COENOSES 2(1): 31-37, 1987

LOCALITY THEORY: THE PHENOMENON AND ITS SIGNIFICANCE

R.C. Jancey and T.C. Wells, Department of Plant Sciences University of Western Ontario London, Ontario N6A 5B7

Keywords: Species, Ranking, Locality, Importance, Pattern, Dynamics, Vegetation, Terraces

Abstract: Species ranking at various dichotomies of the dendrogram has been illustrated. It must be emphasised that not only those species which show marked changes in rank are of interest, but also those which remain more or less consistently important, either in one branch of the dendrogram, or, in some cases, in several branches.
It is felt most strongly that the dynamic changes in species importance direct attention to fundamental questions regarding the current use of species data in the elucidation of vegetation pattern.
The programs were written in BASIC for a PDP-1090 computer, and have now been developed to the point where both speed and core requirements are quite acceptable, though transfer to a micro-computer would require memory expansion. Copies of the programs are available from the authors.

Introduction

The concept of ranking species in terms of their ability to reveal patterns in a study area is not new. The usual intent of ranking by some well defined algorithm is to find that subset of species which is best able to reveal pattern among relevés; the ranking algorithm being matched to whichever clustering or ordination method the investigator chooses to use (Orlóci 1978). An alternative strategy was proposed (Jancey 1980), where the objective was not economy of sampling effort, but rather, more precise pattern elucidation. Ranking would thus exclude those species which contribute little to vegetation pattern, i.e. those which occur in an apparently random fashion.

Species «importance» exists only in the context of a particular study area. To the extent that pattern exists within a set of relevés, the same principle of locality of importance must exist within the study area. This locality problem was addressed (Jancey 1980), using a small, artificial data set. An iterative ranking technique was used to ensure that within successive subsets of relevés, species which were creating random «noise» were excluded no matter how valuable they might be in other subsets of the data. Since that first publication, the program has been extensively modified and run on many sets of real data. We report here an example using data from a set of three river terraces near Hope, B.C. Canada (see Fewster and Orlóci 1978, Jancey 1979).

While the original objective of the technique was increased precision of pattern recognition, we have found, since using real data, that changes in species rank as the clustering process proceeds is a feature of even greater interest. This phenomenon, which might be called pattern dynamics, gives rise in turn to an interest in the implications of locality of species importance. This topic will be explored in the discussion section of this paper.

Methodology

The minimizing of random variation in ecological data sets in order to improve the recognition of group structure through classification is accomplished using a combined ranking and clustering algorithm. The algorithm involves iteration through a chained series of programs as outlined in flowchart form in Figure 1.

Three user supplied files are utilized as input: a raw data matrix DATA, consisting of P sets of species by N relevé scores, TAXON which holds the species names, and GRP containing the relevé identification codes. SRUN is the first of the chaining programs whose function is to initialize the data files and to allow the user to specify several control options used in the program.

The program begins with an initial ranking of the species list. A choice of two ranking algorithms is available. Program SRANK orders species on a sum of squares criterion (Orlóci 1973, 1978) according to their independent contribution to the total shared information. Alternatively, SWIGSS ranks species solely with respect to their unique component of pattern information where shared information is not taken into account (Shaukat 1985). Both ranking algorithms produce a reduced data set of $Pr \times N$ elements, where Pr represents a subset of the original species list accounting for a cumulative percentage of the total variation. The cumulative percentage cutoff level is specified by the user.

Following the species ranking, two resemblance matrices based on euclidean distance are produced. The first matrix is obtained by program SESSA which uses the relevé scores from the reduced species list. The second matrix calculated by STESSA is based upon relevé scores for the entire species list. These resemblance matrices are in turn utilized by program SCHSSA which subdivides the relevés into two vectors (groups) using a sum of squares hierarchial clustering method (Orlóci, 1978). In addition, two fusion levels are calculated: one based on the resemblance matrix using the reduced species list (i.e. the ranked species) and the second fusion level based on the resemblance matrix using all the species. The ratio of SSRANK/SSTOTAL is then calculated to assess the efficiency of the clustering, where SSRANK refers to sum of squares in the relevant set of species and SSTOTAL refers

