ENGLISH NATURE

Statistical advice on dealing with data from permanent plots and other monitoring sites

No. 31 - English Nature Research Reports

working today
for nature tomorrow

ENGLISH NATURE Contract No. F72-07-56

Statistical advice on dealing with data from permanent plots and other monitoring studies

Report by

Gavin J.S. Ross,
Statistics Department,
Rothamsted Experimental Station,
Harpenden, Herts AL5 2JQ.

ENGLISH NATURE Contract No. F72-07-56

Statistical advice on dealing with data from permanent plots and other monitoring studies

Report by Gavin J.S. Ross, Statistics Department,
Rothamsted Experimental Station, Harpenden Herts AL5 2JQ

Contents

Acknowledgements

Thanks are due to Sandra Suarez and Eileen Stoydin for secretarial services, and to Professor Vic Barnett for overall supervision and advice.

0. General Discussion

The statistical aspects of monitoring changes in vegetation with time concern several types of question:

1) *Data presentation and description*
What are the best ways of summarising, tabulating and displaying data to demonstrate the main changes that have been observed?

2) *Statistical modelling, estimating parameters and testing hypotheses*
To what extent can the changes be substantiated against a background of sampling errors and natural variability? What statistical models should be used, and with what accuracy can their parameters be estimated?

3) *Sampling design and adequacy of data*
Are the data adequate in respect of sample size, representativeness, and comparability? Are there more efficient ways of obtaining the information at no extra cost in terms of time and expertise.

4) *Use of statistical software*
Which computer packages are most appropriate for analysis of this type of data?

1. Data presentation and descriptive analysis

Data presentation is in some ways the most important stage in statistical analysis. The large amount of individual recording is to be gathered together and summarized without losing essential information which might invalidate the processes of modelling and hypothesis testing. In many studies there are non statistical inferences that can only be revealed graphically, by maps, scatter diagrams, networks or empirical curves.

Any statistical summary, such as the mean and variance of a sample, or the correlation between two variables, loses some detailed information which may or may not be relevant to subsequent modelling and hypothesis testing. For example if the sample is from a very skew distribution, with mostly small values but the occasional extremely large value, the sample means and variance will generally underestimate the population mean and variance, and will not indicate the skew nature of the distribution. A correlation coefficient may be small, not because there is no relationship between the variables but because the relationship is not a straight line.

Graphical techniques draw attention to details which may be important, such as patterns

among the observations that are most extreme, which may suggest that important extra variables have been omitted from the analysis.

2. Statistical modelling and hypothesis testing

Statistical modelling involves constructing a hypothetical framework by which to measure and explain contrasting sets of observations. If we already know much about the mechanisms underlying the observations it may be possible to incorporate that knowledge into the model, and to assume that the remaining discrepancies are due to random variation. Even so the full model may require external measurements which are not available with the data set, and so a more empirical approach is required.

Empirical models do not require the underlying mechanism to be known, but are chosen to reflect the observed behaviour of the data and of analogous samples from elsewhere. Effects are measured and their standard errors estimated, but the model is only valid in so far as it fits the data, without supporting any particular interpretation of the reasons for the effects.

In the context of vegetation sampling it is particulary difficult to postulate useful explanatory models for the measured variables. The difficulties associated even with the 'simple' count of species present, or species richness (discussed in data set 2), or the asymmetry of the measure presence-absence, means that it is not always possible to apply the more obvious statistical tests. The concept of random or aggregated distribution of individuals in space is more easily analysed with counts of discrete plants rather than with cover measures of spreading clumps.

The number of species recorded in a given area is related to area, and the relationship depends on assumptions about species diversity. The problem is influenced by ideas on species definition, if there are several specimens from a complex genus, or if there is hybridisation. So the numerical counting of species or individuals within species gives rise to a special kind of statistical distribution which is different from a conventional sampling distribution because its shape varies with sample size.

Problems with curve fitting and time series analysis arise because of the correlations over time, whether or not the same plots are sampled on separate occasions. There are several components of variation: pure sampling variation on a particular occasion, extra variation due to aggregation, trends due to competition and spread of individuals, long term trends due to climatic effects, discrete events such as storms, landslips or fires, and interventions such as transplanting, coppicing, grazing or fencing, or application of fertilizers. The problem is often how best to make use of the extra information when the data can only support a small trend or shift in the mean value.

In many situations it is preferable to do statistical analysis on transformed data, especially when the distributions are very skew. The transformations most commonly used are as follows:

Counts (small numbers)	square root (x)
Counts (large numbers)	log (x + 1)
Proportions (r/n)	logit ((r + ½)/(m + 1))
Where logit $x = e^x/(1 + e^x)$	

These transformations make the distribution more symmetrical and equalise the variances, so that significance tests are more valid. (See references 3 and 52.)

3. Statistical design and adequacy of data

The discussion of data sets indicates, where relevant, how it might be possible to increase the amount of information in future, either by redesign of the existing scheme (with no extra resources) or where the data are insufficient as they stand, and require larger samples or more intensive measurement on existing samples.

Past data cannot be repeated, and there is virtue in comparability, as in data set 11. But if the data are not sufficiently representative to allow any inferences to a wider population of sites, or not sufficiently large to detect significant changes that are also ecologically important, then a case must be made for changes in the method of collection.

In all cases the longer the series continues the more reliable will be any statements about change. As it is unreasonable to expect a linear trend to continue for ever, we need to know if it is part of a cycle which may reverse, or an approach to an equilibrium. If there are cycles, regular or irregular, it is necessary to observe several reversals of sign before cyclic behaviour can be demonstrated.

4. Statistical software

The analyses performed in the study of the data sets made use of statistical packages developed at Rothamsted by author and colleagues. It is not the purpose of this report to recommend that English Nature workers should use these programs, but to state that they are very useful for the statistical analyses recommended. Those who are familiar

with other systems may find it convenient to use what they know, provided the required analyses may be done. (See References 51-56.)

GENSTAT is the general statistical package, containing extensive facilities for multivariate and cluster analysis, regression and time series, and the analysis of multiway tables.

MLP is the maximum likelihood program for fitting curves, distributions and general models, with a wide range of related facilities. Both GENSTAT and MLP are distributed by NAG Ltd of Oxford.

GLIM is the well known generalised linear modelling program for regression, multiway tables and the analysis of proportions and counts. It was developed both at Rothamsted and at other sites by a working party of the Royal Statistical Society. Its facilities also exist in GENSTAT.

CLASP is a special purpose Cluster Analysis and ordination package developed at Rothamsted, much of which is incorporated into GENSTAT.

The advantage of the more general packages over special purpose programs such as DECORANA and TWINSPAN is that the assumptions and constraints in these programs may be relaxed. For ordination in particular it may be advisable to use different scaling methods if it is found that the standard output tends to be unduly influenced by rare species or outlying plots. Several alternative methods of cluster analysis may be useful if particular uses of cluster analysis are envisaged.

DATA SET 1. WYTHAM, PERMANENT PLOTS

1. Wytham is predominantly woodland covering c.400 ha. 10 x 10 m plots (permanently marked) were established in 1974 at alternate intersections of a 100m grid - i.e. systematic distribution. 40 plots were re-recorded in 1983, and virtually the full set (159 out of 164) in 1991. A variety of data relating to the tree and shrub layer and to ground flora have been collected (plus soil information for 1974 which may be repeated).

2. The scheme was not set up to monitor any specific change, but rather to provide baseline information on a variety of aspects across the whole wood with the best likelihood that any future change (unspecified) would be detected.

   Questions that we now wish to address relate to the changing abundance of different species and how these might be related to (although not necessarily caused by) changes in canopy conditions (and when the data are available soil conditions).

   The following questions are concerned with how we might assess species change in the data set.

   Comparing species presence/absence

   There is a list of species present in each plot at each time (Table 1). On each occasion species may have been missed when they were really present, either because of errors by the observers, or because minor differences in the time of year when the plots were recorded meant that the species had not yet appeared or had already died back.

   In addition species may genuinely have appeared (or disappeared) from a plot between 1974 and 1991. These latter changes are the ones that we are concerned with, although difficult in practice to separate from the former. Changes for a single plot are difficult to interpret. The following procedures have therefore been proposed for looking at the combined species/plot set. They seem reasonable, but are they?

   (a) Determine the frequency for each species (number of plots in which it occurs) in 1974 and in 1991. Arrange the species in order of their change in frequency (Table 2). Look for common characteristics (defined on the basis of independent information) among those species at the top or bottom of the list. Is there any way of saying that a particular level of change is "significant" or is the decision purely arbitrary?

