



## Multivariate Methods in Plant Ecology: I. Association-Analysis in Plant Communities

W. T. Williams; J. M. Lambert

*The Journal of Ecology*, Vol. 47, No. 1. (Mar., 1959), pp. 83-101.

Stable URL:

<http://links.jstor.org/sici?sici=0022-0477%28195903%2947%3A1%3C83%3AMMPEI%3E2.0.CO%3B2-6>

*The Journal of Ecology* is currently published by British Ecological Society.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/briteco.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# MULTIVARIATE METHODS IN PLANT ECOLOGY

## I. ASSOCIATION-ANALYSIS IN PLANT COMMUNITIES

BY W. T. WILLIAMS AND J. M. LAMBERT

*Botany Department, University of Southampton*

*(With two Figures in the Text)*

### CONTENTS

	PAGE
I. INTRODUCTION ... ..	83
A. General ... ..	83
B. Terminology, symbols and definitions ... ..	84
II. THEORETICAL CONSIDERATIONS ... ..	84
A. The homogeneous grouping ... ..	84
B. The problem of subdivision ... ..	85
III. COMMUNITY ANALYSES ... ..	87
A. Introduction ... ..	87
B. Methods ... ..	88
C. Results ... ..	88
1. A 5-species community ('Beaulieu Road') ... ..	88
2. A 10-species community ('Matley Ridge') ... ..	93
IV. DISCUSSION ... ..	99
SUMMARY ... ..	101
ACKNOWLEDGMENTS ... ..	101
REFERENCES ... ..	101

## I. INTRODUCTION

### A. General

The valuable pioneer work by Goodall (1953) on the use of interspecific associations for sorting quadrats into groups is based on his definition of a homogeneous unit of vegetation as one in which all species-associations are indeterminate or non-significant. The method he eventually recommends, from empirical studies on data from the Australian Mallee, is to sort on the most abundant species involved in positive associations, pooling the *residuum* at each stage. Since statistical methods of this kind, however, inevitably require much large-scale computation, it is necessary both to examine the statistical foundations of any method proposed and to assess whether the ecological information obtained in fact justifies the time and labour involved. Furthermore, the analysis of complex communities can normally only be brought within the reach of the practising ecologist if the method is programmed for a digital computer, and this requirement must be borne continuously in mind if prohibitive expense in computing time is to be avoided.

It is implicit in the use of any sorting method that it is expected to reveal an underlying structure simpler than the raw matrix of associations. This is equivalent to assuming that the associations are not all independent; but if this is the case, the system is multivariate, and must be studied as such. With this in mind, we shall show, on theoretical grounds and by reference to two selected heathland communities, that a method of subdivision more effective than Goodall's can be devised: and we shall then attempt to assess the type of ecological information thereby obtained.

B. *Terminology, symbols and definitions*

(i) *Association* will be used throughout in its statistical sense. It is the traditional term where attributes rather than quantitative measurements are concerned, and we prefer to reserve the term 'correlation' (employed by Goodall) for use in its normally accepted contexts. The risk of confusion with phytosociological 'associations' is slight in the present paper.

(ii) *Community* is used as a convenient neutral term to denote any set of species growing together, without implying a particular statistical or ecological status.

(iii) Individual species will be represented by letters, capitals for presence and lower case for absence.

(iv) Class-symbols will be traditional—*e.g.* (abD) is the class defined by the absence of A and B, the presence of D, and irrespective of the presence or absence of C, E, . . . etc. The complete population is denoted by (N).

(v) Classes defined by one—(B), two—(Ce), . . . etc., letters will be said to be of the first, second, . . . etc., order.

(vi) We shall use  $\Sigma\chi^2$  to denote the sum of all the significant  $\chi^2$  values associated with a particular species in the class under study, both positive and negative associations being included.

## II. THEORETICAL CONSIDERATIONS

A. *The homogeneous grouping*

Goodall's definition of a 'homogeneous grouping' rests on two quite independent concepts: associations may be absent either by being indeterminate—either or both species being present or absent in all quadrats—or by falling below significance. Such groupings can be sought only in presence-or-absence data; it is easily shown theoretically—and Goodall has shown empirically—that a true correlation, based on the *quantities* of the species concerned, cannot in general be removed by subdivision.

Three aspects of the basic definition require examination:

(i) *The use of negative associations.* Goodall rejects negative associations. Admittedly they can be treacherous; if the quadrat size is too small they are apt to appear simply by virtue of the fact that space occupied by one plant cannot be occupied by another. Nevertheless, their presence may greatly strengthen the analysis, for in a large and obviously heterogeneous area—a dry heath bordering a bog, or a wood bordering a pasture—the main subdivision is likely to be reinforced by strong negative associations. We therefore recommend reverting to the original suggestion of Tuomikoski (1942) and regarding positive and negative associations as of equal importance.

(ii) *Homogeneity of a group of associations.* Both Goodall (1953) and Greig-Smith (1957) suggest that at the  $P = 0.05$  level of significance 1 in 20 species-pairs may be expected to reach significance, and that a population can be regarded as homogeneous if this proportion is not exceeded. However, we have already pointed out (Sect. 1 A) that the method tacitly assumes that the associations are not independent; and in this case Goodall's criterion is equivalent to a test of the significance of an entire correlation matrix. Although many—mostly empirical—tests have been proposed for this purpose, no simple valid test concerning the number

of significant entries is in fact known. We therefore prefer at this stage to remove *all* significant associations.

(iii) *Relative importance of the basic criteria.* The alternative criteria—indeterminacy or non-significance—by which a Goodall grouping is defined are not of equal value. A non-significant association has been tested and found wanting; an indeterminate association has not been tested, the circumstances being such that no association can be manifested. We shall therefore take the attitude that, if a choice has to be made, it is more important to reduce the level of associations than to render them indeterminate.

### B. *The problem of subdivision*

(i) *The requirement.* The basic problem is to subdivide a population so that all associations disappear; but there will in general be a large number of alternative subdivisions fulfilling this requirement. We therefore propose the concept of *efficient* subdivision, by which we intend subdivision on that species which, in the two subclasses resulting, produces the smallest total number of residual significant associations. This species can, of course, always be identified by examining every possible means of subdivision at each stage; but, even with a computer, this would be an unacceptably lengthy operation were more than a few species involved. Our immediate problem, therefore, is to find a parameter which, *when applied to the class to be subdivided*, is most likely to bring about efficient subdivision.

(ii) *The available parameters.* The definition of the final groupings sought is such that a test of significance is required. The obvious test for small-scale computation would be that which Goodall uses:  $\chi^2$  (with Yates's correction) where the cell-frequencies are large enough, and Fisher's 'accurate method' where they are small, or where the  $\chi^2$  is marginally significant. However, if the method of subdivision is to be capable of translation into a relatively simple programme for a digital computer, a single test is desirable; and since when the cell-frequencies are large Fisher's method is extremely lengthy, and moreover involves high factorials which would greatly complicate the programme, we propose the use of  $\chi^2$  with Yates's correction for *all* tests. Fortunately, the Yates's correction is known to over-correct, so that it is most unlikely that we shall obtain spurious associations; and in this case we are prepared to sacrifice marginally significant associations to the overriding need for simplicity of analysis.

