for the comparison with the Caatinga. Environmental structure is defined by annual precipitation, altitude, and soil texture. It is noted that the variation in the Chaco/Monte vegetation structure described by eigenordination is closely related to the variation in precipitation (Fig. 7.2.2.2).

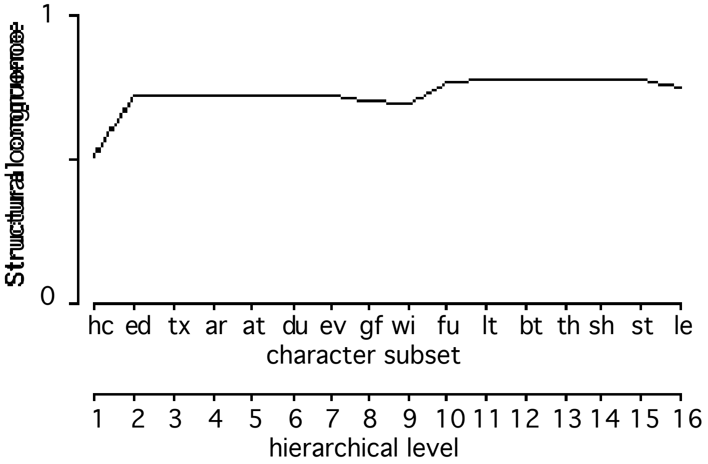


Figure 7.2.2.1. Congruence $\rho(\mathbf{D}_i;\Delta)$ between vegetation and environmental data structures for the Chaco/Monte sample (19 relevés). Character order is optimal (Section 4.3.1). Vegetation data structure is defined by the squared chord distance matrix ($\mathbf{D}_i$) between relevés on each level i. Environmental data structure (Δ) is defined as explained before. The defining character subset on each level is cumulative from the top down in the hierarchy, *e.g.*, only character le on level 16 and the complete character set on level 1. The character labels are identified in Table 7.2.1.1.

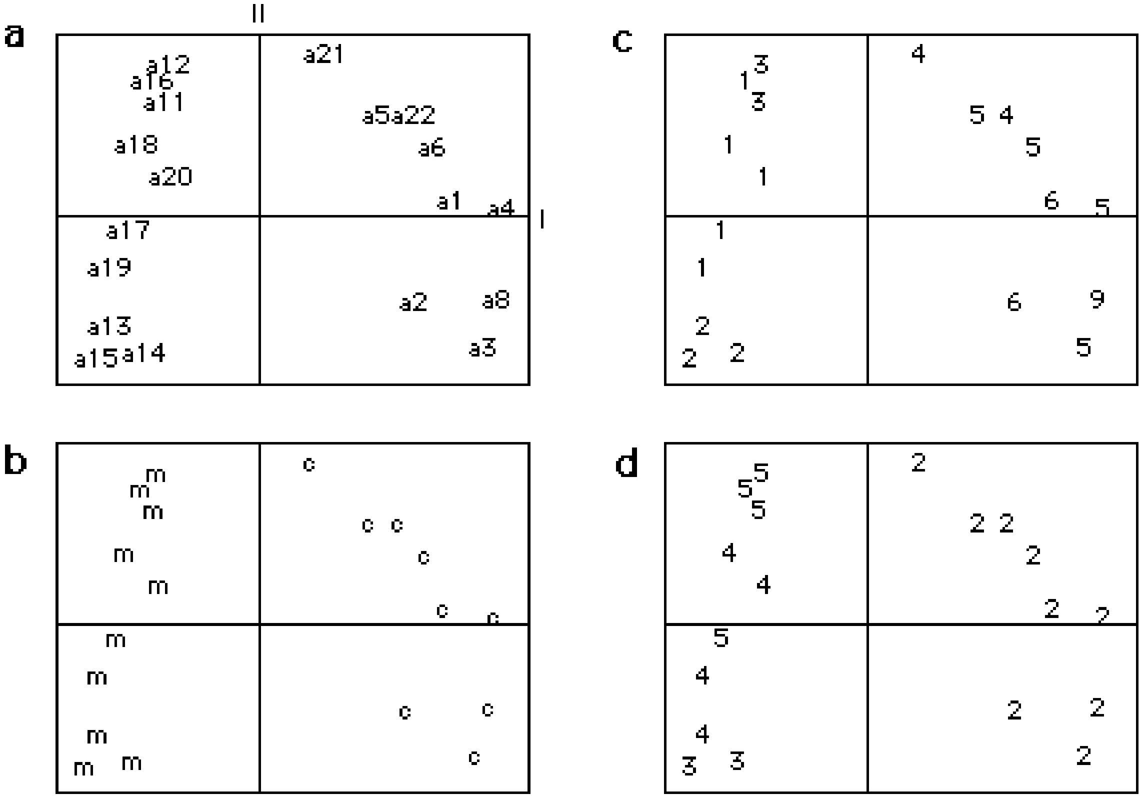


Figure 7.2.2.2. Eigenordination of the Chaco and Monte data set, hierarchical level 11, based on squared chord distances of relevés, using the optimal character order for congruence with the environmental structure (Fig. 7.2.2.1). Eigenaxes 1 and 2 are plotted, representing respectively 71.33% and 11.75% of the total. The labels identify the relevés. In (b) c is Chaco and m is Monte. The variation in annual precipitation within the ordination space is scaled from 1-9 in (c). Soil texture is coded by 2:silt, 3:sand, 4:gravel, 5:rock in (d). Relevés a21 and a22 are in the driest part of the Chaco in a transition with the Monte.

The Chaco/Monte data set is expanded by including the Caatinga sample. The ordination of the expanded data set, using the same character order mentioned above, is shown in Figure 7.2.2.3. It is remarkable that a similar pattern of variation matching the precipitation variation as in Fig. 7.2.2.2 (c) can be observed here (Fig. 7.2.2.3b). This may indicate that the defining characters at hierarchical level 11, or other characters closely related, present a coherent response to precipitation in the two data sets. A good question at this point is whether the response will remain the same if additional data sets from other environmentally similar regions are added. If so, we will find an optimal character set for modelling semi-arid vegetation response to precipitation. We should be aware, however, that annual precipitation is not an ideal index to reflect actual hydric conditions, since it does not take account of evapotranspiration and soil conditions.

Figure 7.2.2.4 displays the dispersion of states of leaf length, which is the character with maximum environmental congruence. Leaf size (indicated by leaf length) tends to increase from the Monte to the Chaco, and to the Caatinga.

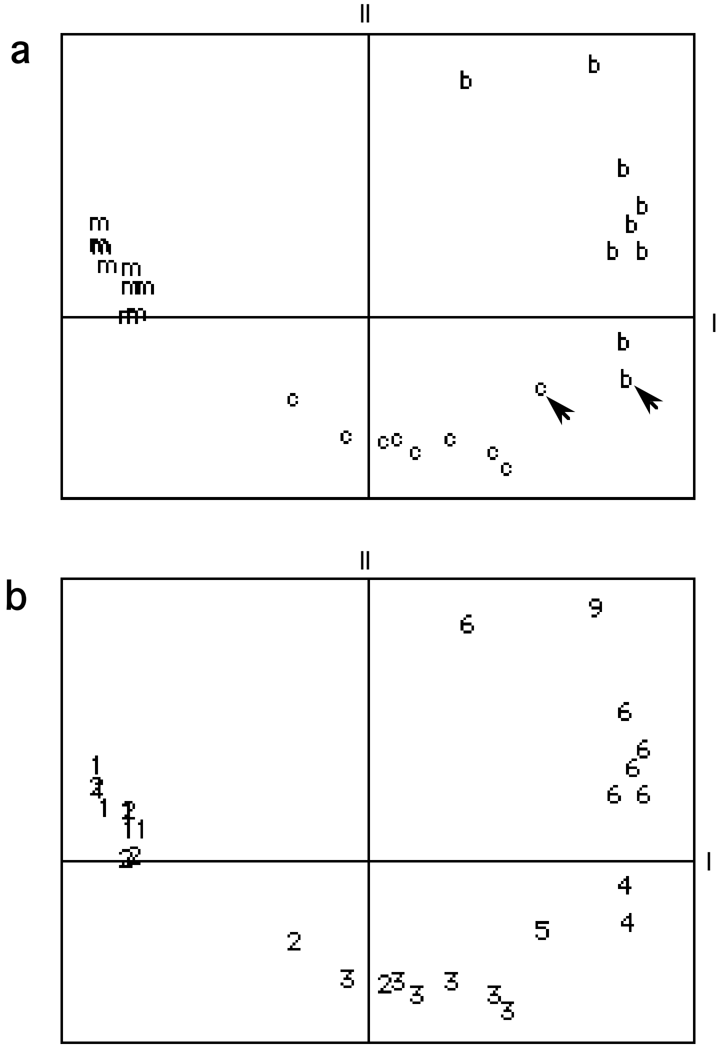


Figure 7.2.2.3. Eigenordination of the Caatinga/Chaco/Monte data set, hierarchical level 11 based on squared chord distances of relevés, using the optimal character order for congruence with the environmental structure (Fig. 7.2.2.1). Eigenaxes 1 and 2 are plotted, which represent 58.09% and 19.23% of the total. The labels identify the relevés. In (a) the symbol b is Caatinga, c is Chaco, and m is Monte. Variation in annual precipitation (graph b) is scaled from 1-9. Note the trended variation in composition from the driest (Monte) to the moistest sites in the Caatinga. The arrows point to relevés b9 in the Caatinga and a8 in the Chaco that are shown in Figure 7.2.3.2.