   (b) The above looks only at nett overall change in frequency. It might be more interesting, particularly if we are concerned with spatial patterns, to classify each plot for each species as showing an increase (+), decrease (-), no change (0). Two possibilities then present themselves (Table 3) - looking, for a given species, at the number and distribution of plots that show each type of change and looking for common patterns of change for all species in a plot (or a set of species defined by some independent set of parameters). Again can anything be said about the significance of such distributions?

   (c) A third approach is to use a classificatory procedure such as TWINSPAN on the plot data, either for each time separately or with both sets of results included. Plots which have changed in

their species composition may be detected by changes in the end-groups to which they belong or by shifts in their location on a DECORANA plot. How can we decide which of such changes are important?

(d) Are there other ways of comparing such lists?

(e) What differences/complications would it introduce if we tried to add a third recording time into the analysis?

(f) What further inferences might we be able to draw if the original sample were randomly placed, rather than systematically distributed, about the changes in the set itself or about the site from which it is drawn? Similarly what conclusions might be drawn if the plots were <u>subjectively</u> placed rather than systematically?

(g) If instead of looking for which species have changed we have <u>a priori</u> reasons for proposing that a small number of species will change in specific ways (e.g. bramble decrease because of grazing, nettles increase. What difference might this make to the analysis and interpretation that we make?

(h) Suppose that for some/all species there is for each plot an estimate of abundance rather than just presence/absence. What additional forms of analysis does this allow

- if % cover is used

- if frequency in a series of sub-plots is measured

- if a restricted scale is used (e.g. scale value 0 = absent, 1-5% cover, 2 = 6-25%, 3 = 26-50%, 4 = 51-75%, 5 = 76-100% cover)

- if a Domin scale is used (or Currall transformation of the latter)

Again how might the analysis be affected if there is <u>a priori</u> interest in particular species and how they might have changed as opposed to looking at the whole data-set to see which have changed?

TABLE 1. WYTHAM — TYPE OF DATA. Presence/absence for

	Plot	452077		452079		452051		452059		453076		. . . . . → 164 plots altogether
	Year	1974	1991	1974	1991	1974	1991	1974	1991	1974	1991	
Angelica sylvestris		✓	✓									
Brachypodium sylvaticum		✓	✓	✓	✓	✓	✓		✓	✓	✓	
Circaea lutetiana		✓	✓	✓	✓	✓	✓	✓		✓		
Deschampsia cespitosa		✓	✓	✓	✓		✓				✓	
Endymion non-scriptus		✓	✓			✓	✓	✓	✓		✓	
Galium aparine		✓	✓				✓	✓		✓		
Geum urbanum		✓				✓			✓		✓	
Glechoma hederacea		✓	✓	✓	✓	✓	✓			✓		
Mercurialis perennis		✓	✓	✓		✓	✓	✓	✓	✓	✓	
Poa Trivialis		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Rubus fruticosus		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Urtica dioica		✓	✓			✓	✓	✓	✓			
Viola riviniana		✓				✓	✓					
Arum maculatum			✓		✓				✓		✓	
Hypericum hirsutum			✓					✓	✓			
Rumex sanguineus			✓									
Cryopteris filix mas			✓	✓			✓			✓		
Carex sylvatica			✓	✓	✓							
Filipendula ulmaria			✓									
Scrophularia nodosa			✓									
Taraxacum officinale			✓									
Ajuga reptans				✓								
Bromus erectus				✓								
Carex acutiformis				✓								
Carex pendula				✓	✓	✓						
Galeobdolon luteum				✓		✓	✓				✓	
Rosa canensis/spp				✓	✓	✓		✓	✓	✓	✓	
Rubus communis				✓						✓	✓	
Truncus effusus							✓					
Tiliber sylvestris							✓					
Chamaenerion angustifolium								✓				
Clematis vitalba								✓	✓			
Epilobium montanum								✓	✓			
" tetra.								✓				
Humulus lupulus								✓				
Poa pratensis								✓				
Pteridium aquilinum								✓	✓			
Rubus caesius								✓				
Solanum dulcamara								✓				
~~Arum maculatum~~									✓			
Arctium minus										✓		
Cryopteris dilatata										✓		
Sonchus oleraceus											✓	

TABLE 2. Comparing overall species frequency (no of plots % in which species occurred)

Species	1974	1991
Total no of plots	159	159.
Arum maculatum	84	120
Desch. cesp.	70	100
Brach. sylv.	66	88
.	.	.
.	.	.
.	.	.
Rubus fruticosus	150	150
Endymion non.	96	100
Poa trivialis	86	80
Epipactis helleborine	6	8
Daphne laureola	4	0
.	.	.
.	.	.
.	.	.
Glechoma hederacea	80	60
Geum urbanum	75	50
Circaea lutetiana	100	60.

↑ species showing increased frequency in 1991

species showing little relative 'change'

↓ species showing decreased frequency in 1991

Possible questions

? What are these common characteristics?

TABLE 1. WYTHAM — TYPE OF DATA. Presence/absence for

Species	Plot 452074 1974	452074 1991	452079 1974	452079 1991	452081 1974	452081 1991	452089 1974	452089 1991	453076 1974	453076 1991
Angelica sylvestris	✓	✓								
Brachypodium sylvaticum	✓	✓	✓	✓	✓	✓		✓	✓	✓
Circaea lutetiana	✓	✓	✓	✓	✓	✓	✓		✓	
Deschampsia cespitosa	✓	✓	✓	✓		✓				✓
Endymion non-scriptus	✓	✓			✓	✓	✓	✓		✓
Galium aparine	✓	✓				✓	✓		✓	
Geum urbanum	✓				✓			✓		✓
Glechoma hederacea	✓	✓	✓	✓	✓	✓			✓	
Mercurialis perennis	✓	✓	✓		✓	✓	✓	✓	✓	✓
Poa Trivialis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rubus fruticosus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Urtica dioica	✓	✓			✓	✓	✓	✓		
Viola riviniana	✓				✓	✓				
Arum maculatum		✓		✓				✓		✓
Hypericum hirsutum		✓					✓	✓		
Rumex sanguineus		✓								
Dryopteris filix mas		✓	✓			✓			✓	
Carex sylvatica		✓	✓	✓						
Filipendula ulmaria		✓								
Scrophularia nodosa		✓								
Taraxacum officinale		✓								
Ajuga reptans			✓							
Bromus erectus			✓							
Carex acutiformis			✓							
Carex pendula			✓	✓	✓					
Galeobdolon luteum			✓		✓	✓				✓
Rosa arvensis/spp			✓	✓	✓		✓	✓	✓	✓
Tamus communis			✓						✓	✓
Juncus effusus						✓				
Ribes sylvestris						✓				
Chamaenerion angustifolium							✓			
Clematis vitalba							✓	✓		
Epilobium montanum							✓	✓		
" tetra.							✓			
Humulus lupulus							✓			
Poa pratensis							✓			
Pteridium aquilinum							✓	✓		
Rubus caesius							✓			
Solanum dulcamara							✓			
~~Arum maculatum~~								✓		
Arctium minus									✓	
Dryopteris dilatata									✓	
Sonchus oleraceus										✓

→ 164 plots altogether

TABLE 2. Comparing overall species frequency (no of plots % in which species occurred)

Species	1974	1991
Total no of plots	159	159
Arum maculatum	84	120
Desch. cesp.	70	100
Brach. sylv.	66	88
.	.	.
.	.	.
.	.	.
Rubus fruticosus	150	150
Endymion non.	96	100
Poa trivialis	86	80
Epipactis helleborine	6	8
Daphne laureola	4	0
.	.	.
.	.	.
.	.	.
Glechoma hederacea	80	60
Geum urbanum	75	50
Circaea lutetiana	100	60

↑ species showing increased frequency in 1991

species showing little relative 'change'

↓ species showing decreased frequency in 1991

Possible questions

- What are there common characteristics?