If we wish to reduce the level of association an actual *measure* of association is required; and the parameter selected for subdivision is inherently likely to be related to the parameter selected for measurement. (It is for this reason that the *abundance* of a single species (cf. Goodall), which is not in itself an associative property, cannot fulfil our requirements for subdivision.) Consider, therefore, the  $s \times s$  matrix whose entries are our selected association-index, say  $I$ , taking the values in the principal diagonal to be invariant. Consider, too, the column sums, which (neglecting the principal diagonal) will represent, for each species, its  $\Sigma I$  value. Now if, in a search for a postulated simpler structure, we subjected this matrix to a factorial analysis, the species with the highest  $\Sigma I$  would necessarily have the highest loading on the first centroid axis, which is itself an approximation to the first principal axis. If, therefore, the associations are due to underlying factors, division on the species with the largest  $\Sigma I$  is likely to produce the greatest

possible discontinuity in the first common factor; it will tend to reduce the residual  $\Sigma I$ —*i.e.* the residual general level of association—to a minimum. Since this is precisely what we require to do,  $\Sigma I$  may be expected to be the parameter we need.

There remains the choice of the index  $I$ . Since we have in any case decided to calculate  $\chi^2$  as a test of significance, it is natural to inquire whether it cannot also serve as an index of association. It is not normally used as such, since the degree of association (but not the significance) which it represents depends on the size of the population in which it is measured; but this is not important in our case, since the values are always compared within a single population or class. The fact that it can serve this double purpose gives it an overwhelming advantage for hand computation over the several alternative indices available. However, Yates's correction is not applicable to  $\chi^2$  when this is used as an index of association, and it would be theoretically preferable to construct  $\Sigma\chi^2$  from uncorrected values, using the corrected values solely as a test of homogeneity. This refinement would, however, effectively put the problem out of reach of hand computation for any but the simplest communities; and we shall in this paper necessarily concentrate on the results obtainable from corrected  $\Sigma\chi^2$  values. It would, of course, be simpler still to select the single greatest  $\chi^2$  value, but this would tend to emphasize trivial features at the expense of the population as a whole. Similarly, it would be simpler to select the species with the greatest *number* of significant associations, which would be equivalent to rendering the greatest possible number of associations indeterminate; but we have already pointed out (Sect. II A (iii)) that this is less fundamental than reducing the level of association.

(iii) *Ambiguity.* Whatever parameter is in use, indeterminate values must be taken as zeros, the species concerned being excluded from the analysis; and if corrected  $\chi^2$  is used, non-significant values must also be taken as zeros. Any path of subdivision may thus terminate in a choice between two or more species of equal importance—for instance, in a class with a single non-zero  $\chi^2$ . Such ambiguities will be rare with uncorrected  $\chi^2$ , since they are then only likely to occur in classes with a high level of indeterminacy. We propose here to resolve all ambiguities by reference to the next highest class in which discrimination is possible.

(iv) *Application of the parameter.* Imagine a population divided on species X into (X) and (x). (X) is still found to be divisible on species Y into (XY) and (Xy). Goodall would 'pool' (Xy) with (x) to form a new population. Our chief objection to this system is that information relating to the discontinuity (X)/(x) is being discarded, and on these grounds we would prefer a 'hierarchical' system—a division, once made, would remain inviolate throughout the analysis. It is conceivable that, from the point of view of the investigation as a whole, the information thus discarded would be better discarded, and an empirical test on actual populations is desirable.

There are, however, other objections to 'pooling'. First, the final groups produced are not necessarily capable of simple statistical definition in terms of presence or absence of key species; secondly, the route by which these groups are obtained is not in itself meaningful, so that *only* the final groups, and not the successive subdivisions which have produced them, are available for examination; and thirdly, the computation is inevitably much longer, since large 'pool' populations are examined at intervals throughout the analysis. With a powerful associative criterion such as  $\Sigma\chi^2$ , it is unlikely that the matter will be important; for

pooling is inherently likely to recreate the intense associations which have previously been removed, so that the pooled group is statistically rejected and the hierarchical pattern retained.

A disadvantage of the use of corrected  $\Sigma\chi^2$  values for hierarchical subdivision is that the path of subdivision may change with significance level; with uncorrected values only the degree of subdivision will change. We shall accept this inconvenience for the purpose of the present paper, and must therefore select a 'standard' level of significance for subdivision. We have chosen to work at  $P = 0.05$ ; only experience on actual populations can show whether this results in an excessive fragmentation of the population.

(v) *Recombination*. Goodall tests his final groupings for their ability to combine without recreating significant associations; but, as with pooling, combination of groups across a hierarchical division is inherently likely to recreate such associations when an associative parameter is used. Recombination might, however, occur in two rather special cases. First, if the subdivision has produced any very small groups—of, say, less than about 8 quadrats—these may well recombine indiscriminately, since the total number of quadrats resulting is too small for any but the most intense associations to be detected by the crude significance test we are using. Such recombinations—and, for that matter, the individual groups themselves—are statistically meaningless, and are merely an indication that the significance level selected is too low. Secondly, recombination may occur if many of the possible associations are indeterminate. Such cases are likely to be rare, but would be statistically meaningful; they represent cases in which the criterion of indeterminacy can override that of non-significance. Nevertheless, since (Sect. II A (iii)) we believe indeterminacy to be of less importance than lack of significance, the final groups so obtained are likely to be less informative than the primary divisions. In any case, a simple recombination test provides little new information; failure to recombine tells us only that two classes are different, and nothing of the extent to which they are related. Methods can be devised for the quantitative comparison of such communities, and we shall return to these in a later paper. Although we shall test our final groupings for recombination here, it would in any case be quite impracticable to incorporate a general test in a computer programme; for  $r$  final groupings the scanning of all  $\frac{1}{2}r(r-1)$  possibilities would lengthen the process to an intolerable extent if  $r$  were large.

### III. COMMUNITY ANALYSES

#### A. *Introduction*

In the experimental section which follows we are testing two independent hypotheses. The first of these is purely statistical: it is that division on the species with the highest  $\Sigma\chi^2$  will subdivide the population in the most efficient manner possible. Theoretical considerations, it will be recalled, tell us that it is *likely* to do so, not that it necessarily will.

The second hypothesis is that the resulting subdivision will provide information of ecological value. It is impossible to examine all possible subdivisions and compare them with a known ecological situation: first, because it is difficult to conceive of a *completely* known ecological situation; and secondly, because the

total number of such subdivisions is prohibitively great.\* We can only divide on a selected statistical basis and ask, 'what ecological information has this division provided?' and thereby try to assess the efficacy of the method.

### B. Methods

(i) *Selection of areas.* The general problem can most easily be investigated in a community with sufficiently few species to allow of a full analysis of all associations in all possible classes within the population.† The test-communities chosen were two areas of New Forest heathland in the proposed Denny Reserve. The first, containing a number of well-defined burnt strips of different ages, not only provided a very simple population for complete analysis, but also promised a reasonably clear-cut ecological comparison of results from different sorting methods; while the second, more complex yet sufficiently similar to permit cross-comparison, gave greater opportunities for investigating the potentialities of the new method.

(ii) *Sampling methods.* For a discussion of random *versus* systematic sampling, see Greig-Smith (1957). Since we were concerned only with presence-or-absence data and therefore required no estimates of density, and since we were concerned with the pattern of an area as a whole, we chose to sample systematically by means of a rectangular grid. The unlikely possibility that our grid might 'resonate' with the vegetation is one we were prepared to accept, in view of its immense practical advantages in setting out, mapping and later revisiting the sample sites. The actual sampling was carried out by M. J. Hudson and P. F. Hunt, both students of the Southampton Botany Department.