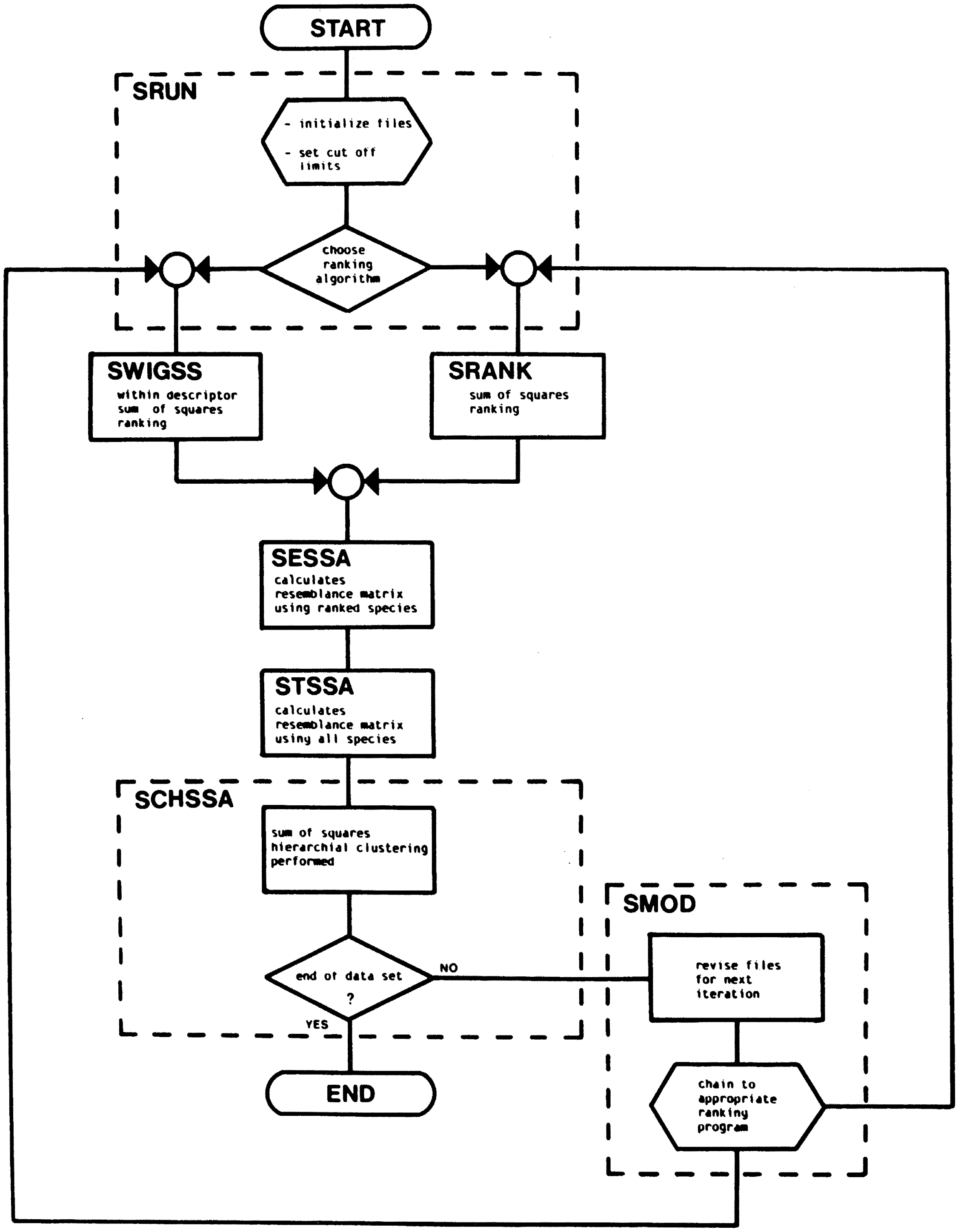

Fig. 1: Flowchart of the iterative ranking and clustering algorithm. Named blocks represent the individual chained programs referred to in the text.

to sum of squares in the complete set of all species.

The two relevé vectors formed in the ranked set represent the initial major branches (or subdivisions) in the final dendrogram. The program proceeds by printing out the ranked species involved in the clustering followed by the two sets of relevé identifiers. Program SMOD then updates the files and chains the appropriate ranking algorithm to start the next iteration. Each subdivision is subsequently partitioned into smaller branches by successive iterations, where each new partition is based on a new ranking of the entire species list for the subset of relevés involved. The ranking of species for the clustering of small numbers of relevés is not particularly meaningful and thus is discontinued once the number of relevés being clustered falls below a user defined cutoff level. The clustering of the remaining relevés then proceeds using the subset of species from the last ranking until all the relevé groups are subdivided to their lowest pairings. Alternatively, a clustering cutoff level can be set which will stop the partitioning of the relevé vector once the number of relevés falls below a critical value. After all the relevés in one branch of the final fusion have been grouped, the program then explores the next branch by the same process until all possible fusions in the dendrogram are formed.