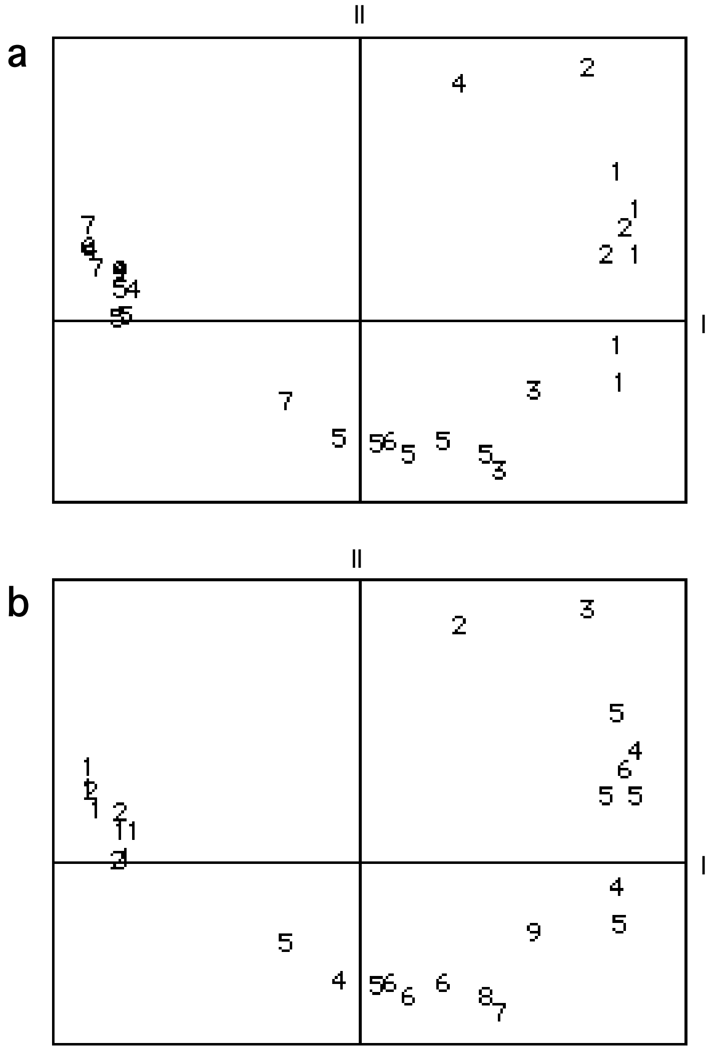


Figure 7.2.2.4. Dispersion of two states of leaf length within the eigenordination scattergram (Fig. 7.2.2.3a). State 2 (5-25mm) is shown in graph (a) and state 3 (25-75mm) in graph (b). The numbers locate the relevés and identify the performance total of the CSTs that contain the character state in the relevé, scaled from 1-9. Note the commonness of shorter leaf in the Monte and longer leaf in the Caatinga. The Chaco is intermediate.

7.2.3. Assessing the convergence of Caatinga and Chaco

Community convergence should not be treated separately from environmental congruence, in the sense that if there is any convergence revealed by using a given character set, the same set should also show strong congruence with the variation in relevant environmental factors. In other words, one needs a two-step procedure before one can draw conclusions about ecological evolutionary convergence: 1. Find an optimal character order for maximum environmental relevance in one or both samples. 2. Evaluate the convergence between the samples using the characters as ordered in (1). These two steps can be reversed.

The Caatinga and Chaco samples do not share species, therefore any structural similarity can be described as convergence, and the correction factor for overlapping species sets (Orlóci *et al.* 1986) is not required. Between samples (9 relevés in the Caatinga and 10 in the Chaco), no significant structural convergence is expressed on any hierarchical level by the taxonomy defined by the characters as ordered for optimal environmental congruence in the Chaco/Monte (see Fig. 7.2.3.1). The moist relevés in the Chaco are the closest to the driest relevés in the Caatinga, but the two groups are clearly distinguished (Fig. 7.2.2.3a). If moister sites in the Chaco were sampled, located more to the east from where the survey was conducted, it is expected that a stronger convergence with the Caatinga sample would have been found.

The lack of convergence between the Caatinga and Chaco sample, with the characters ordered as in Fig. 7.2.3.1, does not preclude the possibility of measuring convergence between specific pairs of relevés, such as relevés b9 in the Caatinga and a8 in the Chaco. In this case, as shown in Figure 7.2.3.2, the structural convergence is significant at an $1–\alpha$ probability of 0.074 on hierarchical level 10. The composition of the relevés in CST and species terms is given in Tables 7.2.3.1 and 7.2.3.2.

Up to this point, the character order is optimal for environmental congruence in the Chaco/Monte data. Now, an optimal character order for convergence between the Caatinga and Chaco samples is found applying the method already explained (Section 4.3.2), and then the local environmental congruence in each sample, and in the Chaco/Monte, is evaluated (Figures 7.2.3.3 and 7.2.3.4). Note that in these terms the structural convergence is significant when using a subset of 7 characters (lt bt th st sh fu ar). This subset coincides with the optimal subset in the previous analysis (Table 7.2.3.1), except for leaf arrangement (ar) and leaf length (le). The environmental congruence, when the optimal character order for convergence between Chaco and Caatinga samples is used, is very low in the Caatinga, and higher in the Chaco and Chaco/Monte samples. This may be explained by the incompleteness of the environmental information in the highly disturbed Caatinga.

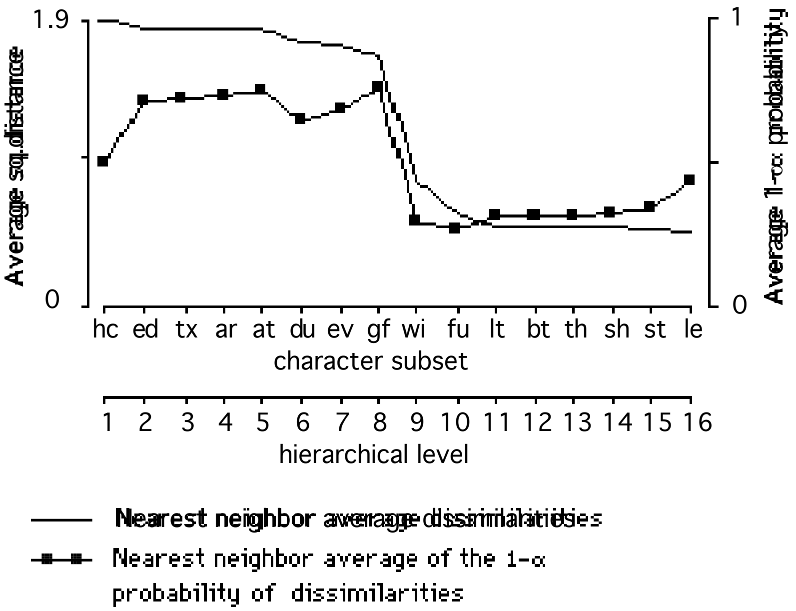


Figure 7.2.3.1. Nearest neighbor dissimilarities (Eq. 4.3) between the Caatinga (9 relevés) and the Chaco (10 relevés) samples (smooth line), and the nearest neighbor average 1–α probabilities of the dissimilarities (connected full squares). Character order is optimal for maximum environmental congruence in the Chaco/Monte data set (Fig. 7.2.2.1). Data structure is defined by the squared chord distance matrix ($\mathbf{D}_i$) between relevés on each level i. The defining character subset on each level is cumulative from the top down in the hierarchy. The character labels are identified in Table 7.2.1.1. The 1–α probabilities of the dissimilarities are generated by randomization carried to 1000 iterations under the random composition hypothesis (Section 2.8). The 1–α probabilities (not shown) of the structural evaluation function itself were all above 0.99. Note the absence of a significant structural convergence.

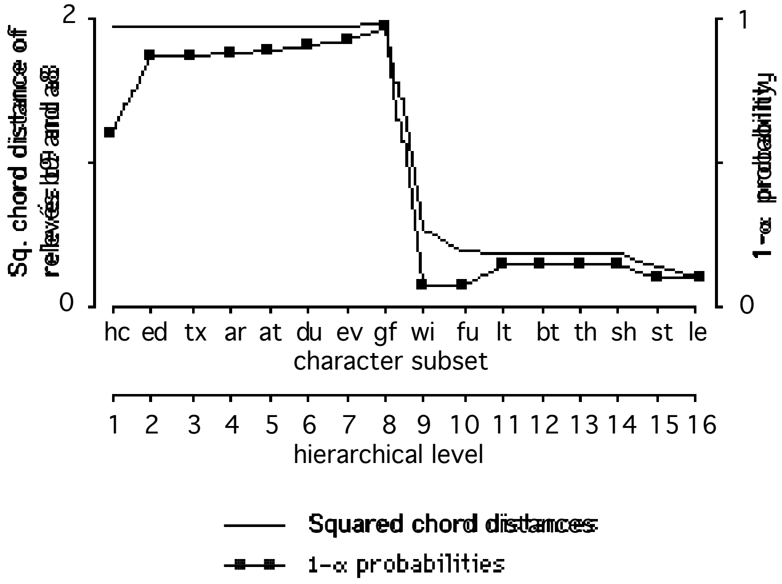


Figure 7.2.3.2. Profiles of the squared chord distances between relevés b9 (Caatinga) and a8 (Chaco) and the respective 1–α probabilities. Character order is optimal for maximum environmental congruence in the Chaco/Monte data set (Fig. 7.2.2.1). The character labels are identified in Table 7.2.1.1. The 1–α probabilities are generated by randomization carried to 1000 iterations under the random composition hypothesis (Section 2.8). Note the stronger convergence than in Fig. 7.2.3.1, since the 1–α probabilities are as low as 0.074 on hierarchical level 10.

Table 7.2.3.1. CST composition of relevés b9 (Caatinga) and a8 (Chaco), on hierarchical level 10. On this level the CSTs are defined by the character subset shown (symbols in Table 7.2.1.1). The species corresponding to the CSTs are indicated (abbreviations in Table 7.2.3.2). Note the matching species. Underlined names are from the Caatinga.

Species	Characters							Relevés	
	le	st	sh	th	bt	lt	fu	b9	a8
Mica Aspy Capy Asqu Cepa Zimi Scqu Acpr	3	3	3	1	7	0	1	12	20
Cole Bach	5	3	3	1	7	0	1	6	0
Miac Acar Prsp Pomi	1	3	3	1	7	0	1	3	10
Ceja Ceva Opqu	0	1	4	0	6	0	2	2	4
Scfa	2	3	3	1	7	0	2	0	3
Trca	5	3	2	1	7	0	1	0	3
Jorh	3	3	3	1	7	0	2	0	2
Trpr unsh Cepa	4	3	3	1	7	0	2	0	7
Clsp unli	3	2	3	1	7	0	2	0	5
Bali	2	3	3	1	7	0	1	0	3

Table 7.2.3.2. Species encountered in relevés b9 (Caatinga) and a8 (Chaco). Note the richer flora in the Chaco relevé. (Authority on species names may be found in Prado 1991).