(iii) *Species considered.* To reduce the species to manageable numbers for hand computation, only vascular plants occurring in more than 2% of the quadrats in either or both of the communities are included in the analyses. They are represented by the following letters:

A, *Calluna vulgaris* (L.) Hull; B, *Molinia caerulea* (L.) Moench; C, *Erica tetralix* L.; D, *E. cinerea* L.; E, *Pteridium aquilinum* (L.) Kuhn; F, *Trichophorum caespitosum* (L.) Hartman; G, *Ulex europaeus* L.; H, *U. minor* Roth; J, *Festuca ovina* L.; K, *Potentilla erecta* (L.) Rausch.; L, *Polygala serpyllifolia* Hose; M, *Carex panicea* L.; P, *C. binervis* Sm.

### C. Results

#### 1. A 5-species community ('Beaulieu Road')

(i) *The area.* A Callunetum in the New Forest, E. of Beaulieu Road station and N. of Bishop's Purlieu (map ref. SU/349054). It lies for the most part directly over Barton Sand, though towards the S. end a slight ridge is covered with a cap of Plateau Gravel, tailing out northwards with some downwash. The soil on both sand and gravel is a well-developed podsol. The area is subjected to an approximately 7-year rotational burning, and is further frequently affected by casual burning through sparks from passing trains; local report and counts from *Calluna*

\* The number of ways in which a population containing  $p$  final classes can be divided into groups is a well-known problem in combinatorial analysis, and no explicit general solution is known. For  $p = 4$  (2 spp.) and  $p = 8$  (3 spp.) it is easily shown by enumeration that the answers are 15 and 4140 respectively.

† For  $s$  species this involves the examination of  $\frac{1}{2}s(s-1)$  associations in  $3^{s-2}$  classes selected from the  $3^s - 2^{s-1}(s+2)$  classes of order  $(s-2)$ .

Table 1.  $\chi^2$  matrix for Beaulieu Road 5-species community

Significant associations are entered as their corrected  $\chi^2$  values, in normal type if the association is positive, in italics if it is negative. Non-significant values are denoted by x, and are treated as zeros for the computation of  $\Sigma\chi^2$ .

	A	B	C	D	E
A	...	51.31	45.66	x	x
B	51.31	...	93.76	12.62	68.64
C	45.66	93.76	...	4.84	14.08
D	x	12.62	4.84	...	6.92
E	x	68.64	14.08	6.92	...
$\Sigma\chi^2$	96.97	226.33	158.34	24.38	89.64

rings suggested that the last deliberate burning of the area as a whole had taken place 4½ years before the survey, but that certain strips had later been accidentally burnt between 1 and 2 years ago and others during the spring of the current year.

The area was sampled in July 1957 by a rectangular grid of 44 × 14 1-metre quadrats spaced 5 m apart in both directions. One record was lost in transferring the data to cards, and (N) is thus effectively 615.

(ii) *The population.* Six vascular species were actually present, A to F in our key. However, F was represented in less than 2% of the quadrats and has been ignored. Of the 32 possible final classes, only the following 12 exist:

$$\begin{aligned}
 (ABCDe) &= 6 & (AbCde) &= 9 & (Abcde) &= 69 \\
 (ABCde) &= 244 & (AbcDE) &= 4 & (aBcde) &= 23 \\
 (ABcDe) &= 8 & (AbcDe) &= 11 & (abcdE) &= 5 \\
 (ABcde) &= 193 & (AbcdE) &= 12 & (abcde) &= 31
 \end{aligned}$$

Of the 131 possible classes of order 3 in which associations are to be sought, 26 are empty; as a result, a further 47 vanish, having become identical with classes of higher order; 3 more are excluded as being below the minimum size (8) in which, using Yates's correction, significant associations can be manifested; and 55 remain.

(iii) *Association-analysis.* Although we have examined all possible associations in all possible classes in this community, we here present only (in Table 1) the  $\chi^2$  matrix for (N), and (in Table 2) the  $\chi^2$  values within the first-order classes. The

Table 2. First-order classes in Beaulieu Road 5-species community

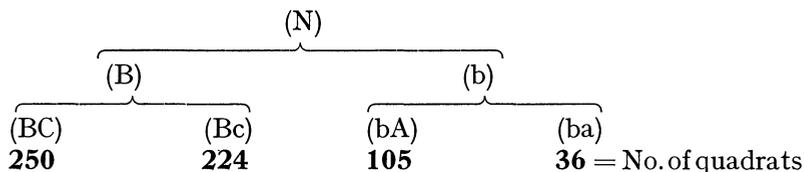
The first column gives the class under examination, the second the number of quadrats in that class. In the association-columns, - denotes that the association is necessarily indeterminate by virtue of the species whose presence or absence defines the class; o denotes that the association is indeterminate in this population; and x denotes that the association fails to reach significance. Significant associations are entered as their corrected  $\chi^2$  values, in normal type if the association is positive, in italics if it is negative.

		{A, B}	{A, C}	{A, D}	{A, E}	{B, C}	{B, D}	{B, E}	{C, D}	{C, E}	{D, E}
(A)	556	-	-	-	-	73.29	19.34	65.41	7.18	12.50	9.25
(a)	59	-	-	-	-	o	o	x	o	o	o
(B)	474	-	24.8	x	o	-	-	-	x	o	o
(b)	141	-	x	4.34	x	-	-	-	x	x	x
(C)	259	o	-	o	o	-	x	o	-	-	o
(c)	356	16.16	-	x	x	-	25.63	35.06	-	-	3.85
(D)	29	o	o	-	o	5.63	-	x	-	x	-
(d)	586	58.12	47.91	-	5.20	83.08	-	59.22	-	11.55	-
(E)	21	o	o	x	-	o	o	-	o	-	-
(e)	594	48.50	43.99	x	-	77.88	7.69	-	x	-	-

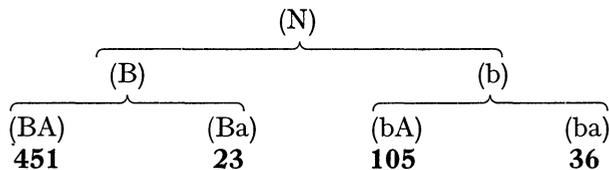
associations in (N) show that the species can be roughly divided into two groups—A, B and C on the one hand, D and E on the other; there are positive associations within each group, negative associations between them. Some of these associations are very strong; {A, B}, {B, C}, {B, D} and {B, E} exist at almost every class-level, and we note that all four involve B. Furthermore, in the second- and third-order classes, the existence of the powerful association {A, C} is manifested only in the presence of B. This suggests that the first division should be on B, which is strikingly confirmed by an examination of the first-order classes. Division on any other species produces two classes, one of which is almost entirely free from associations (though mainly by indeterminacy), the other retaining most or all associations unimpaired. If the population is divided on B, however, all negative associations disappear, and we are left with a single positive association in each class, one of which ({A, D} in (b)) is only marginally significant.

B is therefore the species which our decision-parameter is required to select. Reference to Table 1 shows that B is the species with the highest  $\Sigma\chi^2$ ; this parameter has therefore here proved successful. On the other hand, Goodall's criterion—the most abundant species—would have required the population to be divided on A.