WYTHAM

Table 3a Looking at the "mobility" of species

	No of plots showing change (1974-1991		
	+	-	0
Species			
Arum maculatum	46	10	20
Rubus Fruticosus	9	9	141
Geum urbanum	10	40	60
Blechnum spicant	4	0	0
Polyslichum setiferum	0	4	10
Poa trivalis	24	18	100

etc for about 130 other species

\+ = present in 91, absent in 74

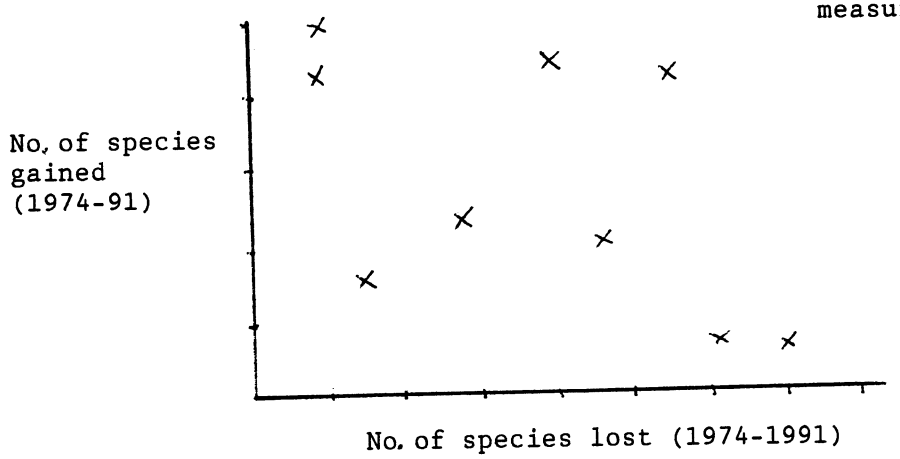
\- = present in 74, absent in 91

0 = present in both years

(Plots in which it is absent in both years are not included)

[Note the numbers in the above table are illustrative, not accurate]

Table 3b Looking at the stability of plots. Plot numbers of losses against gains as a measure of stability?



ENGLISH NATURE Contract No. F72-07-56

Discussion of Data Sets and methods of analysis

Data Set 1

Wytham, Permanent Plots

We have data on 164 permanent plots, systematically arranged at intervals of 100m, sampled fully in 1974, of which 40 were resampled in 1983 and 159 in 1991.

No specific changes are anticipated, but variables have been measured which might be associated with change.

Data consist of presence-absence for each species that ever occurs, and some abundance data.

1. Analysis of Table 2

Using the 159 plots with data in both 1974 and 1991 it is proposed that species should be ranked in order of the differences between the total frequencies for 1991 and 1974. This information is displayed in Table 2.

The information can also be displayed as a simple scatter plot of the 1991 frequencies against the 1974 frequencies. A 45 degree line through the origin indicates the line of equality, while those species plotted above the line show an increase in the period, and those below the line, a decrease. This gives more information than the simple listing of differences which group together all species plotted on the same parallel line regardless of mean frequency.

A simple statistical test for the difference between a proportion r_1/n and a second proportion r_2/n is the Binomial Test. This test assumes that both proportions are samples from a Binomial distribution with mean $P = (r_1+r_2)/2n$. The test does not consider whether the species are observed on the same plots on each occasion, but merely compares the unsigned difference:

$$|r_2-r_1|/2 - .5$$

with its standard error

$$\sqrt{(nPQ)}$$

to obtain a test for the significance of the difference between r_1 and r_2. The ratio Z = value/standard error is tested using tables of the Normal distribution. For a given significance level, such as 5 percent probability of a significant difference in either direction, we can draw contours of equal significance by solving the equation $z = 1.96$ for r_1 given r_2 and n. For example if $n=159$ and $r_2=40$ the critical values of r_1 are approximately 20 and 64. A more exact procedure is to use tables of the binomial distribution and to interpolate for the cumulative probability equalling 0.025 or .975, but this is only necessary when n is small or when the relative frequency is close to zero or 100 percent.

2. Summarising the distribution of change

It is proposed that the distribution of change for each plot be summarised in a 2x2 table.

The statistical significance of change can be assessed by calculating the value of chi-squared (with Yates' correction) for the 2x2 table:

		Year 1	
	Present	Absent	Total
Year 2	A	B	P
Absent	C	D	Q
Total	R	S	N

which is given by the formula

$$(|AD-BC|-N/2)^2 \ *N \ / \ (PQRS)$$

which may be compared with the tables for χ^2 on 1 degree of freedom. This test assumes that the two sets of observations are independent and the samples are random, so it might be questionable if the species were mature trees which might be expected to remain in place for 17 years or more.

If A/C is very close to B/D it is possible that AD-BC is less than N/2 and the correction is not necessary, but the difference is clearly not significant anyway.

An alternative proposal is that for each plot we calculate the number of species that have changed, and to plot the number of species gained against the number of species lost, and to look for patterns.

This is a more difficult problem, because whereas all plots are roughly equivalent in statistical terms, species are certainly not. So if we merely count species, regardless of frequency, we cannot assess significance of change, because the change is likely to affect only the rare or moderately rare species, unless there has been a drastic change in the environment during the period of observation.

Nevertheless, in Cluster Analysis it is common to compare two plots or the same plot at different dates by using a similarity coefficient, such as the Jaccard Coefficient A/(A+B+C) or the Czekanowski Coefficient (A+D)/(A+B+C+D), which is 1 for complete agreement (no change), and otherwise less than 1. These coefficients are means to an end (ranking of similarity) rather than absolute measures, and their statistical distribution depends on the characteristics of each species, and the interactions between species.

Even if we could assume that species were distributed randomly and independently according, say, to the Poisson distribution, we would need to know the mean number M_i of individuals (stems? root centres?) per plot for each species, and calculate the probability $\exp(-M_i)$ of observing no specimens on any given plot. The probability of change between two occasions is greatest when M_i is between 1 and 2, because for small M_i the species is usually absent on both occasions, and for large M_i it is usually present on both occasions.

Thus it is not in general possible to model the frequency of observing particular values of B and C in the 2x2 table. It is necessary to pursue this method of approach via cluster analysis rather than by significance testing.

3. Use of TWINSPAN and DECORANA programs

The use of a particular form of Cluster Analysis is suggested, using TWINSPAN to produce groups and DECORANA to produce plots. The units may be for two years separately or for both years combined.

The results obtained from TWINSPAN and DECORANA depend on what options are used. These procedures are members of a wide class of analyses, provided by many different packages, which aim to reveal data structure by grouping of units and examining the common characteristics of each group.

TWINSPAN stands for Two-Way Indicator Species Analysis and is a hierarchical cluster analysis based on the breaking up of large clusters into two subclusters on the basis of the dominant axis of contrast at any time. It is based on Correspondence Analysis (or

Reciprocal Averaging), and is described in detail by H.G.Gauch in 'Multivariate Analysis in Community Ecology' (CUP, 1982). (Ref. 40.)

DECORANA stands for Detrended Correspondence Analysis. It assumes that the first vector found represents the major trend in the data. In many practical examples the scatter of plots appears curved rather than straight, because plots with little in common are treated as equally unlike, which requires the ends of a single series to curve inwards to map distances adequately. DECORANA divides the first axis into ranges and recentres the plots within each range so that they appear on a straight line.

TWINSPAN and DECORANA obviously work well for some examples, but if it is not clear that there is a dominant trend or a single major axis, the results may be less easy to comprehend.

Correspondence analysis is just one member of a class of two-way analyses which uses a particular weighting of the variables based on their frequencies in the table. It is sometimes argued that it gives too great prominence to rare species, in which case it might be useful to compare the results with those from other procedures.

4. Changes in time on DECORANA plots

Plots which change with time may be detected by changes in their end groups or shifts in their location on the DECORANA plot. How can we decide which of such changes are important?

This question may be illustrated by the analysis of changes of agricultural land use in English Counties (Appendix 1). Here the 10 counties at four dates are clustered and ordinated by methods broadly analogous to TWINSPAN and DECORANA. The counties that were very different at all times remained in a cluster, but where the systematic change with time brought one county in say 1965 to be similar to another

county in 1945, they would appear in the same cluster. The important factor is not the clustering or ordination procedure, but the degree of similarity between the two samples. There is a danger that by looking at DECORANA plots alone one is neglecting the third and higher dimensions, so that while samples which are very similar are plotted close together, the converse is not necessarily true, since they may differ on a dimension not included in the plot.

The statistical problem of assessing the significance of changes in similarity values is that we have no agreed model for the sampling distribution of similarity. It is not even clear that two equal values of similarity have the same meaning in ecological terms. For example if there are only four possible species, a,b,c and d, and two plots in two years have the following compositions:

		Matching coefficient (No. of agreements/No. of comparisons)
Plot 1 Year 1	a b c -	
Plot 1 Year 2	a - c d	2/4
Plot 2 Year 1	a - c -	3/4 3/4
Plot 2 Year 2	- - c -	2/4 2/4 3/4

we can only say that Plot 2 in year 1 was equally similar to Plot 1 (which had more species) and to Plot 2 in Year 2 with fewer species.

The notion of an 'important' change rather than a 'significant' change is obvious when we see it, but in borderline cases requires care in interpretation.