Abbr.	Species name	Abbr.	Species name
CAATINGA:		CHACO:	
Mica	Mimosa caesalpiniaefolia	Asqu	Aspidosperma quebracho-blanco
Miac	Mimosa acutistipula	Acar	Acacia aroma
Ceja	Cereus jamacuru	Cepa	Celtis pallida
Cole	Combretum leprosum	Zimi	Ziziphus mistol
Bach	Bauhinia cheilanta	Ceva	Cereus validus
Aspy	Aspidosperma pyrifolium	Scqu	Schinopsis quebracho-colorado
Capy	Caesalpinia pyramidalis	Prsp	Prosopis sp.
		Pomi	Porlieria microphylla
		Bali	Baccharis like
		Trca	Trithrinax campestris
		Jorh	Jodina rhombifolia
		Trpr	Trixis praestans
		unsh	unidentified shrub
		Clsp	Clematis sp.
		Opqu	Opuntia quimilo
		Acpr	Achatocarpus praecox
		Cepa	Cestrum parqui
		unli	unidentified liana (milky)
		Scfa	Schinus fasciculatus

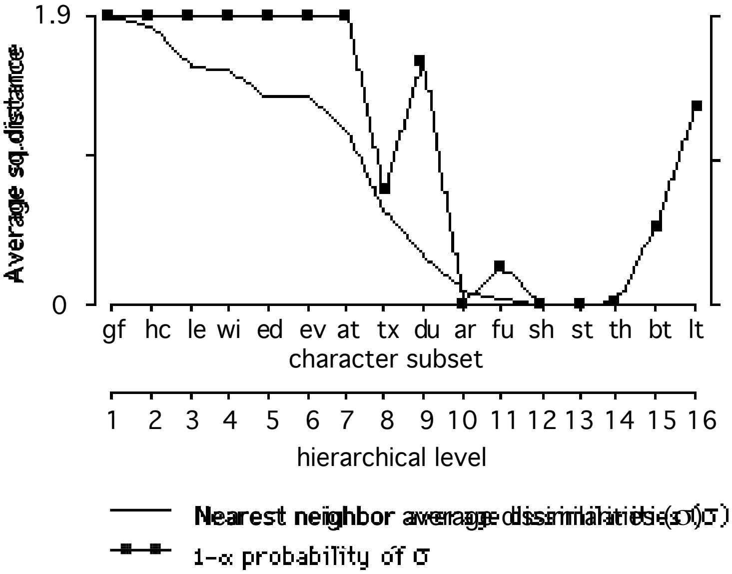


Figure 7.2.3.3. Nearest neighbor dissimilarities (Eq. 4.3) between the Caatinga and the Chaco samples (smooth line), and the corresponding 1–α probabilities (connected full squares). Character order is optimal for maximum structural convergence between the Caatinga and Chaco. Data structure is defined by the squared chord distance matrix between relevés on each level i. The defining character subset on each level is cumulative from the top down in the hierarchy. The character labels are identified in Table 7.2.1.1. The 1–α probabilities are generated by randomization carried to 1000 iterations under the random composition hypothesis (Section 2.8). Structural convergence is significant on hierarchical level 10 (1–α = 0.002), in which case the character subset is lt bt th st sh fu ar.

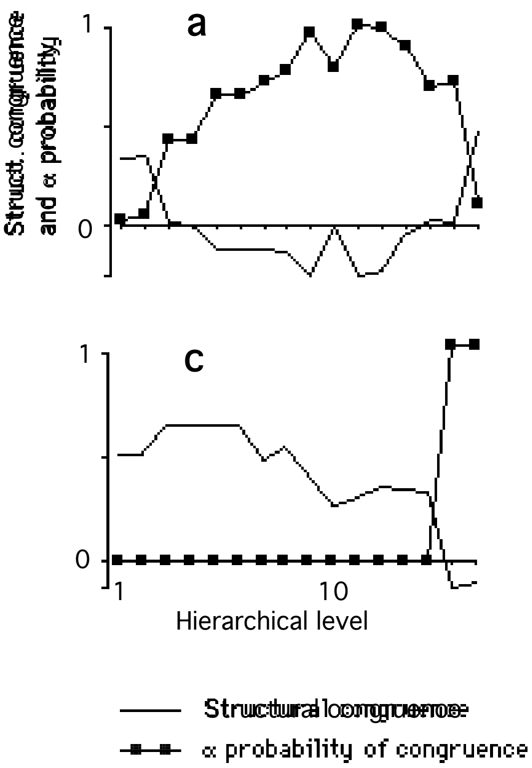


Figure 7.2.3.4. Congruence with environmental structure in the Caatinga (a), Chaco (b) and Chaco/Monte (c) using the optimal character order in Fig. 7.2.3.3, and the corresponding α probabilities. The probabilities (α) were generated by randomization carried on 1000 iterations under the random composition null hypothesis. In (b) the probability on hierarchical level 10 is 0.106. In (c) the probabilities are 0.001 in all levels, except levels 15 and 16 where α is larger than 0.97.

APPENDICES

Appendix A: Contents of data file "ufrgs15re12ch.dat". Note the reminder at the end of the file. Because SYNCSA disregards any comments placed after the data of interest, that is a good place for writing the information the user will need for keyboard input when the session's formatted data file is created from scratch.

```
3 3 2 3 3 2 3 1 2 3 2 3 2
3 1 1 2 3 1 1 3 3 3 2 1 2
3 3 2 3 3 3 1 1 3 3 1 1 1
3 1 1 3 1 1 3 1 1 3 3 1 +
3 3 2 3 3 3 1 1 1 3 1 2 +
5 1 2 3 3 1 1 1 2 3 2 3 1
3 1 1 1 3 2 3 1 3 3 3 1 2
3 1 2 3 1 1 1 1 1 3 1 1 +

3 3 1 3 3 2 3 1 2 3 2 2 1
3 3 2 3 3 3 1 1 3 3 1 2 +
3 1 1 1 3 2 3 1 3 3 3 2 2
3 1 1 2 3 1 2 1 3 3 3 2 4
3 1 1 1 1 1 1 1 1 3 4 1 1
4 2 2 2 3 1 4 1 1 2 2 1 r
3 1 1 3 1 1 3 1 1 3 3 2 1
5 1 2 3 1 1 1 4 1 3 1 1 r
4 1 2 3 3 1 1 1 1 3 3 2 r
3 1 2 3 3 1 1 1 1 3 2 2 +
3 2 1 3 3 1 1 1 1 3 3 1 +

3 3 2 3 3 2 3 1 1 3 1 2 2
3 1 1 2 3 1 1 3 3 3 2 1 3
5 1 2 3 1 1 1 4 1 3 1 1 r
3 2 1 3 3 1 1 1 1 3 3 1 r
3 1 1 2 1 1 2 3 3 3 3 1 r
4 1 2 3 3 1 1 1 1 3 3 2 +
3 3 2 3 3 3 1 1 3 3 1 2 1
5 1 2 3 1 1 1 3 3 3 1 1 r
3 1 1 3 1 1 3 1 1 3 3 1 1
3 3 1 3 3 2 3 1 2 3 2 2 +
3 1 2 3 3 1 1 3 3 3 2 1 +
3 1 1 3 2 1 3 1 1 3 4 1 1
3 1 2 3 3 1 1 1 1 3 2 1 +
5 1 2 3 3 1 1 1 2 3 2 1 +

3 2 2 2 3 1 4 1 1 2 3 4 3
3 1 1 3 1 1 1 3 3 1 4 3 2
3 1 2 3 1 1 1 1 1 3 1 4 r
3 3 2 3 3 1 3 1 1 1 1 4 2
3 1 2 1 3 2 3 1 3 3 3 4 +

3 1 1 2 3 1 1 3 3 3 2 1 4
3 3 2 3 3 3 1 1 3 3 1 1 1
```

```
5 1 2 3 1 1 1 4 1 3 1 1 r
5 1 2 3 1 1 1 3 3 3 1 1 r
3 2 1 3 3 1 1 1 1 3 3 1 r
3 1 1 3 1 1 1 3 3 1 4 1 1
3 1 2 3 3 1 1 1 1 3 2 1 1
3 3 2 3 3 2 3 1 1 3 1 3 +
5 3 2 3 3 1 1 1 2 3 2 2 1
3 1 2 3 3 1 1 3 3 3 2 1 r

3 1 1 2 3 1 1 1 3 3 2 1 3
5 3 1 3 3 1 1 1 2 3 2 1 1
3 3 1 3 3 2 1 1 2 3 2 1 2
3 1 1 2 1 1 2 3 3 3 2 1 r
5 1 2 3 1 1 1 4 1 3 1 1 r
3 1 1 3 1 1 3 1 1 3 3 1 1
3 3 1 3 3 3 1 1 1 3 1 1 1
3 1 1 3 3 2 3 1 3 3 2 1 r
4 2 2 2 3 1 4 1 1 2 2 1 +
4 1 2 3 3 1 1 1 1 3 2 1 r
3 3 2 3 3 2 3 1 1 3 1 3 +
3 3 2 3 3 1 1 1 1 3 2 1 1

3 1 2 2 3 1 1 1 3 3 2 1 3
3 1 1 3 3 2 3 1 3 3 2 2 2
3 2 2 3 3 1 1 1 1 3 3 2 +
3 3 2 3 3 1 1 3 3 3 2 1 +
3 1 2 3 3 1 1 1 1 3 2 1 r
3 3 2 3 3 2 3 1 2 3 2 2 1
3 3 2 3 3 2 3 1 1 3 1 3 1
3 1 1 3 1 1 3 1 1 3 3 1 1
3 1 2 3 1 1 1 1 1 3 1 2 +
3 1 2 3 1 1 1 1 1 3 1 1 r
3 3 2 3 3 3 1 1 3 3 1 2 1
3 1 1 2 1 1 2 3 3 3 3 1 r
4 1 2 3 3 1 1 1 1 3 3 3 r
3 1 1 3 1 1 1 3 3 1 4 1 +
3 1 1 3 2 1 3 1 1 3 4 1 1

3 1 1 3 1 1 1 3 3 1 4 1 +
3 3 1 3 3 2 3 1 2 3 2 2 1
3 1 1 2 3 1 1 3 1 3 2 1 2
5 3 1 3 3 1 1 1 2 3 2 1 +
4 1 2 3 3 1 1 1 1 3 3 2 +
3 3 3 3 3 3 1 1 3 3 1 2 1
3 1 1 3 1 1 3 1 1 3 3 1 1
3 2 1 3 3 1 1 1 1 3 3 1 r
3 3 2 3 3 1 1 3 3 3 2 1 r
3 1 1 3 3 2 3 1 3 3 3 1 +
3 3 2 3 3 2 3 1 1 3 1 3 1
3 1 2 3 3 1 1 1 1 3 2 2 +
3 3 1 3 3 3 1 1 1 3 1 1 1
3 1 1 3 2 1 3 1 1 3 4 1 1

4 2 2 2 3 1 4 1 1 2 3 5 3
3 3 2 3 3 2 3 1 2 3 2 5 2
3 3 2 3 3 3 1 1 1 3 1 2 r
3 1 2 3 1 1 3 1 1 3 3 4 +
2 1 2 3 2 1 2 1 1 3 1 5 2
3 1 2 2 3 1 1 1 3 3 2 3 1
3 1 2 3 1 1 1 1 1 3 1 4 +
```

```
  4 2 2 2 3 1 4 1 1 2 3 4 3
  3 1 1 2 3 1 2 1 3 3 3 1 3
  3 3 1 3 3 2 3 1 2 3 2 1 1
  3 1 3 3 3 1 1 1 1 3 2 1 2
  3 3 3 3 3 2 3 1 1 3 1 2 1
  3 1 1 2 1 1 2 3 3 3 3 1 r
  3 1 1 3 1 1 3 1 1 3 3 1 1
  4 1 2 3 3 1 1 1 1 3 3 2 r
  3 1 1 3 2 1 3 1 1 3 4 1 1
  3 1 2 3 1 1 1 1 1 3 1 1 r
  3 3 3 3 3 3 1 1 3 3 1 2 +
  3 1 1 3 3 3 1 1 1 3 1 1 r

  3 2 1 3 3 1 2 1 1 3 3 1 2
  3 2 1 1 1 1 1 1 1 3 4 1 2
  3 1 1 1 3 2 3 1 3 3 3 1 2
  3 3 2 3 3 1 3 1 1 3 1 2 2
  3 1 1 3 3 2 3 1 2 3 2 1 2
  2 1 2 3 2 1 1 1 1 3 4 1 r

  3 3 2 3 3 2 3 1 2 3 2 3 2
  3 1 1 2 3 1 1 3 3 3 2 1 3
  3 1 1 3 3 2 3 1 3 3 3 2 +
  3 1 2 3 1 1 1 1 1 3 1 1 +
  3 2 1 3 3 1 1 1 1 3 3 1 r
  3 1 1 3 3 1 1 1 1 3 2 1 2
  3 1 1 3 3 2 3 1 1 3 1 2 r
  3 2 1 3 1 1 2 3 3 3 4 2 +
  2 1 2 3 2 1 1 1 1 3 4 2 r
  3 1 1 3 2 1 3 1 1 3 4 1 1
  3 1 1 3 1 1 3 1 1 3 3 1 +
  3 1 1 2 1 1 2 3 3 3 2 1 r

  3 3 3 3 3 2 3 1 2 1 2 3 2
  3 1 2 3 3 1 1 1 1 1 2 3 2
  3 1 2 2 3 1 2 1 3 3 3 1 +
  4 2 2 2 3 1 4 1 1 2 3 3 1
  3 1 1 3 2 1 3 1 1 3 4 1 1
  3 1 1 3 1 1 3 1 1 3 3 1 +
  3 1 2 3 1 1 1 1 1 3 1 3 +
  5 3 1 3 3 1 1 1 2 3 2 2 r
  3 3 2 3 3 3 1 1 1 3 1 3 +
  5 1 2 3 1 1 1 4 1 3 1 1 r

  3 3 2 3 3 2 2 1 2 3 2 2 4
  3 1 1 2 3 1 1 3 1 3 2 1 2
  3 1 1 3 1 1 3 1 1 3 4 1 +
  3 3 2 3 3 3 1 1 3 3 1 1 +
  3 1 2 3 1 1 1 1 1 3 1 1 +
  5 3 3 3 3 1 1 1 2 3 2 3 1
  3 3 2 3 3 3 1 1 1 3 1 1 1

  3 2 1 3 3 1 2 1 1 3 3 1 2
  3 3 2 3 3 1 3 1 1 3 1 1 3
  3 2 1 1 1 1 1 1 1 3 4 1 3
  3 1 1 3 3 2 3 1 3 3 3 1 2
  3 1 2 3 3 2 3 1 2 3 2 1 1


Andropogon   lateralis
Paspalum  notatum
```