We may now continue the subdivision. An ambiguity (A or C) arises in (B); reference to the  $\Sigma\chi^2$  values in the next highest class—(N)—requires division on C. The ambiguity (A or D) in (b) is similarly resolved by division on A. All classes are now homogeneous, and we find that they cannot in fact be recombined without recreating significant associations. The complete hierarchy is therefore as follows:



Division on *uncorrected*  $\Sigma\chi^2$  would have given the following:



(Ba) and (ba) could now be recombined across the hierarchy to produce the homogeneous first-order class (a). Table 1 shows that division of the original population into (A) and (a) is one of the three possible ways of making the greatest possible number of associations indeterminate at the first stage; and that this is a case of true indeterminacy-recombination (Sect. II B (v)) is shown by considering the effect of raising the significance level to  $P = 0.01$ . (b) now fails to subdivide, and the three remaining groups—(BA), (Ba) and (b)—cannot be recombined; recombination is possible only if we can recreate the almost completely indeterminate group (a).

Goodall's method—abundance and pooling—would in this simple case also give the three groupings (AB), (Ab) and (a), though without, of course, any information as to their relative importance.

(iv) *Ecological significance.* Fig. 1 shows the distribution of the corrected  $\Sigma\chi^2$  groupings in relation to the burning lines; Table 3 gives their full species-composition. A valuable feature of a hierarchical system is that its successive subdivisions can be identified on a map at successive levels of importance. Here, the first-order (B)/(b) division forms a well-marked '*Molinia* extinction line' which divides the area cleanly into two; but although it coincides almost exactly with a burning line, the presence of both young and old communities on either side suggests that the burning here has merely sharpened the effect of a more powerful underlying factor. Contouring of the site (Fig. 1—inset) and examination of a limited number of soil pits later established that the *Molinia* boundary is roughly associated with the beginning of the Plateau Gravel ridge. In essence, therefore, this primary division of the population reflects the basic features of topography and soil, rather than the more conspicuous but superficial burning regime.

By contrast, the second-order (final) divisions directly follow the burning pattern. Within the *Molinia* zone, (BC) covers all areas which have not recently been burnt;

Table 3. *Species composition of final groupings at Beaulieu Road*

Ref. No. on map	1	2	3	4
Grouping	(BC)	(Bc)	(bA)	(ba)
No. of quadrats	250	224	105	36
A = <i>Calluna vulgaris</i>	250	201	105	0
B = <i>Molinia caerulea</i>	250	224	0	0
C = <i>Erica tetralix</i>	250	0	9	0
D = <i>E. cinerea</i>	6	8	15	0
E = <i>Pteridium aquilinum</i>	0	0	16	5

it is a fairly closed community consisting almost entirely of *Calluna*, *Molinia* and *Erica tetralix*. Areas accidentally fired during the current year bear only (Bc); *Molinia* has regenerated throughout and *Calluna* is rapidly following—so rapidly that, even during the time that this area was being recorded, a number of quadrats changed from (aBc) to (ABc). *Erica* clearly takes longer to re-establish itself, though the mixture of (BC) and (Bc) over the 1½-year-old areas suggests that it follows *Calluna* fairly soon; it had in fact begun to regenerate from stools even in the youngest areas when the site was revisited 3 months after the survey. South of the *Molinia* boundary, the division between (Ab) and (ab) is again related to a burning line. (Ab) is a 4½-year-old *Calluna* strip with a little *Pteridium*, *Erica cinerea* and *E. tetralix* admixed; (ab) is recently burnt bare ground, with sparse *Pteridium* at one end, but with *Calluna* only regenerating after the survey was complete. The division between these two is clearly, from its significance level, of very minor importance, and relates to the poor degree of association within the parent population (b). In fact, their floristic separation at the time of the survey is entirely dependent on the delayed regeneration of *Calluna* in (ab) compared with its performance in the other recently burnt areas; and this is probably correlated with the position of the (ab) strip on the dry south slope of the gravel ridge.

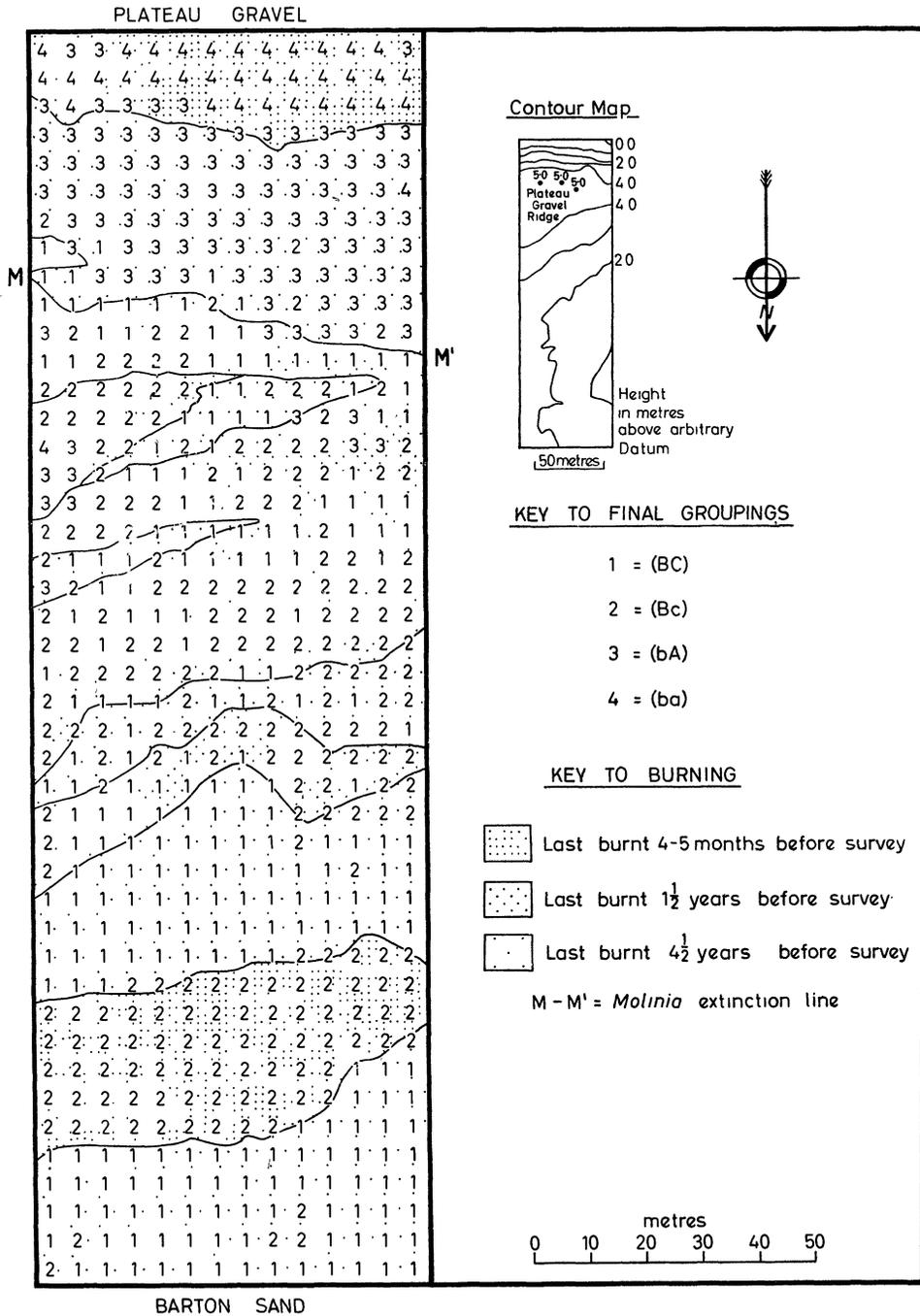


Fig. 1. Map of final groupings at Beaulieu Road.