5. Other ways of comparing lists of species

Are there other ways of comparing lists of species?

There may be useful information from clustering of the species rather than of the plots. An association coefficient based on the 2x2 table above uses the signed square-root of chi-squared to give a positive or negative value depending on whether the number of plots in which both species are present (A) is greater or less than expectation. Rare or very common species tend to be excluded from this analysis as there is no information on their association with other species.

The above method gives no information on changes in time.

6. Introducing a third period

What difference/complications are introduced by adding a third period into the analysis?

Any number of intervening times may be included in a combined cluster analysis or ordination, and the intervening times might be expected to lie inbetween the extreme times, unless changes with time are more random than systematic, in which case the trajectory might appear more like a random walk.

This is illustrated for the county data where four dates are used. In the Wytham plots there are only 40 plots sampled in the intermediate date, so that not all plots have three dates.

The 2 × 2 tables of year 1 against year 2 are now no longer so simple, as for any given species we have eight possibilities, and for the 40 plots where all three years are sampled we could produce a three-way table of counts for each species:

	Year 2	Present		Absent	
	Year 3	Present	Absent	Present	Absent
Year 1	Present				
Year 2	Absent				

or we could look at the marginal totals for each year and test for overall differences in proportions by the chi-squared test (as a 2 × 3 contingency table) as follows:
References 4 - 7 may be useful.

	Year 1	Year 2	Year 3
Present			
Absent			

Compute (Observed - Expected)^2/Expected for each cell, add together, and test for χ^2 on 2 degrees of freedom.

7. Random or systematic sampling

What further inferences might be drawn if the original samples were randomly placed rather than systematically?

The argument as to whether samples should be random or systematic is only relevant when we are trying to estimate some particular quantity and the variance of that estimate. If there are suspected regularities in the wood (e.g. because it is a plantation) and the samples coincide in some way with these regularities, bias would be introduced. On the other hand, if the sampling is random in space and the wood is heterogeneous in character, it is possible that some subclasses might be seriously under-represented. Therefore completely random sampling is only advisable when the area is sufficiently uniform in character.

There is usually some subjectivity about the choice of site, and if it is known that there are sub-areas within the site where it is particularly important to have some information, then there is no objection, at the descriptive stage of analysis, to including them. Thus if we wish to know whether a rare plant such as an orchid is associated with other species, we find some plots with orchids and record all the species we find. If on the other hand we wish to estimate the total number of orchids in the whole site, we should take random samples.

If the site were heterogeneous in a defined way, with some parts treated differently from other parts, we should ensure that each part is sufficiently sampled, usually in proportion to the relative areas, except that the number of samples within any one class should not be too small.

8. Reasons for expecting change

What difference does it make if we have *a priori* reasons for expecting a particular change?

In any statistical test where the direction of change is important, we can replace the even-handed two-sided significance test with a one-sided test in which we are only interested in changes in a particular direction. For example the Normal deviate test in which an absolute difference from the mean of greater than 1.96*s is said to be significant at 5 per cent would be replaced by a one-sided test where only the positive difference is tested and the probability of exceeding 1.96*s is 2.5 per cent.

When we are dealing with a 2 × 2 contingency table it is necessary to check on the sign of change, or the sign of association, before deciding whether the change is one to be tested against a tabular value.

In descriptive analysis it is only necessary to check that the observed changes are of the right magnitude and in the right direction. In the absence of any strong indication of the magnitude of the change it is not possible to propose suitable null-hypotheses, although it might be said that failure to change at all when expected represents a possible significant observation.

9. Use of abundance data

If estimates of abundance are available, what additional forms of analysis does this allow?

The estimates of abundance based on % cover provide for a start an estimate of the % abundance in the whole area. If the distribution is patchy, abundance estimates may be used to form contour maps of the area, using suitable interpolation methods. If the distribution of cover is too variable, some form of smoothing may be advisable before contouring to make the picture more intelligible.

An overall picture may be obtained by a principle components analysis of % cover, and the scores on each component can then be plotted and contoured. These scores represent the relative abundance of groups of associated species, so that a map of the first component might distinguish basically the wet and the dry areas, while the second component might distinguish the inner areas from the marginal areas etc.

Similarity and distance measures may make use of the quantitative information in % cover, and the main problem is how to scale the variable to give equal emphasis to equal change. For this reason it is likely that the Domin scale, or something similar, would be more suitable because we are more likely to be impressed by multiplicative increases than in additive increases in abundance. An alternative method given % cover is to use a logarithmic transformation (adding a suitable small quantity to avoid taking log of zero).

Full quantitative information enables the differences in abundance in the commoner species to be taken into account, whereas presence absence is fairly uninformative if a species is nearly always present.

For species where individual plants are counted it is possible to fit frequency distributions to each species, and to test for non-random density against the Poisson distribution, which would be expected if all plots were equivalent.

For all species together it is possible to estimate diversity indices from the cover information, to examine whether there is an increase or decrease in diversity with time. Diversity from % cover would have to be measured by ranking the species in order of frequency, and using the inter-quartile range (Kempton) or similar score. Diversity is more stable a measure than species richness (simple count of species) and is less affected by plot size.

The alternative restricted scales are presumably used because they are quicker to apply. The loss of information in grouping may not be very serious in relation to the changes being observed. On the other hand if a species consistently scored within one wide class (say 26-50%) it might be advisable to use a more refined scale next time.

The Currall transformation of the Domin Scale attempts to restore the % cover score, for the purposes of mean values. It is therefore better for measuring total cover, but less good as a basis for similarity scores, because the contribution of rarer or smaller species is largely ignored.

10. Interest in particular species

If there is *a priori* interest in a particular species how does this affect the analysis?

There is no difficulty in analysing the distribution of a particular species on its own, and all the resources of conventional single variable statistics are available (regression, curve fitting, analysis of variance), depending of course on the assumptions about randomness and treatment classifications of the plots.

If the suggestion is that certain species be given prominence in the analysis, the that may be achieved by assigning weighting coefficients to each species in the multivariate analysis. It may happen that in the unweighted analysis the species of interest does not feature among any of the dominant contrasts because it is not highly associated with any other set of species, and therefore the clusters and ordinating variables do not take much account of it. Weighted analysis increases the priority attached to that species, and the relationships between other species will be relegated to secondary importance. The justification for this procedure must be clearly stated. (It is more relevant in non-ecological contexts to use weights if the costs of obtaining different measurements vary widely, in which case it would be useful to base classifications on the least costly measurements).

DATASET 2

BUCKINGHAM PLOTS

1. Three 10 x 10 m plots were placed at random in each of three very contrasting parts of a wood (coppice, clear-fell, uncut areas, each about 1-2 ha), permanently marked and then recorded annually for five years. For each plot, for each year, there is a species list. More plots would have been placed in each area if time had permitted, but the approach to answering the questions below is, I think, more or less independent of plot number. Our concern is with changes in species-richness per plot in the different areas and with the differing responses of individual species.

2. What is the best way to compare mean species-richness per plot between the areas and how it changes over time? Possible options seem to be:

   a. Compare initial mean values (as with any other set of independently placed sets of plots).

   b. Determine for each plot the difference in species-richness between the beginning and end of the sampling period (ignoring intermediate values); calculate a mean difference for each area and compare these between areas.

   c. Compare the time curves (Fig 1), so making use of the intermediate data (how?).

3. The permanent plots may change as a result of the recording process. Is it reasonable to take periodic independent random samples of plots (say every five years) to check that the permanent plots remain representative?

4. Suppose the interest is in the differing response of an individual species over time in the three areas. (With only three plots per area there may be little point in this sort of analysis but, in other instances, 5-10 per area may be available.) Curves might be produced showing:

   a. Change in frequency over time for that species.

   b. Change in mean abundance/cover per plot in each area for that species.

   How may these curves be compared?

5. With this approach (one clear-fell area, one coppice area, one uncut area) conclusions can only be drawn about the differences between those <u>particular</u> areas. To generalise to the effects of clear-fells (say) elsewhere in the wood or in other woods, replicates of the clear-fell areas would be needed. Can any advice be given on general principles for deciding how to spread a given number of plots (resources are limited)? Suppose six clear-fells are available (plus six coppice blocks, six uncut blocks) and only twelve plots can be allocated to the clear-fell areas. Is it better to have two plots in each, use only two and have six plots in each area, or somewhere in between?

6. Suppose that with the five year series of data, that year 1 and 2 are felling whereas years 3-5 are post-felling. Two options for data comparison are often proposed:

   a. Compare year 1 with a selected post-felling year (usually 3 or 5).

   b. Compare year 1 with the mean of the post-felling years.