Piptochaetium montevidense
Ruellia sp.
Aristida filifolia
Setaria geniculata
Axonopus affinis
Relbunium hirtum

Andropogon lateralis
Piptochaetium montevidense
Axonopus affinis
Paspalum notatum
Centella biflora
Eryngium horridum
Ruellia sp.
Facelis retusa
Herbertia pulchella
Coelorachis selloana
Sysyrinchium sp.

Sporobolus indicus
Paspalum notatum
Facelis retusa
Sysyrinchium sp.
Richardia humistrata
Herbertia pulchella
Piptochaetium montevidense
Soliva pterosperma
Ruellia sp.
Andropogon lateralis
Eragrostis neesii
Desmodium incanum
Coelorachis selloana
Setaria geniculata

Eryngium horridum
Aspilia montevidensis
Relbunium hirtum
Trachypogon montufari
Axonopus affinis

Paspalum notatum
Piptochaetium montevidense
Facelis retusa
Soliva pterosperma
Sysyrinchium sp.
Aspilia montevidensis
Coelorachis selloana
Sporobolus indicus
Setaria geniculata
Eragrostis neesii

Paspalum notatum
Setaria geniculata
Andropogon lateralis
Richardia humistrata
Facelis retusa
Ruellia sp.
Aristida filifolia
Axonopus affinis
Eryngium horridum
Herbertia pulchella

Sporobolus indicus
Coelorachis selloana

Paspalum notatum
Axonopus affinis
Sysyrinchium sp.
Eragrostis neesii
Coelorachis selloana
Andropogon lateralis
Sporobolus indicus
Ruellia sp.
Relbunium hirtum
Borreria eryngioides
Piptochaetium montevidense
Richardia humistrata
Herbertia pulchella
Aspilia montevidensis
Desmodium incanum

Aspilia montevidensis
Andropogon lateralis
Paspalum notatum
Setaria geniculata
Herbertia pulchella
Piptochaetium montevidense
Ruellia sp.
Sysyrinchium sp.
Eragrostis neesii
Axonopus affinis
Sporobolus indicus
Coelorachis selloana
Aristida filifolia
Desmodium incanum

Eryngium horridum
Andropogon lateralis
Aristida filifolia
Ruellia sp.
Baccharis megapotamica
Paspalum notatum
Relbunium hirtum

Eryngium horridum
Paspalum notatum
Andropogon lateralis
Coelorachis selloana
Sporobolus indicus
Richardia humistrata
Ruellia sp.
Herbertia pulchella
Desmodium incanum
Relbunium hirtum
Piptochaetium montevidense
Aristida filifolia

Paspalum pumilum
Centella biflora
Axonopus affinis
Eleocharis glauco-virens
Andropogon lateralis
Baccharis trimera

```
Andropogon lateralis
Paspalum notatum
Axonopus affinis
Relbunium hirtum
Sysyrinchium sp.
Coelorachis selloana
Sporobolus indicus
Centella biflora
Baccharis trimera
Desmodium incanum
Ruellia sp.
Richardia humistrata

Andropogon lateralis
Coelorachis selloana
Paspalum notatum
Eryngium horridum
Desmodium incanum
Ruellia sp.
Relbunium hirtum
Setaria geniculata
Aristida filifolia
Facelis retusa

Andropogon lateralis
Paspalum notatum
Ruellia sp.
Piptochaetium montevidense
Relbunium hirtum
Setaria geniculata
Aristida filifolia

Paspalum notatum
Eleocharis glauco-virens
Centella biflora
Axonopus affinis
Andropogon lateralis

Data obtained in August 1990, with the help of Mrs. Ilsi Boldrini (Faculdade
de Agronomia, UFRGS).
There are 15 score matrices (relevés), and the corresponding species lists.
In each matrix rows are CSTs and columns are characters (12) followed by
cover-abundance (Braun-Blanquet scale).
The relevé labels are:
a b c d e f g h i j k l m n o, which correspond in the field to labels num-
bered 15 - 29.
The number of CSTs in each matrix is respectively:
   8 11 14 5 10 12 15 14 7 12 6 12 10 7 5
Character set:
Life-form (Raunkiaer 1907)
1.  Form (lf).     1:Phanerophytes, 2:Chamaephytes, 3:Hemicryptophytes,
            4:Geophytes, 5:Therophytes, 6:Other
Growth-form
2.  Form (g1). 1:solitary, 2:rosette, 3:caespitose
3.  Form (g2). 1:prostrated, 2:erect, 3:both
4.  Form (g3). 1:stoloniferous, 2:rhizomatous, 3:else
Stem
5.  Tissue type (st). 1:herbaceous, 2:woody, 3:no stem.
Leaf
6.  Cross section (cr). 1:straight, 2:folded, 3: rolled, 4:else
```

7. Texture (tx). 1-4:herbaceous to leathery, 5:leafless
8. Dorsal epidermal surface (ed). 1:glabrous, 2:glaucous, 3:hairy, 4:tomentose, 5:leafless
9. Ventral epidermal surface (ev). 1:glabrous, 2:glaucous, 3:hairy, 4:tomentose, 5:leafless
10. Armature (ve). 1:prickly, 2:spiny, 3:else
11. Width (wi).1:< 2.5 mm, 2:2.5-5, 3:5-10, 4:10-50, 5:50-100, 6:100<, 0:leafless

Plant height

12. Height class (he). 1: <2.5cm, 2:2.5-5, 3:5-10, 4:10-20, 5:20-50, 6: 50-100, 7:>100cm

Appendix B.Contents of data file "ufrgs15reEnv.dat".

```
 4.4    4.7     4.7     4.8     4.7     4.8     4.8     4.8     4.7     4.8     4.7     4.7
 4.8    4.6     4.6
 3.1    2.5     2.0     1.6     2.5     1.6     1.6     1.2     1.2     2.0     1.6     2.0
 1.7    1.7     2.4
 106    140     138     118     66      130     136     100     66      78      54      70
 86     70      46
 2.9    3.7     2.5     2.5     3       3.2     3.2     2.9     2.8     3.1     3.6     3.2
 3      2.8     3.9
 0.1    0.4     1       0.7     0.9     1.5     0.8     0.5     0.8     1       1.7     1.4
 0.7    0.6     1.4
 1.3    2.1     1.1     1.5     1.9     2       2.1     1.9     1.8     1.9     1       1.5
 1.7    1.6     1.7
 0.7    1.3     0.6     1       1.3     1.3     1.3     1.1     1       0.9     0.7     1.1
 1.1    0.8     0.3
 13.1   42.3    13.1    12.5    15.5    16.7    13.1    12.5    13.1    13.7    10.7    15.5
 13.1   11.3    7.1
 2.2    2.8     1.5     1.6     1.5     1.7     1.9     1.6     2.1     1.9     1.9     2.8
 3      2.5     1.4
 1.4    1.5     0.8     0.9     1.2     1.2     1.3     1       1.7     1.5     1.6     1.5
 1.5    1.7     1.2
 0.2    0.4     0.4     0.4     0.5     0.5     0.5     0.3     0.3     0.2     0.2     0.2
 0.2    0.3     0.3
 41     67      24      28      41      38      46      65      84      90      24      89
 81     80      14
 0.12   0.2     0.06    0.1     0.08    0.1     0.09    0.08    0.11    0.11    0.19    0.25
 0.17   0.14    0.1
 2      3       2       2       1       1       1       1       2       2       4       2
 2      2       5
 1      1       1       1       1       1       1       2       2       2       4       3
 3      3       4
 1      1       1       2       1       1       1       1       2       2       1       1
 2      1       1
```