The four final groupings (BA), (Ba), (bA) and (ba) obtained by the hierarchical use of uncorrected  $\Sigma\chi^2$  without recombination differ only from those set out above by the division of (B) at an earlier phase of the burn succession, *i.e.* the re-establishment of *Calluna* rather than *Erica tetralix* becomes the discriminating feature. Since the time lag between the entry of these two species is very short, the exact course of this secondary division relates only to very minor differences in phase, and there is little to mark either method of division as more important than the other. Subsequent recombination of (aB) and (ab) into the highly indeterminate (a) combines 23 scattered quadrats still lacking *Calluna* from the most recently burnt strips on the Barton Sand with those of the bare area on the south side of the Plateau Gravel ridge. Since the immediate effect of burning is simply to remove species without replacing them by others, the actual number of species is reduced in areas where this second (burning) factor is strong; the sub-population consequently exhibits a level of indeterminacy sufficiently high to override the significance barrier of the first (soil) factor, though the latter is still evident at the earlier stage of the analysis.

A classification directly into (AB), (Ab) and (a) on Goodall's abundance/pooling method is ecologically less convincing and revealing: a major division is made in relation to a markedly evanescent phase of the burn succession and the underlying soil relationships are largely obscured.

## 2. A 10-species community ('Matley Ridge')

(i) *The area.* A *Calluna* heath essentially similar to the foregoing, about 2 miles (3.2 km) S.E. of Lyndhurst on the N. slope of Matley Ridge (map ref. SU/325075). The subsoil is also similar, with Plateau Gravel at the top (S.) end, some gravel downwash, and Barton Sand at the bottom. An old cart-track, intercepting three other minor tracks at right-angles, runs down a central gully from S. to N., and the whole area is much crossed by ruts made by military vehicles in World War II. The slope is also longitudinally divided by a well-marked burning line: the west side bears a fairly open community, known to have been burnt 2½ years before the survey; the east a conspicuously older one which, though much disturbed during the war and not immune from accidental local fires, had not been rotationally burnt for many years.

The area was sampled in November 1957, using a more widely spaced grid of  $36 \times 11$  1-metre quadrats 10 m apart, and (N) is thus 396.

(ii) *The population.* Of the 17 vascular plants which were actually recorded, 4 were present in too few quadrats for study. Of the remainder (A-D and F-P in our key), A and B are statistically inactive in that they each occur in every quadrat but two; they are therefore omitted from the analysis, and the final population to be considered thus consists of 10 statistically effective species with a background of A and B.

Only 63 of the 1024 possible final classes actually exist; and, of these, many are represented by a single quadrat.

(iii) *Association-analysis.* We present the  $\chi^2$  matrix for (N) in Table 4. Once more, the species tend to divide into two groups: C and F, though not mutually associated, are separated from the remaining 8 species by negative associations. It is clear, however, that the main interest in the subdivision of this population will lie in the extent to which it reveals any general structure underlying the positive

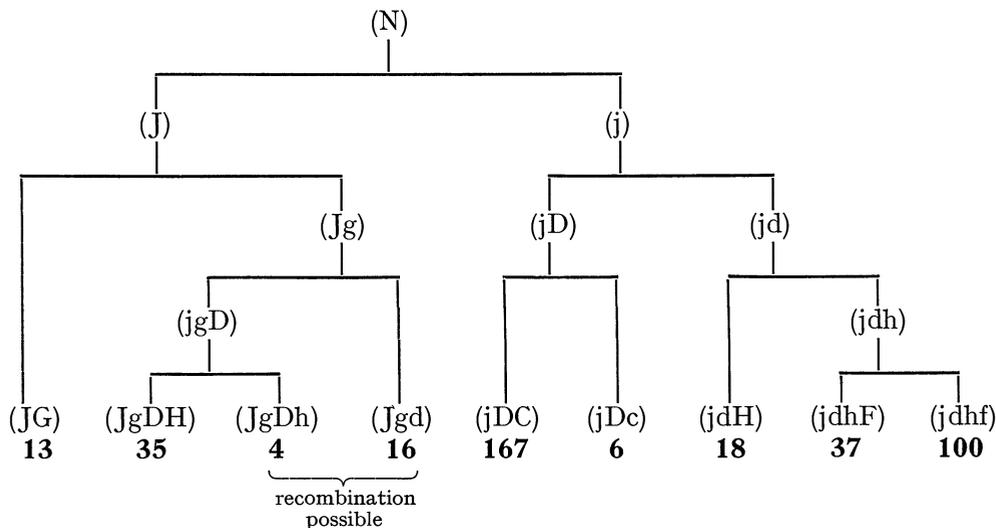
Table 4.  $\chi^2$  matrix and residual 1st-order associations for Matley Ridge 10-species community

Significant associations are entered as their corrected  $\chi^2$  values, in normal type if the association is positive, in italics if it is negative. Non-significant values are denoted by x, and are treated as zeros for the computation of  $\Sigma\chi^2$ .

The last three rows of the table represent the *number* of residual associations at the significance level indicated in both 1st-order classes (taken together) obtained by division on the species heading each column.

	C	D	F	G	H	J	K	L	M	P
C	...	x	x	<i>40.37</i>	x	<i>13.52</i>	<i>23.84</i>	x	x	<i>34.83</i>
D	x	...	<i>18.43</i>	x	39.37	x	x	8.48	x	x
F	x	<i>18.43</i>	...	x	x	x	x	x	x	x
G	<i>40.37</i>	x	x	...	x	36.23	48.62	x	4.23	13.99
H	x	39.37	x	x	...	51.94	x	11.70	x	5.23
J	<i>13.52</i>	x	x	36.23	51.94	...	53.84	4.61	31.82	43.07
K	<i>23.84</i>	x	x	48.62	x	53.84	..	x	33.21	23.84
L	x	8.48	x	x	11.70	4.61	x	...	x	x
M	x	x	x	4.23	x	31.82	33.21	x	...	15.15
P	<i>34.83</i>	x	x	13.99	5.23	43.07	23.84	x	15.15	...
$\Sigma\chi^2$	112.56	66.28	18.43	143.44	108.24	235.03	183.35	24.79	84.41	136.11
	Number of residual associations									
P = 0.05	18	22	25	16	21	15	16	30	18	19
P = 0.01	13	20	17	13	19	8	8	26	14	11
P = 0.001	11	18	13	9	13	4	5	24	10	9

associations within the 8-species group. To avoid presenting a  $45 \times 20$  table, we have included in Table 4 information as to the total *number* of significant associations remaining in both first-order classes when the population is subdivided on each species in turn. At the 0.05 level, 3 species (G, J and K) are almost equal with 16, 15 and 16 residual associations each respectively; at the 0.01 level the choice is reduced to J or K, with 8 each; and at the 0.001 level, J has 4 to K's 5. The level of the residual associations after division on J, moreover, lies slightly below that of those from K. There is no doubt, therefore, that the required first division is on J; and we note with satisfaction that J has the highest  $\Sigma\chi^2$  value. (The most abundant species, C, would clearly produce a relatively inefficient division.) Continued hierarchical division using corrected  $\Sigma\chi^2$  at the  $P = 0.05$  level produces the following final subdivision:



An ambiguity in (Jg)—D or H—is resolved by reference to (J); a second in (Jgd)—H or L—by reference to (Jg); a third in (jd)—H or L—by reference to (j); and a fourth in (jdh)—F or M—remains unresolved in (jd) and is resolved in favour of F by reference to (j). Of all the final groupings, only two—(JgDh) and (Jgd)—can be recombined; and, since (JgDh) is represented by only 4 quadrats, this recombination is without statistical meaning (Sect. II B (v)). It will be noted that two of the groupings contain less than 8 quadrats, and three more are close to the size (15) at which not even perfect associations could be detected at the  $P = 0.001$  level of significance; this fineness of subdivision suggests that the significance level chosen was too low.