   What are the advantages and drawbacks of each method?

FIG 1.

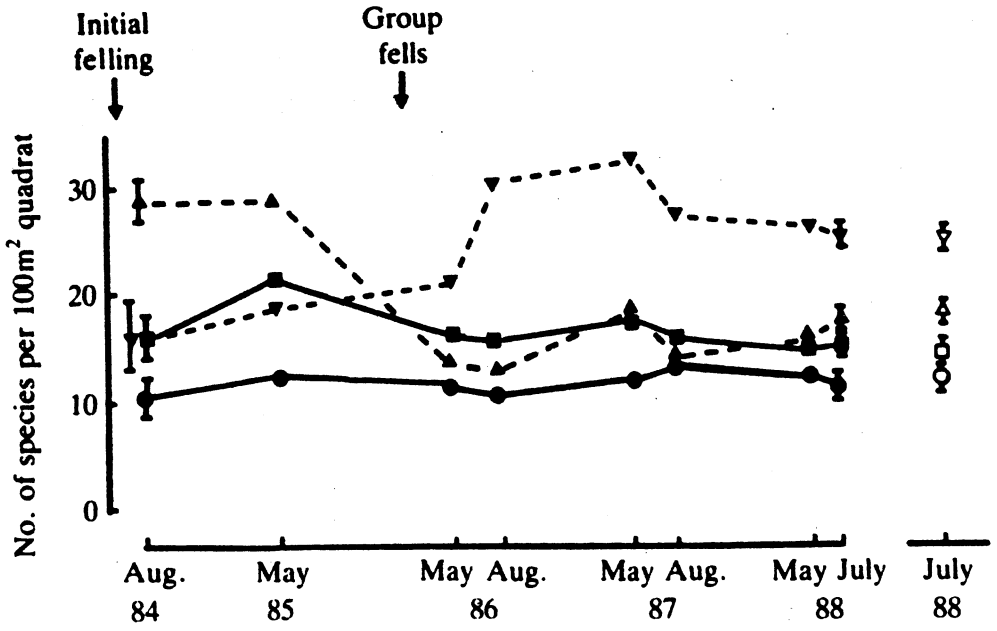


Figure 1. Changes in the field layer between 1984 and 1988 mean no. of species per quadrat. ●——● undisturbed quadrats, ■——■ coppice quadrats, ▼--▼ group fell quadrats, ▲--▲ clear fell quadrats. The values for an independent set of three quadrats in each area recorded in July 1988 only are shown by the open symbols.

Discussion of Data Set 2

Buckingham Plots

1. *Changes in Species Richness*

The statistic 'species richness', defined as the numbers of species recorded on a plot of given size, is difficult to characterise statistically. The problem is that it does not discriminate between different types of species or their behaviour or abundance. If we are comparing similar plots, or the same plots on different occasions, there will be a certain number of dominant species always present. Changes in species richness will be in respect of rarer species, those patchily distributed or those sensitive to environmental or competitive pressure. Two identical values of species richness can represent very different situations, such as the loss of three important species and the gain of three casuals.

If we wished to know the distributional properties of species richness, we would need to know the distribution of each species in turn, so that we could combine the probabilities of presence/absence in each case. Since we do not have this information, we are left with the option of treating species richness as a simple non-negative random variable whose distribution must be inferred from the data being analysed.

If changes are small we can safely proceed as if the variance is constant. If changes are larger (say, the ratio of the largest to the smallest value exceeds 2), it would be more appropriate to use a log transformation.

2a. Comparison of different sets of plots at the same date

The plots are grouped in areas, three per area. So if we had data such as:

Coppice	13,	17,	18
Clear Fell	25,	42,	20
Uncut	13,	8,	12

the analysis of variance of the raw data (R = species richness) is:

Source	*Sum of Squares*	*d.f.*	*Mean Squares*	*F.ratio*
Between groups	518	2	259	5.29
Within groups	294	6	49	
Total	812	8		

The analysis of variance of $\log_e$ (R) is:

Source	*Sum of Squares*	*d.f.*	*Mean Squares*	*F.ratio*
Between groups	1.3423	2	0.67115	8.29
Within groups	0.4857	6	0.08095	
Total	1.8280	8		

This example shows that the difference between groups is less clearly defined on the natural scale because the 'Clear Fell' group has larger variance as well as larger mean. On the log scale the variances are more similar, and the differences between groups are more significant.

2b. *Differences over time for each plot*

The analysis of species richness on the same plot over a short period of time (five years) is clearly affected by the presence of persistent species that should take no part in the analysis. Some analysis could be done in terms of species gained and species lost, since these are the species that constitute the change. A significant change would be indicated if the number of species gained was significantly different from the number of species lost.

After a much longer time period it would be more valid to compare the two mean values as in paragraph 2a, by a t-test or analysis of variance.

2c. *Comparison of time curves*

Comparison of curves is difficult unless they can be fitted by a simple parametric form such as a straight line. The illustrated data do not conform to this because of the discrete event (Group Fells) which produces discontinuity in the response.

If we simply look at linear responses, say to the illustrated solid lines, we could test in sequence

a) that the lines are parallel, indicating movement in the same direction, (otherwise they converge or diverge).

b) that the lines are parallel and not significantly separated, indicating no overall difference, otherwise that they retain significant separation.

The standard regression analysis for this situation is to fit in succession 1) A common line, with residual sum of squares S1, 2) Parallel lines with common slope but different intercepts, using the pooled within line sums of squares and products to calculate the regression slope, with residual sum of squares S2, and 3) Separate lines for each set, with total residual sum of squares S3. The analysis of variance is then set out as follows:

Source	*Sum of Squares*	*d.f.*
Between parallel lines	S1 - S2	1
Between slopes	S2 - S3	1
Within lines (error)	S3	2(n-2)

For more complicated curves a variety of generalisations of this procedure may be done.

However since we are here dealing with repeated observations on the same plot it is not strictly valid to analyse the data in this way. If they were different plots on each occasion then it would be valid.

Otherwise the most telling analysis is the graph itself. The procedure of 2b may be used on any occasion where significance needs to be checked.

The other way of making use of the intermediate data is to collate the species richness over a number of years, so that it refers to the total number of species recorded in at least one of the years in question. This is similar in some ways to the use of larger plots, when the total number of species recorded tends to increase with size of plot. [See, for example, C.B. Williams, 'Pattern in the Balance of Nature'].

3. *Monitoring the permanent plots*
If the permanent plots are in dynamic equilibrium we may find that they gain and lose species over time, or some species may only be observed in certain years. Their changes may be assessed in the same way as in the previous paragraphs. (2a-2c).

4a *Changes of frequency over time for a given species*
If we are only recording presence-absence, then the proportion of plots in which a species is present is a binomial variable. To test whether two binomial samples have the same mean we can use the χ^2 test for the 2 × 2 table of presence/absence, or the

standard binomial test (described in elementary text books such as Snedecor and Cochran). Fairly large samples (n = 10 or more) would be needed to detect any change, and that change would have to be substantial. Alternatively, since the plots are identifiable, we can again record the proportion of plots on which the species is lost or gained between two occasions.

4b *Change in mean abundance or cover*

If abundance refers to separate individuals it is possible to analyse the counts in an analysis of variance (as in 2a), for which the log (n + 1) transformation is appropriate (the + 1 prevents log of zero, and improves the relative constancy of variance). Clearly for long lived species such as trees or shrubs the continued survival of the specimen over five years is of no great interest, but it does again alter one's perception of random variation in this context. For annuals or biennials that are presumably different plants on different occasions, the values are more independent.

The analysis of cover score needs more care, as it is presumably only a grouped value from a DOMIN or other scale. Each of these scales have their own problems in deciding how best to make comparisons. The DOMIN scale is rather similar to a log abundance score. There is a danger that the observations will fall into the same category so that the observed variance is zero even when the plots differ in detail. This is less likely when there are more replications.

5. *Generalisation to other sites*

There is a general problem in experimentation in forming inferences which apply to a wide class of areas and environments.

The analogy with designed experiments is that in an isolated experiment we use local controls and express all our inferences in terms of differences (improvements, changes, responses) rather than in terms of absolute values. If we wish to study differences between years or sites we have to analyse a group of experiments, and estimate

1) an overall mean value.
2) the variance or distribution of mean values between experiments.
3) an overall response or mean effect of treatments.
4) the variance of the response as estimated from the pooled errors.
5) the interaction between response and site, or the variability of the response over different sites.

A mean value will have two independent sources of variation,

1) σ^2_e or the variance within experiments.

2) σ^2_b or the variance between experiments.