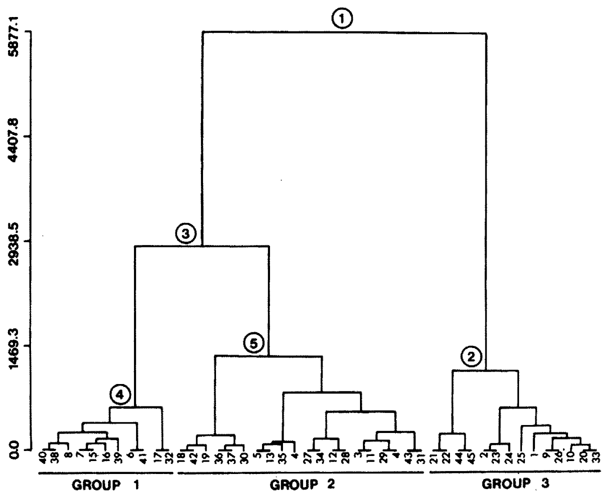

Fig. 2: Dendrogram of the Hope, B.C. data produced by the iterative ranking and clustering method. The vertical axis indicates the sums of squares fusion levels while the horizontal axis contains relevé identifiers. Circled numbers refer to tables 1 to 5, respectively which outline the ranked species used in forming the fusion indicated.

Table 1. Dichotomy Number 1 (see Fig. 2).

SPECIES	RANK	SUM OF SQUARES	PERCENTAGE VARIANCE ACCOUNTED FOR
Polystichum munitum	1	484.80	6.69
Gaultheria shallon	2	427.77	5.90
Hylocomium splendens	3	407.20	5.62
Mnium insigne	4	367.91	5.08
Tsuga heterophylla	5	346.31	4.78
Mahonia nervosa	6	325.24	4.49
Pachistima myrsinites	7	319.20	4.40
Thuja plicata	8	295.24	4.07
Acer circinatum	9	286.00	3.95
Symphoricarpos albus	10	284.97	3.93
Clintonia uniflora	11	276.44	3.81
Disporum hookeri	12	240.57	3.32
Lactuca canadensis	13	220.97	3.05
Achlys triphylla	14	189.20	2.61
Pseudotsuga menziesii	15	153.91	2.12
Hylocomium sp.	16	144.31	1.99
Eurhynchium oreganum	17	140.97	1.94
Pteridium aquilinum	18	127.20	1.75
Circaea alpina	19	114.00	1.57
Linnaea borealis	20	111.77	1.54
Mnium spinulosum	21	110.57	1.52
Dicentra formosa	22	107.91	1.49
Acer macrophyllum*	23	105.64	1.45
Chimaphila umbellata	24	103.91	1.43
Vaccinium parvifolium	25	93.77	1.29
Acer macrophyllum*	26	91.24	1.26

* Juvenile and adult
SSRANK/SSTOTAL= 0.811

It should be emphasized that although at first sight the clustering procedure appears to be divisive, it is actually an iterative, agglomerative technique.

Results

The method is illustrated using a data set of moderate size containing 73 species and 45 relevés, collected from three river terraces differing in soil and moisture content, located near Hope, B.C. (Fewster and Orlóci 1978). Species cover-abundance within relevés was estimated using the Braun-Blanquet scale and subsequently converted to an ordinal scale (van der Maarel 1979) for use in computation.

The program was run using the SWIGSS ranking algorithm with the cumulative percentage cutoff set to account for 80% of the total variation. In all, 19 rankings were performed in creating the dendrogram shown in Figure 2, with the ranking cutoff set at 4 relevés. However, only 5 tables outlining the ranked species used at the primary fusions indicated in Figure 2 are presented, as they are sufficient to illustrate the advantages of the method in elucidating the overall trends in the data.

It can be seen from the dendrogram that 3 distinct groups of relevés are formed, corresponding to the 3 river terraces which were sampled. The species ranked in Table 1 were those used in the first pass to divide the relevés into the 2 major subdivisions. A comparison of Tables 2 and 3 shows that many of the species in Table 1 are present, although few of the highly ranked species occur on both sides of the major split. Some species, such as *Mahonia nervosa*, were highly ranked on both sides of the dendrogram. It should be noted that many species change their ranking at different fusion levels within the dendrogram (e.g. *Tsuga heterophylla, Mnium insigne, Clintonia uniflora, Gaultheria shallon*). New species also became incorporated into the ranking as they became useful in determining the fusion, while other species dropped out as their importance within a branch decreased. A similar series of changes in the species lists occurred again in subsequent fusions as indicated by Tables 4 an 5.

