ANALYSIS AND CLASSIFICATION OF VEGETATION BASED ON FAMILY COMPOSITION*

Roger DEL MORAL & Melinda F. DENTON**

Department of Botany (AK-10), University of Washington, Seattle, Washington 98195, USA

Keywords: Classification, Discriminant analysis, Family composition, Indicator families, Phytogeography, Principal components analysis

Introduction

Many plant species are so consistent in their ecological distributions that they are widely described in vegetation studies as useful indicators of particular conditions. Species have most often been cited for their indicator value, but higher taxonomic levels also have been suggested (Pryor 1959, Kruckeberg 1969, van der Maarel 1972, Dale & Clifford 1976). It appears that certain plant families have evolved by adaptive radiation within particular, circumscribed environments; these families consist of taxa whose salient ecological features are similar. For example, families such as *Cactaceae* and *Agavaceae* are confined to warm deserts, while others, such as *Chenopodiaceae* and *Ericaceae*, are usually restricted to particular soils. The range of environmental conditions spanned by some families is known to be broad (as in *Rosaceae, Fabaceae, Apiaceae* or *Asteraceae*) or narrow (as in *Chenopodiaceae, Lemnaceae*, or *Balanophoraceae*), but the extent to which families serve as ecological indicators in vegetation analysis has not been adequately surveyed.

Attempts to construct a classification of regional vegetation are constrained by the ecological convergence of a number of different species. In the Pacific Northwest, forests characterized by *Abies amabilis* are highly variable at the species level though the generic composition of the communities is similar (Franklin & Dyrness 1973, del Moral 1973, 1974). If data based on genera and families could be utilized, then meaningful ecological patterns might emerge from classification or other analytical methods by revealing patterns submerged at the species level.

Van der Maarel (1972) is among those to suggest the use of higher taxa in ordination of vegetational data. He employed a system parallel to Braun-Blanquet’s hierarchical scheme in which the classifications are based on higher taxa. By performing principal components analysis (PCA) on 51 associations with 32 plant orders, van der Maarel established ecological meaning with respect to the first six components. He also reported some success with multivariate studies at the familial and generic levels. Using a small set of data, Dale & Clifford (1976) compared classifications based respectively on species, subgenera, genera, subfamilies, and families. All of these classifications were similar, but similarity to the species-based classification declined as taxonomic rank increased. At the familial rank, they found a different, but informative classification. We agree with Dale & Clifford that the recovery of any ecological information based on taxonomic ranks, above the species category is remarkable. For that reason, we have pursued this type of investigation by combining ordination and classification.

Classification systems of regional vegetation may be developed if those families with diagnostic ecological characteristics can be identified. In this report, we try to identify such families in the vegetation of the Pacific Northwest. Several multivariate methods are used to address the following questions: What success can plant taxa above the rank of species be used to indicate particular ecological conditions? Can taxonomic ranks at the familial level form the basis for the organization of a classification of regional vegetation? Can the information loss intrinsic
in the use of aggregated data be compensated for by the emergence of other patterns, such as correlations of families with specific habitat properties?

Methods

Data selection

Compositional data were gathered from an extensive geographical range to represent several physiognomic types and successional states. Examples of forests, steppes, shrublands, grasslands, and meadows are included. While most samples represent mature vegetation, some represent pioneer or seral stages. Fifty-three families of seed plants and the order Filicales were encountered. The primary data matrix was obtained from various published and unpublished sources and consists of 54 taxa by 105 samples.

Sources are as follows: five samples (1-5) from serpentinite soil, Wenatchee Mountains, Washington, and nine samples (6-14) from normal soils, Wenatchee Mountains (del Moral 1972, 1974); 14 samples (15-28) along a transect over Stevens Pass, Washington and seven samples (29-35) from lowland forests, Olympic Mountains, Washington (del Moral unpubl.); five samples (36-40) along a primary successional series, Hoh River Valley, Washington (Fonda 1974); 11 samples (41-51) from montane areas, Alsek River, Southwest Yukon Territory (Douglas 1974); 15 samples (52-66) along a transect across the central Washington Cascade Mountains (del Moral 1976, del Moral et al. 1976); seven samples (67-73) from subalpine forests, Olympic Mountains (Fonda & Bliss 1969); five samples (74-78) from the Cedar River, Washington (del Moral & Long, unpubl.); six samples (79-84) from Findley Lake Basin, upper Cedar River, Washington (del Moral 1973); eight samples (85-92) from subalpine meadows, eastern Olympic Mountains (Kuramoto & Bliss 1970); six samples (93-98) from the Rattlesnake Hills, Washington (del Moral unpubl.); two samples (99-100) from the Mima Mounds Prairie, Washington (del Moral & Deardorff 1976); and five samples (101-105) from subalpine meadows, dry eastern Olympic Mountains (J. A. Belsky, unpubl.).

In these studies, data concerning dominant species were reported in several ways. To make the samples comparable, relative dominance scores for each species were summed by families. Rare species were not reported by most authors, but such are usually in families already well represented in the sample. The failure to include rare species does not materially influence principal components analysis or classification methods emphasizing dominance.

Principal components analysis

The structure of the data matrix was analyzed by PCA of the covariance matrix using CEP-7 from Cornell University (Gauch 1972). PCA provides several mutually orthogonal axes of variation by a least squares criterion such that each axis in the sequence accounts for more variance than the one that follows. PCA is useful for the detection of ecological trends and provides unique information about the eigenstructure of the variables.

Classification

All samples were drawn from studies in which similarity relationships between samples were stated explicitly. A subjective classification based on species composition was readily accomplished, and was used for comparison with a numerical classification of associations based on families as the variables. The latter was generated by MINFO, an agglomerative, hierarchical, and polythetic method based on minimum gain of information (Orlóci 1969, Goldstein & Grigal 1972). Each of these classifications was then subjected to stepwise multiple discriminant analysis (del Moral 1975, Grigal & Ohmann 1975, Denton & del Moral 1976) to determine whether provisional groups were statistically distinct and to reallocate problematic associations. The MINFO clustering of associations was terminated at $2 \times 10^2$ bits, resulting in 23 classes. These classes were revised by two discriminant analyses to provide the following: an analysis of familial composition of the resultant groups, the discrimination and the classification functions, and the rank order of families used in discrimination.

Results and discussion

Environmental characteristics

The characteristics of the samples used in this study appear in Table 1. Samples are grouped according to the final classification provided by discriminant analysis. Reallocated samples are indicated by an asterisk. The most abundant families and general habitat conditions are noted.

Principal components analysis

Preliminary studies indicated that rare families did not contribute significantly to an understanding of the overall data structure. Accordingly, centered, nonstandardized