The overall variance $\sigma^2_e + \sigma^2_b$ is thus a function of the number of experiments (for σ^2_b) and the number of plots per experiment (for σ^2_e), and the optimum allocation if we have a fixed number of plots that can be used (or can be afforded) depends on the relative importance of these two sources of error. If we have too many small experiments we will not be able to measure the responses accurately, whereas if we have too few then the overall variance may not be measured accurately. If the interaction between response and experiment is large we will also need more sites.

In the present context it is not possible to say exactly whether it is preferable to use 2 × 6 or 6 × 2 or 3 × 4 plots, without some idea of the variability to be expected, or if the importance of the different kinds of question. But without some replication you cannot make any inferences about a wider class of sites.

6. *Effect of discrete events (felling)*

This has been discussed before. If the effect of felling is permanent, and the curve represents a change from one plateau level to another, one could compare the mean of all years before felling with the mean of all after. Or one could analyse the total number of species in either of two years, to increase the sample size somewhat.

But if the effect is temporary the plot may recover over time, and the difference in the years immediately closest to felling might be most significant. A graphical presentation is important in making these judgements.

DATASET 3

COTSWOLDS PLOTS

1. An area of woodland (10 ha) is likely to be thinned and we want to measure the response of the ground flora. In this instance, eighty temporary 5 x 5 plots have been recorded at random through the area. This process was repeated in successive years with a fresh set of samples each time. Species cover has been recorded. As in the previous set, our concern is with changes in species-richness per plot and in the frequency and mean cover of individual species.

2. What is the best approach to comparing mean species-richness over the years, assuming one year's "before thinning" records and 2-5 "post thinning" records? If there is more than one year's pre-thinning, how does this affect the approach? There may not be recording every year. What can be done if there is a break in the time series - how does this alter the analysis? In one year only forty rather than eighty plots were recorded (but still an independent random sample of the whole area). Does this matter in terms of what analysis is possible?

3. The usual questions about the increase or decrease in individual species frequency and cover arise. Curves showing change in the frequency or mean abundance of each species might be produced and compared. However since different plots are recorded each year and a large number of species are involved, the likelihood that some changes are purely sampling effects increases.

4. All the species in a plot could be classified into one of Grime's life categories (Ruderals, Competitors, Stress-tolerators) and the mean number per plot for each year calculated. Can trends in the number of "ruderals" be examined independently of the other two, then "competitors" independently of the other two and then "stress-tolerators"? My feeling is that this depends on the size of the quadrat to some degree: in a very "large" quadrat, space is not limiting the number of species present, so "ruderals" could go up independently of changes in the other groups, at the smaller scales, there are very significant interactions between species, so that changes in one group almost have to involve a corresponding decrease in another.

Thinning ↓

Year:	1987	1988	1989	1990	[illustrative data only]
Mean no per plot	2.0 ± 0.3	2.9 ± 0.4	3.0 ± 0.4	3.8 ± 0.4	
No of plots recorded	80	80	40	80	
No of plots with :					
Rubus	21	33	19	37	)
Dryopteris	10	8	8	17	) and so on for every other
Hedera	5	4	2	3	) species (20 in total)
Fagus (seedling)	31	37	20	45	)

Only ½ density of sampling

Data Set 3

Cotswold Plots

1. *Sampling Scheme*
A discrete event (thinning) separates the series of observations into two phases. 80 random temporary plots are chosen each year, with fresh samples each time. The randomisation scheme is not described.

Species cover and species richness are recorded. Since the plots are small the number of species is also small (means given range from 2 to 38) so some plots may be all a single species such as Rubus.

For spreading species such as Rubus we have a direct estimate of the percentage of the total area covered, as given by the mean cover for each sample. The distribution is presumably patchy, but the variance of the estimate may be completed directly.

For separate individuals such as Fagus seedlings, one would expect their total counts, if available, to follow a Negative Binomial distribution, or similar contagious distribution, in which the mean frequency and a measure of aggregation may be estimated. The aggregation is relative to the excess variance over the theoretical Poisson variance if all seedlings were randomly distributed over the whole area.

Presumably, being woodland floor, there is a varying proportion of litter/bare ground not contributing to the species count.

2. *Comparison over years*
The statistic 'mean species richness' has been discussed under Data Set 2. In this data set the distribution is clearly very skew, and using the illustrative data to estimate the variances of the individual counts, we find for 1987 the mean is 2.0 and the variance is 7.2 (= 80 × .09), while for 1990 the mean is 3.8 and the variance is 12.8, although the rounding error is too great to allow much to be deduced from this. However, these figures suggest that a square root transformation would be appropriate when testing the significance of changes of mean species richness. (See General Discussion.)

The difference in sampling numbers from 80 to 40 can be taken account of in the two sample t test.

Any pair of years may be compared. If more than two years are to be compared, and one is prepared to regard the effect of thinning as static, then it might make sense to test 1987 against the weighted mean of the three years 1988-1990. However, if it is assumed that the changes will be dynamic and the flora will develop over time after equilibrium has been distributed by thinning; then it would be necessary either to look at individual years (the first year at which significant differences are recorded) or to test for a trend. It is unlikely that any trend would continue to be linear for a long time, as more likely a new mean level would be reached or approached, so that an exponential or logistic curve might be expected to describe the rate of change.

The break in the series does not matter unless one is interested in serial correlations, the relationships between changes between successive years when trends have been accounted for. Time series analysis is hardly possible over such a short period.

3. *Analysis of individual species*
Since the plots are sampled independently there is no objection to applying standard binomial tests to the proportion of plots on which a particular species occurs.

For example we have for Rubus, in 1987 and 1989

Differences in proportions = 19/40 - 21/80 = 17/80 = 0.2125
Variances of difference = $(19 \times 21)/40^3 + (21 \times 59)/80^3 = .00865 = (.093)^2$
Students t (on 120 d.f.) = 2.285

So we would deduce that there was a just significant increase in frequency of Rubus.

Curves of mean abundance can certainly yield graphical information about the nature of the changes observed. The problem is to find suitable curves with sufficiently simple parametric form to express the typical shapes observed. After a number of years the new equilibrium proportions may be estimated, pooling data over several years if necessary.

4. *Use of species categories*
If the Grime Categories mean what they say then we would expect different behaviour from each of the three main categories, so it would make sense to analyse each separately.

The problem of small sample plots does mean that an individual plot with a dominant species will exclude others. But the overall statistics which refer to the distributions over the whole site will simply have these effects reflected in their variance, which would be higher than if all species behaved independently.

Larger plots would include more species, but unless there were the same number of plots, there would be a decrease in precision of estimates of overall mean abundance.

DATA SET 4

POTATOPOT GRASSLAND TRANSPLANT, 90 x 30M PLOT

Background

A programme of grassland transplantation monitoring was started in 1987 by the England Field Unit of the NCC at 8 sites in England, in order to indicate whether such attempts are 'successful' according to nature conservation criteria. Potatopot is one of the 8 sites and is an area of acid grassland with a range of grassland and mire communities as defined by the National Vegetation Classification. Part of the grassland was moved before development of the site by British Coal.

Objectives

The overall aim has been to see if the species assemblage which was present before the transplantation remains intact in its new location or can recover to its previous composition. Alternatively the grassland may lose species, resulting in a less rich assemblage or species may invade that are of lower conservation interest.

The specific objectives are to compare data from before and after transplantation to assess:

a) If individual species change significantly in their relative abundance, measured by frequency of occurrence in the plot.

b) Which species appear and disappear.

Sampling Design

One plot measuring 90 by 30 metres was marked out in 1988, before transplantation occurred. The plot was subdivided into 9 10m x 30m strips. Within each strip, 20 'mini' 10cm x 10cm temporary quadrats were randomly located and species presence recorded in each quadrat. A total of 180 mini-quadrats were recorded. In 1989 the sampling was repeated in each strip but this time 24 mini-quadrats per strip were recorded giving a total of 216 quadrats. After the recording in 1989, the entire plot was transplanted. The integrity of the strips was maintained so that Strip 1 in 1990 had the same turf as Strip 1 in 1988 and 1989. In 1990 and 1991 24 quadrats per strip were again recorded. The intention is to gather 3 years post transplant data (1990, 1991 and 1992) and then re-record at longer intervals if resources allow.

Information Collected

The attached data sheet illustrates the information recorded in each mini-quadrat i.e. Species present
Presence of litter
Presence of bare ground
Vegetation height
Total number of species per quadrats
The species which has the greatest cover in a quadrat (can be one or more species)

From this information species frequencies (number of quadrats per strip with a species/total number of quadrats per strip) have been calculated.