16 variables in rows, 15 relevés in columns.
Variables are pH, P (ppm), K (ppm), soil organic matter (%), Al (me/dl), Ca (me/dl), Mg (me/dl), S (ppm), Zn (ppm), Cu (ppm), B (ppm), Mn (ppm), Fe (%), soil moisture (1.very dry, 2.dry, 3.mesic, 4.moist, 5.wet), landscape position (1.flat top, 2.convex slope, 3.concave slope, 4.lowland), and grazing intensity (1.grazed, 2.ungrazed)
Taken from Pillar (1988), except grazing intensity, which is updated (August 1990)

Appendix C. Contents of data file "ufrgs15reSp.dat".

```
5 3 2 0 0 5 3 3 5 3 5 5 5 8 3
5 8 7 0 8 7 7 5 3 7 0 7 2 5 5
3 2 3 0 3 0 3 3 0 2 0 0 0 2 0
2 3 3 0 0 3 3 3 2 3 0 2 2 2 0
2 0 0 0 0 3 0 3 1 1 0 0 2 3 0
3 0 2 0 3 3 0 2 0 0 0 0 1 3 0
5 5 0 2 0 1 5 2 0 0 5 2 0 0 5
2 0 0 1 0 0 2 0 2 1 0 2 2 2 0
0 3 0 0 0 0 0 0 0 0 5 2 0 0 7
0 1 0 7 0 2 0 0 7 7 0 0 3 0 0
0 1 1 0 1 1 0 0 0 0 0 0 1 0 0
0 1 2 0 0 1 1 2 0 1 0 0 0 0 0
0 2 2 0 3 3 1 2 0 5 0 5 5 0 0
0 2 1 0 1 0 2 1 0 0 0 1 0 0 0
0 0 5 0 2 2 3 3 0 3 0 1 0 0 0
0 0 1 0 0 1 1 0 0 1 0 1 0 0 0
0 0 1 0 1 0 0 0 0 0 0 0 0 0 0
0 0 2 0 1 0 2 1 0 0 0 0 0 0 0
0 0 3 0 0 0 3 3 0 3 0 3 3 0 0
0 0 0 5 3 0 2 2 0 0 0 0 0 0 0
0 0 0 5 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 5 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 5 0 0 0 0
0 0 0 0 0 0 0 0 0 0 5 0 0 0 7
0 0 0 0 0 0 0 0 0 0 1 1 0 0 0

Andropogon  lateralis
Paspalum  notatum
Piptochaetium  montevidense
Ruellia  sp.
Aristida  filifolia
Setaria  geniculata
Axonopus  affinis
Relbunium  hirtum
Centella  biflora
Eryngium  horridum
Facelis  retusa
Herbertia  pulchella
Coelorachis  selloana
Sysyrinchium  sp.
Sporobolus  indicus
Richardia  humistrata
Soliva  pterosperma
Eragrostis  neesii
Desmodium  incanum
Aspilia  montevidensis
Trachypogon  montufari
Borreria  eryngioides
Baccharis  megapotamica
Paspalum  pumilum
Eleocharis  glauco-virens
Baccharis  trimera

C/A matrix with 26 species (rows) and 15 relevés (columns).
Species name are listed in the same order.
```

Appendix D. Formatted data file "ufrgs5re5ch.formda". The meaning of the contents in each line is within parentesis and is not part of the file.

```
ufrgs15re12ch.formda                                  (original data file name)
5                                                     (number of characters)
g3 cr g1 st wi                                        (character labels)
 2  2  2  2  3                                        (character types)
 0  0  0  0  3                                        (character ranges)
 3  3  3  3  4                                        (number of states found per character)
5                                                     (number of relevés)
a b c m k                                             (relevé labels)
20                                                    (total number of different CSTs)
  7  11  11   9   6                                   (number of CSTs in each relevé)
1 2 3 5 7 8 9                                         (c/a code replacement values)
0302030302     5     3     2     5     0              (CST 1 and its performances)
0201010302     5     0     7     0     0
0303030301     5     2     3     2     0
0301010103     2     3     3     2     0
0301010302     3     2     6     5     0
0102010303     5     5     0     0     5
0301010101     2     1     2     3     0
0201010303     0     8     0     2     0
0101010104     0     3     0     0     0
0201020302     0     1     0     0     0
0301010303     0     1     2     0     0
0301020303     0     2     1     0     5
0302030301     0     0     5     0     0
0201010103     0     0     1     0     0
0301010204     0     0     3     3     1
0101020104     0     0     0     0     5
0301030301     0     0     0     0     5
0302010302     0     0     0     0     5
0201020303     0     0     0     3     0
0301030302     0     0     0     1     0              (CST 20 and its performances)
16                                                    (number of environmental variables)
pH P K c Al Ca Mg S Zn Cu B Mn Fe hu e gr             (environmental variable labels)
ufrgs.env.dat                                         (original environmental data file)
   4.4    4.7    4.7    4.8    4.7                    (pH in the the relevés)
   3.1    2.5      2    1.7    1.6
   106    140    138     86     54
   2.9    3.7    2.5      3    3.6
   0.1    0.4      1    0.7    1.7
   1.3    2.1    1.1    1.7      1
   0.7    1.3    0.6    1.1    0.7
  13.1   42.3   13.1   13.1   10.7
   2.2    2.8    1.5      3    1.9
   1.4    1.5    0.8    1.5    1.6
   0.2    0.4    0.4    0.2    0.2
    41     67     24     81     24
  0.12    0.2   0.06   0.17   0.19
     2      3      2      2      4
     1      1      1      3      4
     1      1      1      2      1                    (gr in the the relevés)
22                                                    (number of different species names)
ufrgs15re12ch.dat                                     (original file with species names)
Anla Andropogon lateralis                             (label and name of species 1)
Pano Paspalum notatum
Pimo Piptochaetium montevidense
Rusp Ruellia sp.
```

```
Arfi Aristida filifolia
Sege Setaria geniculata
Axaf Axonopus affinis
Rehi Relbunium hirtum
Cebi Centella biflora
Erho Eryngium horridum
Fare Facelis retusa
Hepu Herbertia pulchella
Cose Coelorachis selloana
Sysp Sysyrinchium sp.
Spin Sporobolus indicus
Rihu Richardia humistrata
Sopt Soliva pterosperma
Erne Eragrostis neesii
Dein Desmodium incanum
Papu Paspalum pumilum
Elgl Eleocharis glauco-virens
Batr Baccharis trimera                          (label and name of species 22)
11111                                           (presence of species 1 in the relevés)
11110
11100
11110
10010
10110
11001
10010
01001
01010
01110
01100
01110
01100
00100
00100
00100
00100
00110
00001
00001
00001                                           (presence of species 22 in the relevés)
1000000000000000000000                          (species corresponding to CST 1)
0100000000000000000000
0010100000000000000000
0001000000000000000000
0000010000001000010000
0000001000000000000000
0000000100100000100000
0100000000000000000000
0000000010000000000000
0000000001000000000000
0000000000010000000000
0100000000000100000100
0000000000000010000000
0000000000000001000000
0000000000000000001001
0000000010000000000000
0000000000000000000010
1000000000000000000000
0000000001000000000000
0000010000000000000000                          (species corresponding to CST 20)
Thu Jul 23 10:53:35 1992                        (creation date)
```

REFERENCES

Adamson, R. S. 1939. The classification of life-forms of plants. *Botanical Review 5*: 546-561.

Andrade Lima, D. de 1954. *Contribution to the Study of the Flora of Pernambuco, Brazil*. MSc. Thesis, State University of New York, College of Forestry. 154 pp.

Andrade Lima, D. de 1982. Present-day forest refuges in Northeastern Brazil. In: Prance, G. T. (ed.) *Biological Diversification in the Tropics*, pp. 245-251. Columbia University Press, New York.

Banyikwa, F. F., E. Feoli and V. Zuccarello. 1990. Fuzzy set ordination and classification of Serengeti short grasslands, Tanzania. *Journal of Vegetation Science 1*: 97-104.

Barbour, M. G. and D. V. Diaz. 1973. *Larrea* plant communities on bajada and moisture gradients in the United States and Argentina. *Vegetatio 28*: 335-352.

Barkman, J. J. 1988a. New systems of plant growth forms and phenological plant types. In: M. J. A. Werger, P. J. M. van der Aart, H. J. During and J. T. A. Verhoeven (eds.). *Plant Form and Vegetation Structure; adaptation, plasticity and relation to herbivory*, pp. 9-44. SPB Academic Publishing, The Hague. 356 pp.

Barkman, J. J. 1988b. Some reflections on plant architecture and its ecological implications; a personal view demonstrated on two species of Quercus. In: M. J. A. Werger, P. J. M. van der Aart, H. J. During and J. T. A. Verhoeven (eds.). *Plant Form and Vegetation Structure; adaptation, plasticity and relation to herbivory*, pp. 1-7. SPB Academic Publishing, The Hague. 356 pp.

Benzécri, J. P. 1969. Statistical analysis as a tool to make patterns emerge from data. In: S. Watanabe (ed.). *Methodologies of Pattern Recognition*. p 35-74. Academic Press, New York.

Bezdek, J. C. 1974. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology 1*: 57-71.

Bezdek, J. C. 1987. Partition structures: a tutorial. In: Bezdek, J. C. (ed.). *Analysis of Fuzzy Information*, vol. 3, pp. 81-107. CRC Press, Boca Raton, Florida. 296 pp.

Bongers, F. and J. Popma. 1988. Is exposure-related variation in leaf characteristics of tropical rain forest species adaptive? In: Werger, M. J. A., P. J. M. van der Aart, H. J. During and J. T. A. Verhoeven (eds.). *Plant Form and Vegetation Structure*. p. 191-200. SPB Academic Publishing, The Hague. 356 pp.

Box, E. O. 1981. *Macroclimate and Plant Forms*. Junk, The Hague. 258 pp.

Bradfield, G. E. and N. C. Kenkel. 1987. Nonlinear ordination using flexible shortest path adjustment of ecological distances. *Ecology 68*: 750-753.

Bradley, J. V. 1968. *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, New Jersey. 388 pp.

Braga, R. 1960. *Plantas do Nordeste, especialmente do Ceará*. 2nd. ed. Imprensa Oficial, Fortaleza, Brazil. 540 pp.

Braun-Blanquet, J. 1928. Pflanzensoziologie. Springer-Verlag, Berlin. English translation by Fuller, G. D. and H. S. Conard, *Plant Sociology*. McGraw-Hill, New York. 439 pp.

Braun-Blanquet, J. 1964. Pflanzensoziologie; Grundzüge der Vegetationskunde. 3rd ed. Springer, Vienna, New York. Spanish translation, *Fitosociologia; bases para el estudio de las comunidades vegetales*. Blume, Madrid, 1979. 819 pp.

Bray, J. R. and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs 27*: 325-349.