This population proves, incidentally, to be one in which the hierarchy changes with significance level; at the 0.01 level (J) divides on C, though division on G is re-established (via a C/G ambiguity) at the 0.001 level. We shall confine our attention to the 0.05 level result given above.

Table 5. *Species composition of final groupings at Matley Ridge*

Ref. No. on map	1'	2'	3'	4'	5'	6'	7'	8'
Grouping	(JG)	(JgDH)	(Jgd) + (JgDh)	(jDC)	(jDc)	(jdH)	(jdhF)	(jdhf)
No. of quadrats	13	35	20	167	6	18	37	100
A = <i>Calluna vulgaris</i>	11	35	20	167	6	18	37	100
B = <i>Molinia caerulea</i>	13	35	20	165	6	18	37	100
C = <i>Erica tetralix</i>	7	33	20	167	0	18	37	100
D = <i>E. cinerea</i>	4	35	4	167	6	0	0	0
F = <i>Trichophorum caespitosum</i>	2	4	3	17	0	6	37	0
G = <i>Ulex europaeus</i>	13	0	0	3	0	0	0	2
H = <i>U. minor</i>	5	35	8	58	5	18	0	0
J = <i>Festuca ovina</i>	13	35	20	0	0	0	0	0
K = <i>Potentilla erecta</i>	6	4	2	0	0	0	0	0
L = <i>Polygala serpyllifolia</i>	8	25	8	89	0	14	11	34
M = <i>Carex panicea</i>	3	3	6	1	0	0	4	0
P = <i>C. binervis</i>	4	7	1	2	0	0	0	0

The greater number of species involved at Matley compared with Beaulieu Road made it impracticable to analyse this community by more than one method without the aid of a computer. The following assessment of ecological results from the analysis above must therefore omit any immediate comparison with results from the use of uncorrected  $\Sigma\chi^2$ .

(iv) *Ecological significance.* The map of the final groupings and their species lists are given in Fig. 2 and Table 5 respectively. The mosaic of communities on the map partly reflects the patchy disturbance of the ground, and its ecological interpretation is clearly less easy than in the almost diagrammatic Beaulieu Road community. Nevertheless, an underlying pattern can be seen, which was by no means obvious when the area was surveyed.

The 68 quadrats separated as (J) (*Festuca ovina*) at the important first dichotomy are ringed on the map for clarity. They are obviously concentrated into two main clusters, both lying athwart the  $2\frac{1}{2}$ -year burning line. One occurs near the top of the slope, with extensions along the line of the old cart-track to end in a

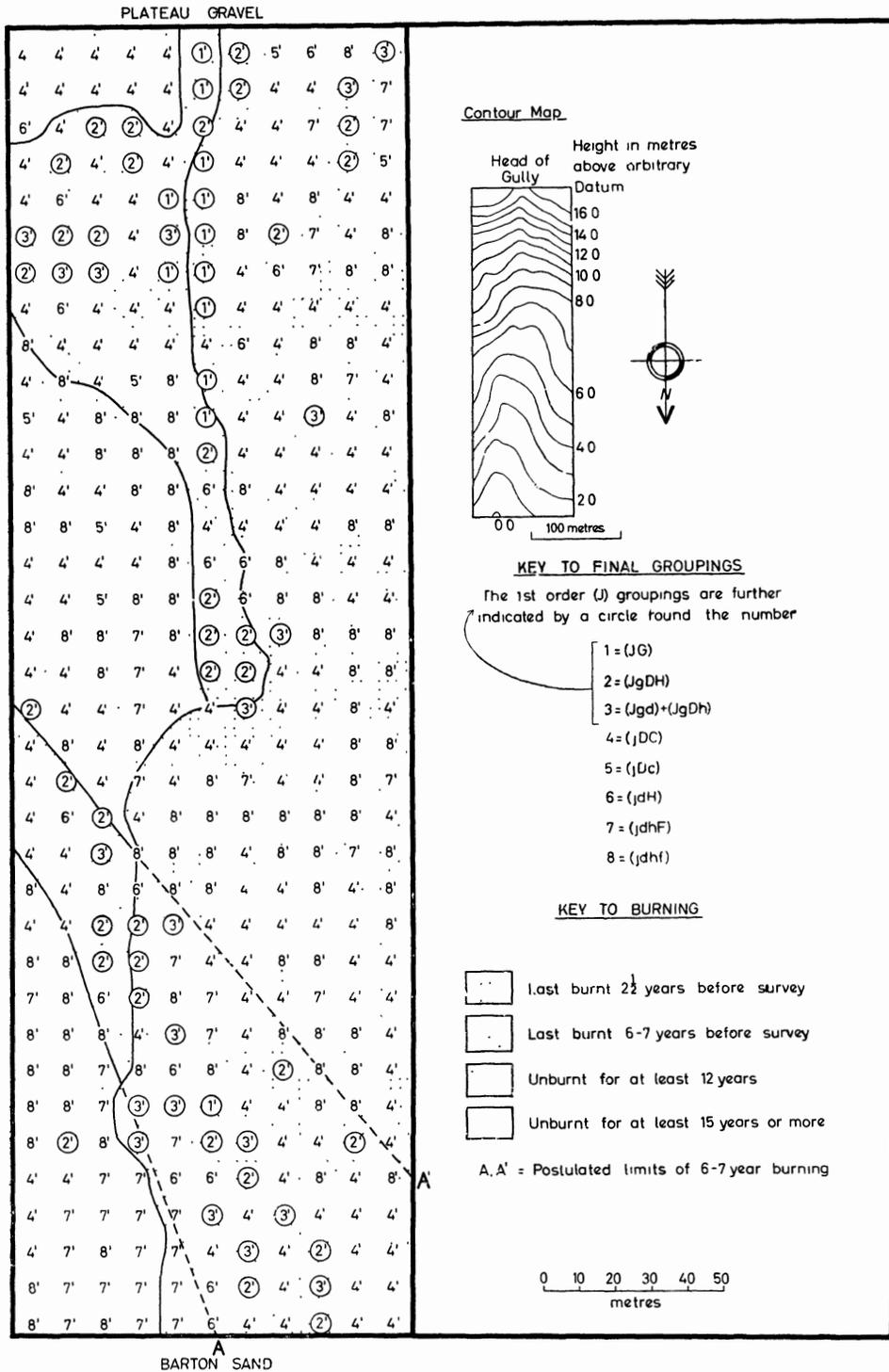


Fig. 2. Map of final groupings at Matley Ridge.

small outlier further down the gully; the other crosses the lower part of the area obliquely in a widening S.E./N.W. strip. Contouring of the site (Fig. 2—inset) and a few exploratory soil pits showed that neither bears any consistent relationship with an obvious soil or topographical feature, apart from the local focus on the gully; and the real key to the distribution of (J) was only revealed when the burning history of the area was later reconstructed in detail from counts of *Calluna* rings. Except for a very old pocket of 15-20 (-25)-year-old plants at the wettest end, most of the Callunetum on the east side of the area was roughly 12-13 years old, and clearly indicated recolonization after the war-time disturbance ceased. But two fairly large patches of 6/7-year-old *Calluna* also occurred within this zone, presumably as the result of later localized fires; they were not markedly different from the rest in either cover or size of plants, and had not been specifically noted in the original survey. On the west side of the 2½-year burning line, any direct indication of these accidental fires had naturally been lost. Nevertheless, it is possible to suggest the path of at least one of them by extrapolation; and it will be noted from the map that the boundaries thus defined strikingly enclose the whole of the oblique strip containing (J). Moreover, the eastern part of the other (J) cluster similarly coincides with the position of the other 6/7-year-old patch; and there is no reason to suppose that, despite its N./S. deflection along the gully, this fire did not also cross to the western side to affect that part of the area now bearing the rest of (J).