Statistical analysis to date has investigated changes in individual species frequencies i.e. all species that have been recorded in the plot have been subject to analysis simultaneously.

The summary data is the form:

(Example only)
Species : Ranunculus acris

Strip	1	2	3	4	5	6	7	8	9
% frequency 1988	50	30	44	60	52	47	61	58	39
% frequency 1989	50	25	51	41	59	34	55	62	36
% frequency 1990	25	16	11	22	19	17	25	31	14
% frequency 1991	5	2	0	10	9	6	3	0	2

Data has been analysed using ANOVA to compare 1988 with 1989 and 1989 with 1990. The frequencies were transformed (arcsinesquare root transformation) before analysis. The ANOVA table was:

Source	df
Strips	8
Years	1
Error	8
Total	17

$\frac{\text{ms years}}{\text{ms error}}$ - F value

Questions

1. Is the statistical technique analysis appropriate? If so, would a comparison of strips ($\frac{\text{ms strips}}{\text{ms error}}$ F value) also be legitimate?

2. Does it matter that % frequencies were calculated from different sample numbers i.e. 20 quadrats in 1988 and 24 in 1989-91?

3. Can more than 2 years be included in the 'treatments'? If so can contrasts be constructed e.g. pre-transplant year(s) versus post-transplant year(s).

4. What other statistical techniques might be used?

5. Can all species be looked at simultaneously? What limitations on statistical tests might be necessary?

Frequency of species within a sample area

Recorders: J. COX

Site: POTATO POT Date: 2/7/90

Sample area: 90 X 30m PLOT

STRIP 1 (16 out of 24 quadrats on this sheet)

Quadrat number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
Species / Random numbers	02	49	91	163	248	283	288	160	C7	23	25	170	23C	256	257	293					16
AGROSTIS CANINA	⊘	✓	✓	✓		⊘				✓	⊘	✓	✓		✓	✓					11
AGROSTIS CAPILLARIS							✓														1
FESTUCA RUBRA	✓		✓		✓	✓	✓							✓	✓	✓					8
HOLCUS LANATUS	✓	⊘					✓									⊘					4
CAREX PANICEA			✓	✓		✓	✓	⊘													5
LOTUS ULIGINOSUS	✓	✓																			2
NARDUS STRICTA	✓							✓	⊘												3
JUNCUS CONGLOMERATUS	✓	✓					✓				✓				✓	✓					6
CAREX NIGRA	✓	✓		✓		✓		✓			✓				✓						7
DESCHAMPSIA CESPITOSA	✓	✓							✓	✓											4
POTENTILLA ERECTA		✓	✓	✓				✓	✓		✓	✓	✓	✓	✓						10
SUCCISA PRATENSIS		✓	✓	✓	⊘		⊘			⊘	✓	⊘		✓							9
PLANTAGO LANCEOLATA			✓	✓												✓					3
CAREX FLACCA		✓					✓						✓								3
POA PRATENSIS		✓																			1
FESTUCA OVINA			⊘	⊘	✓	✓	✓						⊘	✓	⊘						8
GALIUM SAXATILE				✓																	1
MOLINIA CAERULEA					✓				✓	✓	✓	✓		⊘							6
LOTUS CORNICULATUS					✓																1
CIRSIUM PALUSTRE								✓													1
ANTHOXATHEMUM ODORATUM										✓		✓				✓					3
JUNCUS ACUTIFLORUS											✓										1
CAREX PILULIFERA												✓									1
DANTHONIA DECUMBENS												✓									1
Litter	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					16
Bare ground									✓												1
Total number of species	8	10	7	8	5	5	6	5	4	5	7	7	4	5	6	6					
Vegetation height	35	23	22	25	30	28	18	19	27	28	30	19	20	17	13	17					

Key: presence of mature plant ✓ | presence of seedling s | presence of mature plant and seedling ✓s | species with greatest cover ⊘

Record bare ground and litter as amounts visible from directly overhead and as '50' if cover 50 % or more

DATA SET 5

POTATOPOT GRASSLAND TRANSPLANT, BLADED PLOT

Background and Objectives

As for Data set 2. In this case however, no pre-transplant monitoring was done of the grassland to be moved. The source of the material was grassland and topsoil scraped up before the development occurred and spread out on a receptor site. The objective was to see how species composition changed as vegetation became re-established.

Sampling design

Scraped up material was placed on a 20 by 30 metre plot. Within this plot, 50 temporary randomly-located mini-quadrats (10 x 10 cm) were recorded as for the 90 x 30 m plot. A sample data sheet is attached. It has been proposed that the species frequency changes be analysed using X^2 tests, for example:

Ranunculus acris	present	absent	total
1989	32	18	50
1990	12	38	50

or as in the test described in Appendix 9 of EFU report 103 (attached) where 'absent' quadrats are not included.

Questions

1. Are either of these analyses appropriate?

2. Are there other techniques that would be more illuminating?

3. The question of simultaneous tests raised for Data Set 4 is also applicable for this data set.

Frequency of species within a sample area

Recorders: C. BLAKE

Site: POTATO POT Date: 3/7/90

Sample area: BLADED PLOT A

(20 out of a total of 50 quadrats)

Quadrat number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
Species / Random numbers	2 10	26 0	1 3	8 17	19 4	9 4	12 3	17 12	24 8	4 0	12 15	16 16	24 16	17 18	18 5	6 9	11 6	26 17	14 2	19 9	20
CAREX PANICEA			⊘				✓	✓			✓	✓				✓					6
JUNCUS ACUTIFLORUS		✓	✓									✓	⊘					✓	⊘		6
FESTUCA OVINA			✓										✓					✓			3
ULEX EUROPAEUS			s		s		s														3
CAREX DEMISSA	✓			✓				✓	✓		✓	✓		✓				✓		✓	9
AGROSTIS CAPILLARIS	✓	⊘		✓		✓			✓			✓		✓	✓	✓	✓	✓	✓		12
PLANTAGO LANCEOLATA	✓	✓		✓		✓		✓						✓		✓		✓	✓		9
AGROSTIS CANINA	✓	✓		✓	✓	⊘		✓	✓	✓	✓		✓	✓	✓	⊘		✓		✓	15
POTENTILLA ERECTA	✓	✓					✓						✓			✓			✓		6
HOLCUS LANATUS		✓				✓	✓			✓			✓				✓	✓	✓		8
EQUISETUM ARVENSE										✓											1
RUMEX ACETOSA							✓			✓											2
SUCCISA PRATENSIS																✓					1
RANUNCULUS REPENS		✓				✓			✓												3
ANTHOXANTHUM ODORATUM							⊘														1
DESCHAMPSIA CAESPITOSA																	✓				1
CAREX NIGRA								✓													1
CIRSIUM PALUSTRE		✓						✓													2
MOLINEA CAERULEA								✓													1
HYPERICUM PULCHRUM																				✓	1
RANUNCULUS FLAMMULA																				✓	1
CAREX FLACCA					✓							✓							✓		3
CIRSIUM ARVENSE																			✓		1
JUNCUS ARTICULATUS											✓										1
ANGELICA SYLVESTRIS														✓							1
HYPOCHAERIS RADICATA												✓									1
TRIFOLIUM REPENS												✓									1
ACHILLEA PTARMICA		✓																			1
Litter																					
Bare ground	50	✓	✓	50	50	✓	✓	50	50	50	50	50	✓	50	50	✓	50	50	✓	50	20
Total number of species	5	9	4	4	3	5	6	7	4	4	4	7	5	5	2	6	3	7	7	4	
Vegetation height	1	10	11	2	4	5	24	3	3	3	2	3	10	3	4	3	5	7	35	41	

Key: presence of mature plant ✓; presence of seedling s; presence of mature plant and seedling ✓s; species with greatest cover ⊘

Record bare ground and litter as amounts visible from directly overhead and as '50' if cover 50 % or more

Appendix 9: Worked example of chi-squared test used to assess significance of changes in species frequency between samples from different years.

The test is only suitable for data from strictly randomly located samples; not for data from non-randomly located samples or from restricted randomly located samples.

The null hypothesis (Ho) under test is that the proportion of quadrats in which a species is present is the same in each year.

1 For each species make a table of OBSERVED values, including total number of quadrats recorded and number of quadrats in which the species was present.