Bucher, E. H. 1982. Chaco and Caatinga - South American arid savannas, woodlands and thickets. In: Huntley, B. J. and B. H. Walker (eds.). *Ecology of Tropical Savannas*, pp. 48-79. Springer-Verlag, New York.

Burgos, J. J. 1963. El clima de las regiones áridas de la República Argentina. *Revista de Investigaciones Agricolas 17*: 385-405.

Cabido, M., S. Diaz and A. Acosta. 1991. Morphological analysis of herbaceous communities and their response to disturbance. In: *Abstracts of the IAVS Symposium on Mechanisms in Vegetation Dynamics*, Eger, Hungary, p. 62.

Cabrera, A. L. 1971. *Fitogeografia de la Republica Argentina*. Sociedad Argentina de Botanica, Buenos Aires. 57pp.

Cain, S. A. 1950. Life-forms and phytoclimate. *Botanical Review 16*: 1-32.

Cain, S. A., G. M. de O. Castro, J. M. Pires and N. T. da Silva. 1956. Application of some phytosociological techniques to Brazilian rain forest. *American Journal of Botany 43*: 911-941.

Clements, F. E. 1916. *Plant Succession; an analysis of the development of vegetation*. Carnegie Institution, Washington, Publication No. 242.

Cody, M. L. 1974. Optimization in Ecology. *Science 183*: 1156-1164.

Cody, M. L. and J. M. Diamond (eds.). 1975. *Ecology and Evolution of Communities*. Harvard University Press, Cambridge. 545 pp.

Constance, L. 1953. The role of ecology in biosystematics. *Ecology 34*: 642-649

Dale, M. B. 1968. On property structure, numerical taxonomy and data handling. In: Heywood, V. H. (ed.). *Modern Methods in Plant Taxonomy*. pp. 185-197. Academic Press, London.

Dale, M. B. 1978. Planning an adaptive numerical classification. *Vegetatio 35*: 131-136.

Dale, M. B. 1988. Some fuzzy approaches to phytosociology: ideals and instances. *Folia Geobotanica et Phytotaxonomica 23*: 239-274.

Dale, M. B. and M. T. Clifford. 1976. On the effectiveness of higher taxonomic ranks for vegetation analysis. *Australian Journal of Ecology 1*: 37-62.

Dansereau, P. 1951. Description and recording of vegetation upon a structural basis. *Ecology 32*: 172-229.

Dansereau, P. 1957. *Biogeography; an ecological perspective*. Ronald Press, New York. 394 pp.

Davis, P. H. and V. H. Heywood. 1963. *Principles of Angiosperm Taxonomy*. Oliver & Boyd, Edinburgh. 556 pp.

De Candolle, A. P. 1818. *Regni vegetabilis systema naturale.* Vol. 1. Paris. *Apud* Du Rietz (1931).

Drude, O. 1887. *Die Systematische und Geographische Anordnung der Phanerogamen.* Schenck, Handbuch der Botanik, Bd. III, 2. Berlin. *Apud* Du Rietz (1931).

Drude, O. 1896. *Deutschlands Pflanzengeographie.* Handbücher zur Deutschen Landes- und Volkskunde. Stuttgart. *Apud* Du Rietz (1931).

Drude, O. 1913. Die Ökologie der Pflanzen. Die Wissenschaft, 50. F. Vieweg, Braunschweig. *Apud* Clements (1916) p. 267-268.

Du Rietz, G. E. 1931. Life-forms of terrestrial flowering plants. *Acta Phytogeographica Suecica 3* : 1-95.

Dwass, M. 1957. Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics 28*: 181-187.

Edgington, E. S. 1969a. *Statistical Inference: The Distribution-Free Approach.* McGraw-Hill, New York. 211 pp.

Edgington, E. S. 1969b. Approximate randomization tests. *The Journal of Psichology 72*:143-149.

Edgington, E. S. 1987. *Randomization Tests.* Marcel Dekker, New York. 341 pp.

Egler, W. A. 1951. Contribuição ao estudo da Caatinga pernambucana. *Revista Brasileira de Geografia 13*: 577-590

Ehleringer, J. R. and K. S. Werk. 1986. Modification of solar-radiation absorption patterns and implications for carbon gain at the leaf level. In: Givnish, T. J. (ed.). *On the Economy of Plant Form and Function.* p. 57-82. Cambridge University Press, Cambridge. 717 pp.

Ehleringer, J., H. A. Mooney, S. L. Gulmon and P. W. Rundel. 1981. Parallel evolution of leaf pubescence in *Encelia* in coastal deserts of North and South America. *Oecologia 49*: 38-41.

Ehrlich, P. R. and R. W. Holm. 1962. Patterns and populations. *Science 137*: 652-657.

Eiten, G. 1982. Brazilian "savannas". In: Huntley, B. J. and B. H. Walker (eds.). *Ecology of Tropical Savannas*, pp. 25-47. Springer-Verlag, New York.

Fekete, G. and J. S. Lacza. 1970. A survey of the plant life-form systems and the respective research approaches II. *Annales Hist.-Natur. Musei Nation. Hungarici, Pars Botanica 62*: 115-127.

Feoli, E. 1984. Some aspects of classification and ordination of vegetation data in perspective. *Studia Geobotanica 4*: 7:21.

Feoli, E., M. Lagonegro and L. Orlóci. 1984. *Information Analysis of Vegetation Data.* Junk, The Hague. 143 pp.

Feoli, E. and L. Orlóci. 1979. Analysis of concentration and detection of underlying factors in structure tables. *Vegetatio 40*: 49-54.

Feoli, E. and L. Orlóci. 1985. Species dispersion profiles of anthropogenic grasslands in the Italian Eastern Pre-Alps. *Vegetatio 60*: 113-118.

Feoli, E. and L. Orlóci. 1991. *Computer Assisted Vegetation Analysis.* Kluwer, Dordrecht. 498 pp.

Feoli, E., L. Orlóci and M. Scimone. 1984. Comparison of ordinations of some Apennine forest communities based on different characters and methods. *Studia Geobotanica 4*: 137-145.

Feoli, E., L. Orlóci and M. Scimone. 1985. Measuring structural convergence of vegetation types on the basis of floristic data. *Abstracta Botanica 9*: 17-32.

Feoli, E. and M. Scimone. 1984. A quantitative view of textural analysis of vegetation and examples of application of some methods. *Archivo Botanico e Biogeografico Italiano 60*: 72-94.

Feoli, E. and M. Scimone. 1987. Phenetic similarity and coexistence in plant communities: some tests and brief discussion. *Giornale Botanico Italiano 121*: 197-199.

Feoli, E. and V. Zuccarello. 1986. Ordination based on classification: yet another solution? *Abstracta Botanica 10*: 203-219.

Feoli, E. and V. Zuccarello. 1988. Syntaxonomy: a source of useful fuzzy sets for environmental analysis? *Coenoses 3*: 141-147.

Fisher, R. A. 1951. *The Design of Experiments.* 6th ed. Oliver and Boyd, Edinburgh. 244 pp.

Ghiselin, M. T. 1987. Species concepts, individuality, and objectivity. *Biology and Philosophy 2*: 127-143.

Gimingham, C. H. 1951. The use of life form and growth form in the analysis of community structure, as illustrated by a comparison of two dune communities. *Journal of Ecology 39*: 396-406.

Gittins, R. 1965. Multivariate approaches to a limestone grassland community. I. A stand ordination. *Journal of Ecology 53*: 385-401.

Givnish, T. J. 1978. On the adaptive significance of compound leaves, with particular reference to tropical trees. In: Tomlinson, P. B. and M. H. Zimmermann (eds.). *Tropical Trees as Living Systems*. p. 351-380. Cambridge University Press, Cambridge.

Givnish, T. J. 1979. On the adaptive significance of leaf form. In: Solbrig, O. T., P. H. Raven, S. Jain and G. B. Johnson (eds.). *Topics in Plant Population Biology,* pp. 375-407. Columbia Univ. Press, New York.

Givnish, T. J. (ed.). 1986. *On the Economy of Plant Form and Function*. Cambridge University Press, Cambridge.

Givnish, T. J. 1987. Comparative studies of leaf form: assessing the relative roles of selective pressures and phylogenetic constrains. *New Phytologist 106*(suppl.): 131-160.

Givnish, T. J. and G. J. Vermeij. 1976. Sizes and shapes of lianes leaves. *American Naturalist 100*: 743-778.

Gomes, M. A. F. 1979. *Padrões da Caatinga nos Cariris Velhos, Paraiba*. M. Sc. Thesis, Universidade Federal Rural de Pernambuco, Departamento de Botânica, Recife, Brazil.

Gomez Sal, A., J. M. De Miguel, M. A. Casado and F. D. Pineda. 1986. Successional change in the morphology and ecological responses of a grazed pasture ecosystem in Central Spain. *Vegetatio 67*: 33-44.

Good, R. 1974. *The Geography of the Flowering Plants*. 4th. ed. Longman, London. 557 pp.

Goodal, D. W. 1963. The continuum and the individualistic association. *Vegetatio 11*: 297-316.

Goodall, D. 1966. A new similarity index based on probability. *Biometrics 22*: 882-907.

Goodal, D., P. Ganis and E, Feoli. 1987. *Probabilistic Methods in Classification: A Manual for Seven Computer Programs*. Universita di Trieste, Trieste. 52 pp.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53*: 325-338

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics 27*: 857-871.

Gower, J. C. 1975. Generalized Procrustes analysis. *Psychometrika 40*: 33-51.

Greig-Smith, P. 1952. The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany 16*: 293-316.

Greig-Smith, P. 1983. *Quantitative Plant Ecology*. 3rd ed. Blackwell, Oxford. 359 pp.

Grime, J. P. 1979. *Plant Strategies and Vegetation Processes*. John Wiley & Sons, New York.

Grime, J. P., R. Hunt and W. J. Krzanowski. 1987. Evolutionary physiological ecology of plants. In: P. Calow (ed.). *Evolutionary Physiological Ecology*. p. 105-125. Cambridge University Press, New York. 239 pp.

Grisebach, A. 1872. *Die Vegetation der Erde nach ihrer klimatischen Anordnung*. 2nd ed. 2 Vols. W. Engelmann, Leipzig, 1884.

Guárdia, R. and J. M. Ninot. 1991. Plant strategies in the badlands of the upper Llobregat basin (Pyrenees). In: *Abstracts of the IAVS Symposium on Mechanisms in Vegetation Dynamics*, Eger, Hungary, p. 74.

Halloy, S. 1990. A morphological classification of plants, with special reference to the New Zealand alpine flora. *Journal of Vegetation Science 1*: 291-304.

Harper, J. L. 1982. After description. In: Newman, E. I. (ed.). *The Plant Community as a Working Mechanism*. p. 11-25. Blackwell Scientific, Oxford. 128 pp.