It should be noted that the fusion level at any point will be lower using the ranked species as compared to the level obtained when all species are used. While several correction factors were explored, it was found that the fusion levels obtained using the reduced data set were satisfactory indicators of the importance of the fusion despite the fact that different species are used in calculating each fusion point. However, since the species were qualified by a consistent, objective ranking criterion, the portion of the total variance explained by those species should be comparable at all the fusion levels where a new ranking occurs. This can be verified by examining the SSRANK/SSTOTAL ratio, which indicates the degree to which the program has achieved this goal of consistency. In this data set the ratio was indeed highly stable at all fusions where rankings were performed.

By comparing the species in Tables 1 to 5, the differences in vegetative composition which separate the 3 terraces are elucidated as a result of changes in the ranking pattern. A dendrogram which used all the species throughout in classifying the relevés yielded a similar result, but was more susceptible to misclassification of relevés because of the large element of randomness incorporated into the solution by sporadically occurring species.

Table 2. Dichotomy Number 2 (see Fig. 2).

SPECIES	RANK	SUM OF SQUARES	PERCENTAGE VARIANCE ACCOUNTED FOR
Mnium insigne	1	152.92	10.91
Thuja plicata	2	134.85	9.62
Hylocomium splendens	3	107.21	7.64
Mnium spinulosum	4	96.92	6.91
Mahonia nervosa	5	89.21	6.36
Circaea alpina	6	69.71	4.97
Acer macrophyllum	7	67.21	4.79
Symphoricarpos alba	8	64.92	4.63
Eurhynchium oreganum	9	62.85	4.48
Acer macrophyllum	10	47.71	3.40
Lactuca canadensis	11	46.92	3.34
Disporum hookeri	12	45.71	3.26
Claytonia sibirica	13	33.71	2.40
Hylocomium sp.	14	32.92	2.34
Streptopus amplexifolius	15	29.42	2.09
Osmorhiza chilensis	16	26.35	1.88
Rhytidiadelphus loreus	17	26.35	1.88

SSRANK/SSTOTAL=0.809

Table 3. Dichotomy Number 3 (see Fig. 2).

SPECIES	RANK	SUM OF SQUARES	PERCENTAGE VARIANCE ACCOUNTED FOR
Tsuga heterophylla	1	308.96	8.70
Gaultheria shallon	2	270.96	7.63
Clintonia uniflora	3	203.93	5.74
Acer circinatum	4	178.77	5.03
Achlys triphylla	5	159.54	4.49
Mahonia nervosa	6	149.09	4.20
Pachistima myrsinites	7	147.54	4.15
Polystichum munitum	8	139.67	3.93
Lactuca canadensis	9	127.93	3.60
Pseudotsuga menziesii	10	119.93	3.37
Linnaea borealis	11	99.48	2.80
Pteridium aquilinum	12	97.93	2.75
Disporum hookeri	13	93.35	2.63
Chimaphila umbellata	14	89.41	2.51
Hylocomium splendens	15	88.77	2.50
Vaccinium parvifolium	16	81.48	2.29
Thuja plicata*	17	79.35	2.23
Hylocomium sp.	18	76.19	2.14
Vaccinium membranaceum	19	62.83	1.77
Thuja plicata*	20	61.93	1.74
Amelanchier alnifolia	21	49.41	1.39
Symphoricarpos albus	22	48.38	1.36
Lonicera ciliosa	23	46.38	1.30
Eurhynchium oreganum	24	45.67	1.28
Rhytidiadelphus loreus	25	45.35	1.27

* Adult and juvenile
SSRANK/SSTOTAL= 0.809

The method described in this paper has been applied to other real data sets of varying sizes with similar results. The largest data set run to date consisted of 200 species and 66 relevés, while another had 139 species and 118 relevés. It has been found that the ranking algorithm of Shaukat (1985) is very time efficient and thus can provide significant savings in analyzing large data sets with excellent recovery of the original group structure.

Discussion

The concept of the locality of importance is almost as old as ecology itself. We have found that the extension of this approach to subsets within the study area to be useful in itself, but, perhaps more importantly, the technique has pointed the way to a new view of species as indicators of pattern.

When this iterative technique was first developed, the argument was raised that any species occurring in the study area was intrinsically important. At the time, our reaction was that this might be so, but given realistic levels of sampling intensity and sampling precision, data from some species would only create «noise». We still believe this to be true, but would now go further and suggest that some species in some areas occur in an essentially random pattern, no matter how intensive the data recording effort might be.