Again, as at Beaulieu Road, we find that the first stage of the analysis has ignored the most conspicuous feature of the area—in this case the 2½-year burning line—in favour of a more subtle but ecologically legitimate division. It is outside the scope of this paper to inquire closely into the specific connection between (J) and the 6/7-year burning. *Festuca* is notoriously resistant to trampling and similar disturbance; the most economical hypothesis on present information is that it spread over the area during the military occupation, that it was later ousted by taller competitors except where the intermediate burning reopened the community sufficiently for *Festuca* to persist, and that the surrounding parts completed their reversion to more typical heath before the later 2½-year burning could act in the same way. Even so, there is still some circumstantial evidence, both from inspection of the site and from examination of air photographs showing the direction of former tank tracks, that a more complex story may be partly involved. The fact that a simple causal explanation for the distribution of *Festuca* and its associates is not possible on the available data by no means diminishes the value of the statistical analysis; on the contrary, it has directed attention to a phase of the burning pattern which might otherwise have been missed, and has succeeded in extricating a concrete ecological problem from an otherwise incoherent community mosaic.

It will be noted, incidentally, that, although the Matley and Beaulieu areas both lie across a boundary between Plateau Gravel and Barton Sand, the striking primary vegetation/soil relationships exposed at Beaulieu are not reiterated by the first division here: this was already foreshadowed by the presence of *Molinia*—the operative species at Beaulieu—throughout the Matley area. The explanation probably lies in the fact that the range of drainage conditions may be less extreme at Matley; it lies entirely on a damp north slope, and has no topographical counterpart to the dry, raised Beaulieu ridge. Moreover, the geological boundary at

Matley is very patchy and indistinct, with correspondingly less clear-cut subsoil differences.

To return to the map, we find that, within (J), the small (JG) final grouping is very sharply circumscribed; apart from one isolated quadrat, it is restricted to the top part of the disused cart-track down the gully, which has become filled with *Ulex europaeus*. The division between (JgDH) and the recombined group (Jgd) and (JgDh) is unimportant; it is based primarily on a small positive association between *U. minor* and *Erica cinerea*, and the 35 quadrats separated as (JgDH) have no easily discernible ecological focus.

It is the negative (j) side of the original dichotomy which provides the main bulk of the Matley population. Here the first division is on *E. cinerea*, mainly by virtue of a fairly strong positive association with *Ulex minor* and a slightly weaker negative one with *Trichophorum*. All but 6 of the (jD) quadrats emerge at the next division as the final grouping (jDC). Remembering that these 167 quadrats also contain the universal A and B, we note that (jDC) is defined by the presence of four species—*Calluna*, *Molinia*, *Erica cinerea* and *E. tetralix*—very widespread in New Forest heaths in general, so we are not surprised to find it abundant and widely distributed here. It is most poorly represented in the oldest and densest vegetation at the N.E. corner, and this would accord with the observations of Fritsch & Parker (1913) at Hindhead that *E. cinerea* is gradually eliminated in the late phases of a burn succession. The separation of (jDc) is based only on two very small associations; nevertheless, we find that its 6 quadrats are confined to the top half of the slope, suggesting here, at least, the possible operation of a soil or drainage factor.

Within (jd), the 18 quadrats of (jdH) have particular interest in that they nearly all lie within the postulated limits of the 6/7-year burning and partly fill the spaces between the *Festuca* quadrats. But since there is a strong positive association between *Festuca* and *Ulex minor* in the population as a whole, it is reasonable to find (jdH) sharing the same habitat zones which favour (J).

Lastly, we have a minor division of (jdh) on *Trichophorum*. Although separated on a single small positive association, the 37 quadrats of (jdhF) have a recognizable ecological niche in that they lie preponderantly in the lower and wetter part of the area, with nearly half of them clustered in the pocket of rank vegetation at the N.E. corner. The big group (jdhf) thus remains to fill the gaps left after all the disturbing elements in the original community have been allotted their places. All its 100 quadrats contain *Erica tetralix* as well as *Calluna* and *Molinia*; we recognize it in fact as virtually identical in composition with the (BC) community of Beaulieu Road, but this time appearing at the extreme negative end of the hierarchy.

The emergence of this same triple community from the independent Beaulieu and Matley analyses is perhaps one of the more encouraging results of the method. In each case, the *Calluna-Molinia-Erica tetralix* grouping forms an appreciable part of the total population; and in certain respects it may be regarded as the basic heathland community of the region. It has become further impoverished in parts of Beaulieu Road by unfavourable soil conditions or by very recent burning: it has become locally enriched or modified at Matley by the entry of other disturbing species to give the more complex Matley mosaic.

## IV. DISCUSSION

It is arguable that vegetation as an organic whole consists, not of individual species, but of relationships between them. A painting is *made* of pigments, oil and canvas, but it may reasonably be said that it *is* a set of relationships between forms and colours; similarly, vegetation is *made* of plants, but it *is* a matrix of correlation coefficients. It is for this reason that we agree with Greig-Smith (1957, p. 163) as to the importance of factor analysis, although in the only published example known to us—Goodall (1954)—we believe that the method adopted is inappropriate to the study of vegetation (we shall support this view in a later communication). We have already pointed out that association-analysis is an approximation to factor analysis, since it seeks to divide the population by causing the greatest possible discontinuity in a postulated first factor. (A closer approximation would be obtained by summing, not  $\chi^2$  values, but the corresponding correlation coefficients  $\left(\sqrt{\frac{\chi^2}{N}}\right)$  taken regardless of sign; but for hand computation this would

be extremely laborious.) Of the various forms of factorial analysis, we have invoked the very flexible, and now conventional, multifactorial analysis; the methods of Goodall (1953) and Hopkins (1957) are, however, approximations to the older and rather more crude method of group factor analysis. Both attempt to extricate groups of species internally linked only by positive associations. Goodall seeks these directly, but by a method we have shown to be inefficient; Hopkins seeks them by the active elimination of negative associations. Goodall's quadrats, like ours, become members of discrete groups. Hopkins, on the other hand, 'scores' his quadrats according to the groups represented within them. This we believe to be well intentioned but ill advised; the problem of specification of individuals when qualitative data have been factorized is controversial and difficult, and is not to be solved by a simple arbitrary scoring system.