Hayashi, I. and M. Numata. 1976. Structure and succession of Caatinga vegetation in the Brazilian Northeast. *Tokyo Geographical Papers 20*: 23-44.

Heywood, V. H. 1973. Chemosystematics, an artificial discipline. In: Bendz, G. and J. Santesson (eds.). *Chemistry in Botanical Classification*. p. 41-54. Academic Press, New York. 320 pp.

Heywood, V. H. 1976. *Plant Taxonomy*. 2nd. ed. E. Arnold, London. 63 pp.

Hope, A. C. A. 1968. A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society 30*: 582-598.

Horikawa, Y. and A. Miyawaki. 1954. Studies on the growth forms of weeds as related to community structures. *Japanese Journal of Ecology 4*: 79-88.

Horn, H. S. 1971. *The Adaptive Geometry of Trees*. Princeton University Press, Princeton, New Jersey. 144 pp.

Hueck, K. 1966. *Die Wälder Südamerikas*. G. Fisher, Stuttgart. 422 pp.

Hueck, K. and P. Seibert. 1972. *Vegetationskarte von Südamerika*. G. Fischer, Stuttgart. 71 pp.

Humboldt, A. von. 1806. Ideen zu einer Physiognomik der Gewächse. Cotta, Stuttgart. English translation by E. C. Otté and H. G. Bohn, as Ideas for a physiognomy of plants. In: Humboldt, A. von. *Views of Nature*. p. 210-352. H. G. Bohn, London, 1850. 452 pp.

Jancey, R. C. and T. C. Wells. 1987. Locality theory: the phenomenon and its significance. *Coenoses 2*: 1-8.

Jardine, N. and R. Sibson. 1971. *Mathematical Taxonomy*. Wiley, London. 286 pp.

Juhász-Nagy, P. and J. Podani. 1983. Information theory methods for the study of spatial processes and succession. *Vegetatio* 51: 129-140.

Karr, J. R. and F. C. James. 1975. Eco-morphological configurations and convergent evolution in species and communities. In: Cody, M. L. and J. M. Diamond (eds.). *Ecology and Evolution of Communities*, pp. 258-291. Harvard University Press, Cambridge, Massachusetts.

Kempthorne, O. 1952. *The Design and Analysis of Experiments*. J. Wiley, New York. 631 pp.

Kempthorne, O. 1955. The randomization theory of experimental inference. *Journal of the American Statistical Association 50*: 946-967.

Kenkel, N. C. 1984. *Boreal Vegetation of a Lacustrine Surface Sand Belt, Elk Lake, Ontario: Types and Environmental Gradients*. Ph.D. Thesis, University of Western Ontario, London, Canada. 380 pp.

Kenkel, N. C., P. Juhász-Nagy and J. Podani. 1989. On sampling procedures in population and community ecology. *Vegetatio 83*: 195-207.

Kerner von Marilaun, A. 1863. Das Pflanzenleben der Donauländer. Reprint by F. Vierhapper, Vienna, 1929. English translation by H. S. Connard, *The Background of Plant Ecology*. Iowa State College, Ames, 1950.

Kerner von Marilaun, A. 1891. Pflanzenleben. English translation by Oliver, F. W., *The Natural History of Plants*. Blackie & Son, London, 1894-1895.

Kernighan, B. W. and D. M. Ritchie. 1988. *The C Programing Language*. 2nd. ed. Prentice-Hall, Englewood Cliffs, N.J. 272 pp.

Kershaw, K. A. 1964. *Quantitative and Dynamic Ecology*. Arnold, London. 183 pp.

Kirmse, R. D., J. A. Pfister, L. V. Vale and J. S. de Queiroz. 1983. *Woody Plants of the Northern Ceará Caatinga.* Small Ruminant Collaborative Research Support Program, Technical Report Series no. 14, Utah State University, Department of Range Science, Logan, Utah.

Knight, D. H. 1965. A gradient analysis of Wisconsin prairie vegetation on the basis of plant structure and function. *Ecology 46*:744-747.

Knight, D. H. and O. L. Loucks. 1969. A quantitative analysis of Wisconsin forest vegetation on the basis of plant function and gross morphology. *Ecology 50*: 219-234.

Lacza, J. S. and G. Fekete. 1969. A survey of the plant life-form systems and the respective research approaches I. *Annales Hist.-Natur. Musei Nation. Hungarici, Pars Botanica 61*: 129-139.

Lambert, J. M. and M. B. Dale. 1964. The use of statistics in phytosociology. *Advances in Ecological Research 2*: 59-99.

Lausi, D. and P. L. Nimis. 1985. The study of convergent evolution in plants and plant communities; a quantitative approach. *Abstracta Botanica 9*: 67-77.

Lausi, D. and P. L. Nimis. 1986. Leaf and canopy adaptations in a high-elevation desert on Tenerife, Canary Islands. *Vegetatio 68*: 19-31.

Lausi, D., P. L. Nimis and M. Tretiach. 1989. Adaptive leaf structures in a *Myrica-Erica* stand on Tenerife (Canary Islands). *Vegetatio 79*: 133-142.

Maarel, E. van der. 1972. Ordination of plant communities on the basis of their plant genus, family and order relationships. In: Tüxen, R. and E. vand der

Maarel (eds.). *Grundfragen und Methoden in der Pflanzensoziologie*. p. 183-190. Junk, The Hague.

Maarel, E. van der. 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio 39*: 97-114.

Maarel, E. van der (ed.). 1980. *Classification and Ordination*. Junk, The Hague. 188pp.

Marsili-Libelli, S. 1989. Fuzzy clustering of ecological data. *Coenoses 4*: 95-106.

Mayr, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York. 334 pp.

Mayr, E. 1961. Cause and effect in biology. *Science 134*: 1501-6.

Mayr, E. 1982. *The Growth of Biological Thought*: *diversity, evolution and inheritance*. Belknap Press, Cambridge, Massachusetts. 974 pp.

McMillan, C. 1969. Discussion. In: *Systematic Biology; Proceedings of an International Conference*, held at The University of Michigan, Ann Harbor, June 1967. p. 203-206. Washington, National Academy of Sciences. 632 pp.

Michener, C. D. 1970. Diverse approaches to systematics. *Evolutionary Biology 4*: 1-38.

Michener, C. D. and R. R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution 11*: 130-162.

Minchin, P. R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio 69*: 89-107.

Mooney, H. A. 1974. Plant forms in relation to environment. In: Strain, B. R. and Billings, W. D. (eds.). *Vegetation and Environment*. Handbook of Vegetation Science 6, pp. 111-122. Junk, The Hague.

Mooney, H. A. and E. L. Dunn. 1970. Convergent evolution in mediterranean-climate evergreen sclerophyll shrubs. *Evolution 24*: 292-303.

Mooney, H. A., E. L. Dunn, F. Shropshire and L. Song. 1970. Vegetation comparisons between the mediterranean climatic areas of California and Chile. *Flore 159*: 480-496.

Mueller-Dombois, D. and H. Ellenberg. 1974. *Aims and Methods of Vegetation Ecology*. Wiley, New York. 547 pp.

Naveh, Z. 1967. Mediterranean ecosystems and vegetation types in California and Israel. *Ecology 48*: 445-459.

Noble, I. R. and R. O. Slatyer. 1980. The use of vital attributes to predict successional changes in plant communities subject to recurrent disturbances. *Vegetatio 43*: 5-21

Numata, M. 1970. Flora and vegetation in the North-Eastern and Central Brazil. *Tokyo Geographical Papers 14*: 43-74.

Orians, G. H. and R. T. Paine. 1983. Convergent evolution at the community level. In: Futuyma, D. J. and M. Slatkin (eds.). *Coevolution*, pp. 431-457. Sinauer, Sunderland, Massachusetts.

Orians, G. H. and O. T. Solbrig (eds.). 1977a. *Convergent Evolution in Warm Deserts*. Dowden, Hutchison and Ross, Stroutdsburg, Pennsylvania. 333 pp.

Orians, G. H. and O. T. Solbrig. 1977b. A cost-income model of leaves and roots with special reference to arid and semiarid areas. *American Naturalist 111*: 677-690.

Orlóci, L. 1966. Geometric models in ecology; the theory and application of some ordination methods. *Journal of Ecology 54*: 193-215

Orlóci, L. 1967a. An agglomerative method for classification of plant communities. *Journal of Ecology 55*: 193-205.

Orlóci, L. 1967b. Data centering: a review and evaluation with reference to component analysis. *Systematic Zoology 16*: 208-212.

Orlóci, L. 1971. An information theory model for pattern analysis. *Journal of Ecology 59*: 343-349.

Orlóci, L. 1972. On objective functions of phytosociological resemblance. *The American Midland Naturalist 88*: 28-55.

Orlóci, L. 1975. Measurement of redundancy in species collections. *Vegetatio 31*: 65-67.

Orlóci, L. 1978. *Multivariate Analysis in Vegetation Research*. 2nd ed. Junk, The Hague. 451 pp.

Orlóci, L. 1988a. Detecting vegetation patterns. *ISI Atlas of Science 1*: 173-177.

Orlóci, L. 1988b. Community organization: recent advances in numerical methods. *Canadian Journal of Botany 66*: 2626-2633.

Orlóci, L. 1990. Statistics in ecosystem survey: computer support for process-based sample stability tests and entropy/information inference. *Abstracta Botanica 14*: 31-49.

Orlóci, L. 1991. On character-based plant community analysis: choice, arrangement, comparison. *Coenoses 5* (in press).

Orlóci, L. and E. Beshir. 1976. A heuristic test for homogeneity in species populations. *Vegetatio 31*: 141-145.

Orlóci, L., E. Feoli, D. Lausi and P. Nimis. 1986. Estimation of character structure convergence (divergence) in plant communities; a nested hierarchical model. *Coenoses 1*: 11-20.

Orlóci, L. and N. C. Kenkel. 1985. *Introduction to Data Analysis; with examples from population and community ecology*. International Co-operative Publishing House, Fairland, Maryland. 340 pp.

Orlóci, L. and M. M. Mukkattu. 1973. The effect of species number and type of data on the resemblance structure of a phytosociological collection. *Journal of Ecology 61*: 37-46.

Orlóci, L. and M. Orlóci. 1985. Comparison of communities without the use of species: model and example. *Annali di Botanica 43*: 275-285.

Orlóci, L. and M. Orlóci. 1990. Edge detection in vegetation: Jornada revisited. *Journal of Vegetation Science 1*: 311-324.

Orlóci, L. and V. D. Pillar. 1989. On sample size optimality in ecosystem survey. *Biometrie-Praximetrie 29*: 173-184.

Orlóci, L. and S. L. Stofella. 1986. The taxon-free approach to the study of plant communities. *Annals of Arid Zone 25*: 111-131.

Parsons, D. J. 1976. Vegetation structures in the mediterranean scrub community of California and Chile. *Journal of Ecology 64*: 435-447.