There are familiar early examples of local importance: calcicole vs. calcifuge species, indicator species, and faithful species are all classic examples. It is reasonable to suppose, and indeed, it is borne out by our results, that the same phenomenon is true on an even more local level. The case of a species which is highly ranked on only one set of dichotomies of the dendrogram is easily understood in the above context. More interesting is a species which is revealed as being important in a number of major branches. Any area of vegetation is a multivariate mosaic of environmental factors. Hence, a given species, important in a number of disparate branches of a dendrogram might indicate only that it is restricted with respect to one environmental factor, but is otherwise quite undemanding of its environment. The species would, therefore, occur wherever the limiting factor existed, regardless of any other factor interactions which would typify a set of relevés. On the other hand, it may be that the species possesses a set of plastic phenotypic responses, different aspects of which play a decisive role in different subsets of the study area. We believe this to be a point of particular potential.

While the technique outlined represents an operational

Table 4. Dichotomy Number 4 (see Fig. 2).

SPECIES	RANK	SUM OF SQUARES	PERCENTAGE VARIANCE ACCOUNTED FOR
Mahonia nervosa	1	70.90	9.44
Hylocomium splendens	2	58.54	7.79
Thuja plicata*	3	58.18	7.75
Thuja plicata*	4	46.18	6.15
Acer circinatum	5	40.54	5.40
Linnaea borealis	6	36.00	4.79
Amelanchier alnifolia	7	28.72	3.82
Achlys triphylla	8	28.72	3.82
Vaccinium parvifolium	9	26.18	3.48
Vaccinium membranaceum	10	26.18	3.48
Pyrola grandiflora	11	24.54	3.26
Chimaphila umbellata	12	22.90	3.05
Lonicera ciliosa	13	22.72	3.02
Hylocomium sp.	14	22.54	3.00
Holodiscus discolor	15	19.63	2.61
Betula papyrifera	16	16.90	2.25
Rubus spectabilis	17	14.72	1.96
Corylus cornuta	18	13.63	1.81
Rhytidiadelphus loreus	19	12.90	1.71
Goodyera oblongifolia	20	12.54	1.67

* Adult and juvenile
SSRANK/SSTOTAL= 0.803

Table 5. Dichotomy Number 5 (see Fig. 2).

SPECIES	RANK	SUM OF SQUARES	PERCENTAGE VARIANCE ACCOUNTED FOR
Tsuga heterophylla	1	242.95	14.74
Polystichum munitum	2	102.55	6.22
Clintonia uniflora	3	91.80	5.57
Pseudotsuga menziesii	4	90.95	5.52
Acer circinatum	5	89.75	5.44
Lactuca canadensis	6	76.20	4.62
Gaultheria shallon	7	68.55	4.16
Achlys triphylla	8	60.20	3.65
Pachistima myrsinites	9	59.75	3.62
Hylocomium sp.	10	52.95	3.21
Disporum hookeri	11	49.00	2.97
Pteridium aquilinum	12	46.20	2.80
Eurhynchium oreganum	13	42.20	2.56
Symphoricarpos albus	14	39.80	2.41
Mnium insigne	15	28.55	1.73
Hylocomium splendens	16	28.55	1.73
Rhytidiadelphus loreus	17	28.55	1.73
Trillium ovatum	18	27.80	1.68
Betula papyrifera	19	27.75	1.68
Mahonia nervosa	20	26.80	1.62
Betula papyrifera	21	26.55	1.61
Galium triflorum	22	25.80	1.56

SSRANK/SSTOTAL= 0.809

approach which will, at least in part, enable ecologists to gain an understanding of the dynamics of the revealed pattern, it is also a key to the investigation of the way in which particular species play their role in pattern creation.

It is hoped that these results will call into question the practice of using a single species list for the analysis of pattern in a study area, whether arrived at by a single ranking procedure or a list representing the totality of practically recordable species from the area. Much more importantly, however, the technique may allow an investigation of the ways in which some species may play different pattern forming roles in different sites of the study area.

REFERENCES

FEWSTER, P. and L. ORLÓCI. 1978. Stereograms to aid group recognition and trend identification in vegetation data. Can. J. Bot. 56: 162-165.

JANCEY, R.C. 1979. Species ordering on a variance criterion. Vegetatio 39: 59-63.

JANCEY, R.C. 1980. The minimisation of random events in the search for group structure. Vegetatio 42: 99-101.

ORLÓCI, L. 1973. Ranking characters by a dispersion criterion. Nature 244: 371-373.

ORLÓCI, L. 1978. Multivariate analysis in vegetation research. 2nd ed. Junk, The Hague. 451 pp.

SHAUKAT, S.S. 1985. A technique for species weighting and its utility in data reduction and minimization of misclassification. (manuscript to be submitted).

VAN DER MAAREL, E. 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. Vegetatio 39: 97-144.

Manuscript received: January 1986