In short, therefore, we believe that the foregoing pages have shown that our method fulfils its statistical function of dividing a population as efficiently as possible, and that it does this by selecting the greatest possible discontinuity in the first centroid axis of the association-matrix under scrutiny. However, even a complete multifactorial analysis does not *necessarily* provide any useful ecological information; the 'factors' it discloses are mathematical postulates, arising from a transformation designed to express the information contained in the matrix in the most economical form possible. The solution is not unique, nor do the 'factors' necessarily relate to any real entities in the outside world. Whether such corresponding entities can be recognized, what their nature may be, and whether the information they provide is useful for other purposes, can only be decided by appeal to the area under investigation. In the two test-communities studied here, we have found that our major groupings do in fact correspond with recognizable habitat factors, and that the analysis moreover directs attention to important points which—even in these very restricted communities—might otherwise have remained obscured by more conspicuous but superficial features.

But, although a sorting system may be only a crude approximation to a factor analysis, it provides, in compensation, additional information of a different kind; for the homogeneous groups which result may be used in their own right for phytosociological study. The groups, considered as subclasses of the original

population, are likely to be most meaningful if defined in their lowest possible order; for example, at Beaulieu Road, (BC) happens also to be (ABCe), and at Matley (jdhf) is equivalent to (ABCdfhjkm). The positively defining species of a class may together be regarded broadly as its 'characteristic species'. They provide its 'constants'; and in our hierarchical system they also carry a strong element of the 'fidelity' concept. It is tempting to try to identify the various classes further in terms of their total species. Percentage composition will not do; the fiducial limits of a percentage depend markedly both on its value and on the size of the parent population, and the danger of comparative lists of percentage compositions is that they suggest a precision which they do not in fact possess. We have preferred in this paper to publish only the figures of actual occurrences in quadrats; an approximate comparison between the abundance of a species in different communities can then be made by the  $\chi^2$  test.

It must be emphasized that any one set of final groupings relates only to the particular population which has been examined. The findings from one area thus cannot be used directly to provide the basis for a more generally applicable phytosociological classification. Nevertheless, where certain species are consistently associated in the vegetation of a region, the group containing them is likely to reappear in a number of independent analyses from separate areas. The recognition of the substantially similar (ABC) grouping at both Beaulieu Road and Matley is a case in point; it provides a common focus to which specific local variations can be related.

The number of final groupings to be considered will necessarily change with the precise definition of homogeneity adopted. If non-significance is, as we recommend, to be the major criterion, then our method may be expected to give the smallest possible number of final groups; indeterminacy may, if a population contains several abundant and highly associative species, reduce the number slightly—as it does at Beaulieu Road—but is objectionable on other grounds. In any case, the number of groupings by any system is a direct reflection of the heterogeneity of the vegetation; and if in a very complex area the complete separation becomes too unwieldy to be appraised, it is possible either to raise the significance level or to examine only the major subdivisions. At Matley, for instance, the first-order division makes a useful distinction between the groupings which approach a poor 'grass-heath' in composition and those of more typical heath.

The use of association-analysis in ecological research is most likely to be in connection with the primary survey of an area rather than in its later more detailed investigation. It will serve to *expose* the problems, not to *solve* them. For this purpose a comparatively crude but easily applicable method is likely to be of more value than a more precise but less straightforward technique; and this is an additional reason for our advocacy of statistical approximations where feasible. The method is at the moment being programmed for a 'Pegasus' digital computer; when the programme is fully tested we propose to carry out a comparative investigation of the Matley data with a view to making definite recommendations as to the basic parameter to be used (corrected  $\chi^2$ , uncorrected  $\chi^2$ , or

$\sqrt{\frac{\chi^2}{N}}$ ), the path of subdivision (hierarchical or by pooling), and the value of the

different significance levels. We shall also examine populations involving larger numbers of species. The results presented in this paper already suggest that association-analysis of this type will prove a powerful ecological tool; and it seems likely, even at the present stage of this investigation, that we shall not require to change the basic principles of the method.

## SUMMARY

1. Goodall's method of subdividing a set of sample quadrats into homogeneous groups, in which all species-associations have been made non-significant or indeterminate, is subjected to a theoretical analysis.

2. A new sorting method is described, consisting of hierarchical division on the species with the highest aggregated value of the chosen association-index in the class under study. The properties of suitable indices are briefly considered, but in order to bring this exploratory study within the reach of hand-computation, the index used ( $\Sigma\chi^2$ ) was constructed from corrected  $\chi^2$  values, non-significant and indeterminate values being taken as zero, and ambiguities being resolved in the next highest class in which discrimination is possible. Equal weight is given to positive and negative associations.

3. The statistical efficiency of the method is confirmed by its application to populations from two heathland communities.

4. The nature of the ecological information thus obtained is assessed, and it is concluded that a method of this type is likely to prove a very useful tool in primary survey.

## ACKNOWLEDGMENTS

It is with pleasure that we record our indebtedness to the following: To Mr M. J. Hudson and Mr P. F. Hunt, for the entire collection of the quadrat data, and for help in the preliminary sorting of the record cards; to Mr J. A. Bailey, for other assistance in the field; to Mr A. C. Clark, of the Geography Department, for his careful drawing of the maps; to Mr E. Wynne Jones, the Deputy-Surveyor of the New Forest, for permission to work on the areas; to Dr J. G. Manners, for placing his extensive knowledge of New Forest heaths at our disposal; and to Dr G. N. Lance, of the Mathematics Sub-Department of Computation, for advice on the application of a digital computer to this type of work.

## REFERENCES

- FRITSCH, F. E. (1913). The heath association on Hindhead Common. *New Phytol.*, **12**, 149-63.  
 GOODALL, D. W. (1953). Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. *Aust. J. Bot.*, **1**, 39-63.  
 GOODALL, D. W. (1954). Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Aust. J. Bot.*, **2**, 304-24.  
 GREIG-SMITH, P. (1957). *Quantitative Plant Ecology*. Butterworths, London.  
 HOPKINS, B. (1957). Pattern in the plant community. *J. Ecol.*, **45**, 451-63.  
 TUOMIKOSKI, R. (1942). Untersuchungen über die Untervegetation der Bruchmoore in Ost-finnland. I. Zur Methodik der pflanzensoziologischen Systematik. *Ann. Soc. Zool.-Bot. Fenn. Vanamo*, **17**, 1-203. (*ex* Goodall 1953).

(Received 8 March 1958)

## LINKED CITATIONS

- Page 1 of 1 -



You have printed the following article:

### **Multivariate Methods in Plant Ecology: I. Association-Analysis in Plant Communities**

W. T. Williams; J. M. Lambert

*The Journal of Ecology*, Vol. 47, No. 1. (Mar., 1959), pp. 83-101.

Stable URL:

<http://links.jstor.org/sici?sici=0022-0477%28195903%2947%3A1%3C83%3AMMPEI%3E2.0.CO%3B2-6>

---

*This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.*

## References

### **The Heath Association on Hindhead Common**

F. E. Fritsch; Winifred M. Parker

*New Phytologist*, Vol. 12, No. 4/5. (Apr. - May, 1913), pp. 148-163.

Stable URL:

<http://links.jstor.org/sici?sici=0028-646X%28191304%2F05%2912%3A4%2F5%3C148%3ATHAOHC%3E2.0.CO%3B2-I>

### **Pattern in the Plant Community**

Brian Hopkins

*The Journal of Ecology*, Vol. 45, No. 2. (Jul., 1957), pp. 451-463.

Stable URL:

<http://links.jstor.org/sici?sici=0022-0477%28195707%2945%3A2%3C451%3APITPC%3E2.0.CO%3B2-K>