Peters, R. H. 1976. Tautology in evolution and ecology. *American Naturalist 110*: 1-12

Pielou, E. C. 1966. The measurement of diversity in different types of biological collection. *Journal of Theoretical Biology 13*: 131-144.

Parkhurst, D. F. and O. L. Loucks. 1972. Optimal leaf size in relation to environment. *Journal of Ecology 60*: 505-537.

Pillar, V. De Patta. 1988. *Fatores de Ambiente Relacionados à Variação da Vegetação de um Campo Natural*. M.Sc. Thesis, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Pillar, V. DePatta. 1992. *The Theory of Character-Based Community Analysis*. Ph.D. Thesis, The University of Western Ontario, London, Canada.

Pillar, V. D. 1999. On the identification of optimal plant functional types. *Journal of Vegetation Science 10*: 631-640.

Pillar, V. D., A. V. A. Jacques and I. I. Boldrini. 1989. Environmental Related Variation in a Natural Grassland of Rio Grande do Sul, Brazil. *Proceedings of the 16th International Grassland Congress*, Nice, France, pp. 1527-1528.

Pillar, V. D. and L. Orlóci. 1991. Fuzzy components in community level comparisons. In: Feoli, E. and L. Orlóci (eds.). *Computer Assisted Vegetation Analysis*. p. 87-93. Kluwer, Dordrecht. 498 pp.

Pillar, V. D. and E. E. Sosinski Jr. 2003. An improved method for searching plant functional types by numerical analysis. *Journal of Vegetation Science 14*: 323-332.

Pitman, E. J. G. 1937. Significance tests which may be applied to samples from any populations. II. The correlation coefficient. *Journal of the Royal Statistical Society (Series B) 4*: 225-232.

Podani, J. 1982. *Spatial Processes in the Analysis of Vegetation*. Ph.D. Thesis, University of Western Ontario, London, Canada. 337 pp.

Podani, J. 1984. Spatial processes in the analysis of vegetation: theory and review. *Acta Botanica Hungarica 30*: 75-118.

Podani, J. 1986. Comparison of partitions in vegetation studies. *Abstracta Botanica 10*: 235-290.

Podani, J. 1989. Comparison of ordinations and classifications of vegetation data. *Vegetatio 83*: 111-128.

Podani, J. 1991. On the standardization of Procrustes statistics for the comparison of ordinations. *Abstracta Botanica 15*: 43-46.

Podani, J. and T. A. Dickinson. 1984. Comparison of dendrograms: a multivariate approach. *Canadian Journal of Botany 62*: 2765-2778.

Poore, M. E. D. 1956. The use of phytosociological methods in ecological investigations. IV. General discussion of phytosociological problems. *Journal of Ecology 44*: 28-50.

Poore, M. E. D. 1962. The method of successive approximation in descriptive ecology. *Advances in Ecological Research 1*: 35-69, Academic Press, New York.

Pound, R. and F. E. Clements. 1898. *Phytogeography of Nebraska*. 1st ed. Lincoln, Nebraska. Reprinted by Arno Press, New York, 1977.

Prado, D. E. 1991. *A Critical Evaluation of the Floristic Links Between the Chaco and the Caatingas Vegetation in South America*. Ph.D. Thesis, University of St. Andrews, St. Andrews, Scotland, UK.

Queiroz, J. S. de 1985. *The Acarau Valley in Northeast Brazil: Vegetation, Soils and Land Use*. Ph.D. Thesis. Utah State University, Logan, Utah.

Rambo, Pe. B. 1956. *A Fisionomia do Rio Grande do Sul*. 2nd. ed. Selbach, Porto Alegre. 456pp.

Rao, C. R. 1952. *Advanced Statistical Methods in Biometric Research.* Hafner Press, New York. 390 pp.

Raunkiaer, C. 1904. Om biologiske Typer, med Hensyn til Planternes Tilpasning til at overleve ugunstige Aarstider. Botanisk Tidsskrift 26. English translation as "Biological types with reference to the adaptation of plants to survive the unfavourable season" in: *The Life Forms of Plants and Statistical Plant Geography; the Collected Papers of C. Raunkiaer.* p. 1. Clarendon Press, Oxford, 1934.

Raunkiaer, C. 1907. Planterigets Livsformer og deres Betydning for Geografien. Munksgaard, Copenhagen. English translation as "The life-forms of plants and their bearing on geography" in:*The Life Forms of Plants and Statistical Plant Geography; the Collected Papers of C. Raunkiaer*. p. 2-104. Clarendon Press, Oxford, 1934.

Raunkiaer, C. 1908. Livsformernes Statistik som Grundlag for biologisk Plantegeografi. Botanisk Tidsskrift 29. English translation as "The statistics of life-forms as a basis for biological plant geography" in: *The Life Forms of Plants and Statistical Plant Geography; the Collected Papers of C. Raunkiaer*. p. 111-147. Clarendon Press, Oxford, 1934.

Raunkiaer, C. 1916. Om Bladstørrelsens Anvendelse i den biologisk Plantegeografi. Botanisk Tidsskrift 34. English translation as "The use of leaf size in biological plant geography" in: *The Life Forms of Plants and Statistical Plant Geography; the Collected Papers of C. Raunkiaer*. p. 368-378. Clarendon Press, Oxford, 1934.

Reiter, H. 1885. *Die Consolidation der Physiognomik. Als Versuch einer Oekologie der Gewächse*. Graz. *Apud* Du Rietz (1931).

Roberts, D. W. 1986. Ordination on the basis of fuzzy set theory. *Vegetatio 66*: 123-131.

Roberts, D. W. 1989. Fuzzy systems vegetation theory. *Vegetatio 83*: 71-80.

Rohlf, F. J. 1965. A randomization test of the nonspecificity hypothesis in numerical taxonomy. *Taxon 14*: 262-267.

Rohlf, F. J. 1977. A note on the measurement of redundancy. *Vegetatio 34*: 63-64.

Rosenberg, A. 1985. *The Structure of Biological Science.* Cambridge University Press, N. York. 281 pp.

Salisbury, E. J. 1940. Ecological aspects of plant taxonomy. In: Huxley, J. (ed.). *The New Systematics*. p. 329-340. Clarendon Press, Oxford. 583 pp.

Sarmiento, G. 1972. Ecological and floristic convergence between seasonal plant formations of tropical and subtropical South America. *Journal of Ecology 40*: 367-410.

Schimper, A. F. W. 1898. Pflanzengeographie auf physiologischer Grundlage. Fischer, Jena. English translation as *Plant Geography upon a Physiological Basis*. Clarendon Press, Oxford, 1903. 839 pp.

Shimwell, D. W. 1971. *The Description and Classification of Vegetation.* University of Washington Press, Seattle, 1972. 322 pp.

Shöneman, P. H. and Carroll, R. M. 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika 35*: 245-256.

Shreve, F. 1942. The desert vegetation of North America. *Botanical Review 8*: 195-246.

Simpson, G. G. 1961. *Principles of Animal Taxonomy*. Columbia University Press, New York. 247 pp.

Snaydon, R. W. 1973. Ecological factors, genetic variation and speciation in plants. In: Heywood, V. H. (ed.). *Taxonomy and Ecology*. p. 1-29. Academic Press, London. 370 pp.

Sneath, P. H. A. and R. R. Sokal. 1973. *Numerical Taxonomy*. Freeman, San Francisco. 573 pp.

Sokal, R. R. 1965. Statistical methods in systematics. *Biological Reviews 40*: 337-391.

Sokal, R. R. and T. J. Crovello. 1970. The biological species concept: a critical evaluation. *American Naturalist 104*: 127-153.

Sokal, R. R. and F. J. Rohlf. 1962. The comparison of dendrograms by objective methods. *Taxon 11*: 33-39.

Soriano, A. and Prego, A. J. 1963. *Las tierras áridas y semiáridas de la República Argentina*. Conferencia Latinoamericana para el Estudio de las Regiones Aridas. Buenos Aires, Informe Nacional. 59 pp.

Specht, R. L. 1969. A comparison of the sclerophyllous vegetation characteristic of mediterranean type climates in France, California, and southern Australia. I. Structure, morphology, and succession. *Australian Journal of Botany 17*: 277-292.

Stace, C. A. 1989. *Plant Taxonomy and Biosystematics*. 2nd ed. E. Arnold, London. 264pp.

Stowe, L. G. and J. L. Brown. 1981. A geographic perspective on the ecology of compound leaves. *Evolution 35*: 818-821.

Stuessy, T. F. 1990. *Plant Taxonomy*. Columbia University Press, New York. 514 pp.

Thompson, D'Arcy. 1917. *On Growth and Form*. Cambridge University Press, Cambridge. 793 pp.

Tilman, D. 1988. *Plant Strategies and the Dynamics and Structure of Plant Communities*. Princeton University Press, Princeton, New Jersey. 362 pp.

Walter, H. 1973. Vegetationszonen und Klima. 2nd ed.. English translation by J. Wieser. *Vegetation of the Earth in Relation to Climate and the Eco-Physiological Conditions*. Springer-Verlag, New York, 1973. 273 pp.

Walters, S. M. 1963. Methods of classical plant taxonomy. In: Swain, T. (ed.). *Chemical Plant Taxonomy*. p.1-15. Academic Press, London. 543 pp.

Warming, E. 1884. *Om Skudbygning, Overvintring og Foryngelse.* Festskrift Naturhistorisk Forenings, Copenhagen. *Apud* Du Rietz (1931).

Warming, E. 1895. *Plantesamfund; Grundtraek af den økologiske Plantegeografi*. Philipsens, Copenhagen.

Warming, E. 1909. *Oecology of Plants: An Introduction to the Study of Plant Communities*. Clarendon Press, Oxford. 422 pp. (modified version of Plantesamfund, 1895)

Whittaker, R. H. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. *Ecological Monographs 22*: 1-44.

Wildi, O. and L. Orlóci. 1990. *Numerical Exploration of Community Patterns*. SPB Academic Publishing, The Hague. 124 pp.

Wilkins, D. A. 1968. The scale of genecological differentiation. In: Heywood, V. H. (ed.). *Modern Methods in Plant Taxonomy*. pp. 227-239. Academic Press, London.

Williams, W. T. and M. B. Dale. 1962. Partition correlation matrices for heterogeneous quantitative data. *Nature 196*: 602.

Wilson, E. O. 1975. *Sociobiology: The New Synthesis*. Harvard University Press, Cambridge, Massachusetts. 697 pp.

Wood, J. G. 1934. The physiology of xerophytism in Australian plants. The stomatal frequencies, transpiration and osmotic pressures of sclerophyllous and tomentose-succulent leaved plants. *Journal of Ecology 22*: 69-87.

Zadeh, L. A. 1965. Fuzzy sets. *Information and Control 8*: 338-353.

INDEX