Year	No of quadrats with *Primula veris*	Total no of quadrats
1	36	50
2	8	100
3	2	100
Total	46	250

2 Calculate EXPECTED values for each year.

Expected value (E) for year 1 =

$$\frac{\text{total quadrats in which Primula present}}{\text{overall total number of quadrats}} \times \text{total quadrats in year 1}$$

so expected value here for year 1 is $\frac{46}{250} \times 50 = 9.2$

expected value in year 2 is $\frac{46}{250} \times 100 = 18.4$

expected value in year 3 is $\frac{46}{250} \times 100 = 18.4$

Data Sets 4 and 5

Potatopot Grassland transplant

1. *Validity of Analyses of Variance*
The proposed analysis is to record the proportion of plots in which each species occurs (p) and to use the angular transformation (arc sine square root (p)) before analysis.

The angular transformation is used to equalise variances when the observation has a binomial distribution with fixed sample size but variable mean. It also improves the additivity and linearity of the scale of measurement, so that systematic differences caused by treatments or blocks may be more equally treated. This latter consideration is usually more important than the variance stabilising property, particularly when observed proportions are below 20% or above 80%.

The assumption of a binomial distribution may not be strictly valid, in particular for presence-absence data, where the criterion of presence is not symmetrical with absence. If the species occurs as random or aggregated groups of individual plants at a particular density, we require the probability that no plant occurs within a given plot. If the species has a spreading habit, we require the probability that a random plot does not overlap any of the regions covered by the species. In the aggregated case we would expect the variance of the proportion present to exceed the theoretical binomial variance. This does not however invalidate the angular transformation, but would generally result in a residual mean squares in excess of its theoretical expectation for a pure binomial model.

To test for variation between years is quite valid, particularly as different plots are chosen on each occasion. Before transplant we would only expect non significant random effects (unless a particular season was adverse for the species in question). Comparison before and after transplanting is clearly of interest, and in the example given there is no doubt that a significant effect would be shown (a non-parametric test giving 10/10 changes in the same direction is highly significant).

The test for differences between strips is valid, but possibly less interesting, because the strips are identically treated, and we are only estimating a possible excess of spatial variation as the area sampled increases.

2. *Difference in sample numbers*

The difference between binomial samples of sizes 20 and 24 is unlikely to affect the validity of the analysis of variance. If the ratio was more than two to one it might cause a small bias in the F ratios, but there are much larger effects (such as due to deviations from the pure binomial distribution)which are unimportant.

However, it is now quite simple to avoid the problem altogether by fitting a Generalised Linear Model (as provided by the programs GLIM or GENSTAT) which takes into account different sample sizes. The analysis is slightly less straightforward because the tests of strips and years are no longer completely independent, but there is greater freedom to choose different transformations, such as the logit or probit transformation instead of the angular transformation.

3. *Contrasts for more than two years*

A single contrast for comparing groups of years may easily be constructed, and its contribution separately recorded in the analysis of variance. If four years are used we can divide the three degrees of freedom for 'between years' into three orthogonal contrasts

$$\tfrac{1}{2}((y_1 + y_2)-(y_3 + y_4)),\ (y_1 - y_2) \text{ and } (y_3 - y_4)$$

which may be tested independently. For different numbers of years it may be necessary to check the orthogonality carefully, although it is often of sufficient interest to remove a single contrast such as (mean of earlier years) - (mean of later years), in which case it is not necessary to partition the remainder orthogonally.

Alternatively, it may be appropriate to use a parametric curve, such as the logistic curve (or a straight line if changes are confined to a small range). This can be done quite easily using the GLIM or GENSTAT packages.

4. *Other statistical techniques*

The main alternative to the analysis of variance with transformed percentages is the Generalised Linear Model analysis referred to above. In this analysis the choice of the response scale is separated from that of the distribution of the random variation, and the significance tests are based on the theory of Likelihood rather than of Least Squares.

The main features of the theory are described in the program manuals for GLIM and GENSTAT, and in the text by McCullagh and Nelder (1982, 1990). (References 12, 51, 52, 54.) (See General Discussion, Section 2.)

In the present context it might well be appropriate to use the logit scale for transforming the percentages, using the binomial error model. The analysis of variance becomes an 'analysis of deviance', and terms may be tested against the chi-squared distribution, although if the residual deviance is large this is an indication that the variance is greater than expected for 'pure' binomial error, in which case the ratios of deviances are compared with F distribution, as for a standard analysis of variance.

5. *Analysis of all species simultaneously*
Cluster analysis applied to the individual plots would produce too much detail, much of which is subject to statistical fluctuation. But a multivariate analysis of strips within years, in which the variables are the percent frequency of each species, together with the mean scores for litter and bare ground, would provide the basis for an overall study of the main contrasts in composition and how they change over years. This is similar to the analysis of agricultural acreages in English Counties shown in Appendix 1.

The analysis could combine cluster analysis and ordination, and by relating the major components of contrast to groups of species the overall changes with time can be illustrated.

5.1 *Bladed Plots*
1. The standard binomial test for comparing proportions of presence/absence of a species on two different occasions is equivalent to the 2 × 2 table chi-square test.

There may be substance to the argument that presence and absence is not a symmetrical relationship, but the analysis omitting the 'absent' column, as suggested in Appendix 9 of the EFU report, is incorrect, as the calculated chi-squared does not follow the chi-squared distribution. If the proportions of 'presence' are very small compared with 'absence' it may be found that the additional contribution of the omitted terms is small, but that does not apply to the example in their Appendix 9, where the first year records 36/50 present. The analysis given is in fact using the Poisson approximation to the binomial distribution, and will tend to be too conservative, rejecting the null hypothesis on too few occasions.

2. The proportions of each species may be displayed graphically on a labelled plot in which the proportions in year are plotted against those in year 2. Guidelines for significant differences can be drawn on these plots, and species which are gaining in prominence will appear above the 45 degree line, while those that are declining appear below the line.

It is possible to include further years on these plots, joining the points for the same species with arrows showing the possibility of recovery or further change in the same direction.

3. As there is only one plot it is only possible to do a multivariate analysis of single plots (labelled by year) in which plots do not correspond in pairs. The contrast in years is therefore represented by the change in the clusters or labelled ordinated points, which should reproduce the information already found by individual analysis.

It can be argued that the cluster analysis is unnecessary if the full analysis of each species has already been done. If, however, it is done first, it quickly identifies those species in which it is worth proceeding with the full statistical analysis.

DATA SET 6

JNNABRIDGE AND SCRATCHY BOTTOM GRASSLANDS

Background

Dunnabridge is a meadow (NVC type MG5) and Scratchy Bottom is a calcareous grassland (CG4). Recently, management of both sites has changed. At Dunnabridge, cattle grazing all year was replaced 3 years ago by a regime of hay cutting followed by 'aftermath' grazing in the autumn. At Scratchy Bottom, management by cattle grazing has been introduced in 1991.

Objectives

The overall aim is to assess the impact of change in management on species assemblages, but not all species will be studied. Broadleaved herbs and one or two grass species have been selected as the species most likely to indicate positive or negative effects of management.

Sampling Design

Dunnabridge.

A 40 by 30 plot was chosen within the meadow, and divided into 3 strips (see map attached). It was felt that the impact of management might be seen as grading out from the gateway to the field so the strips were aligned more or less perpendicular to this assumed gradation. Within each strip, 10 randomly located temporary quadrats were recorded. Each quadrat was subdivided in a 'nested' way into units of 10 x 10cm, 20 x 20cm, 50 x 50cm and 100 x 100cm (see diagram on map sheet attached). Subdivisions were inspected in order $10cm^2$, $20cm^2$, $50cm^2$, $100cm^2$ and species present were recorded. If a species was found in the $10cm^2$ quadrat it was therefore recorded as present in all the other subdivisions. (See data sheet example and tabulation of species frequencies).

Scratchy Bottom.

A 100 by 70m plot was chosen within the site and randomly sampled with 35 nested quadrats which were searched in the same way as at Dunnabridge. The plot was laid out perpendicular (on its shorted side) to the slope of the ground (see map attached).

Additional monitoring at Dunnabridge

In parallel with the above approach, another sampling scheme is being continued. It was begun when management of the site changed 3 years ago and consists of recording species present in 30 temporary quadrats, 50 x 50cm in size which are located by 'throwing over the shoulder' technique!

Information Collected

The only data collected are the presence of particular species within the subdivisions of the nested quadrat (see sample data sheet).

Analysis

Depending on frequencies recorded for particular species at particular sub-division sizes, analysis of changes in these frequencies through time will be wanted, in a similar way as for Data sets 4 and 5. It is hoped that the 'nested' design will allow choice of appropriate quadrat size for

species which are distributed with different patterns e.g. sparse throughout, a few large clumps etc. so that for instance Ranunculus bulbosus would be analysed using the 10 x 10cm sub-divisions each year while Platanthera chlorantha would be analysed using the 50 x 50cm sub-divisions.

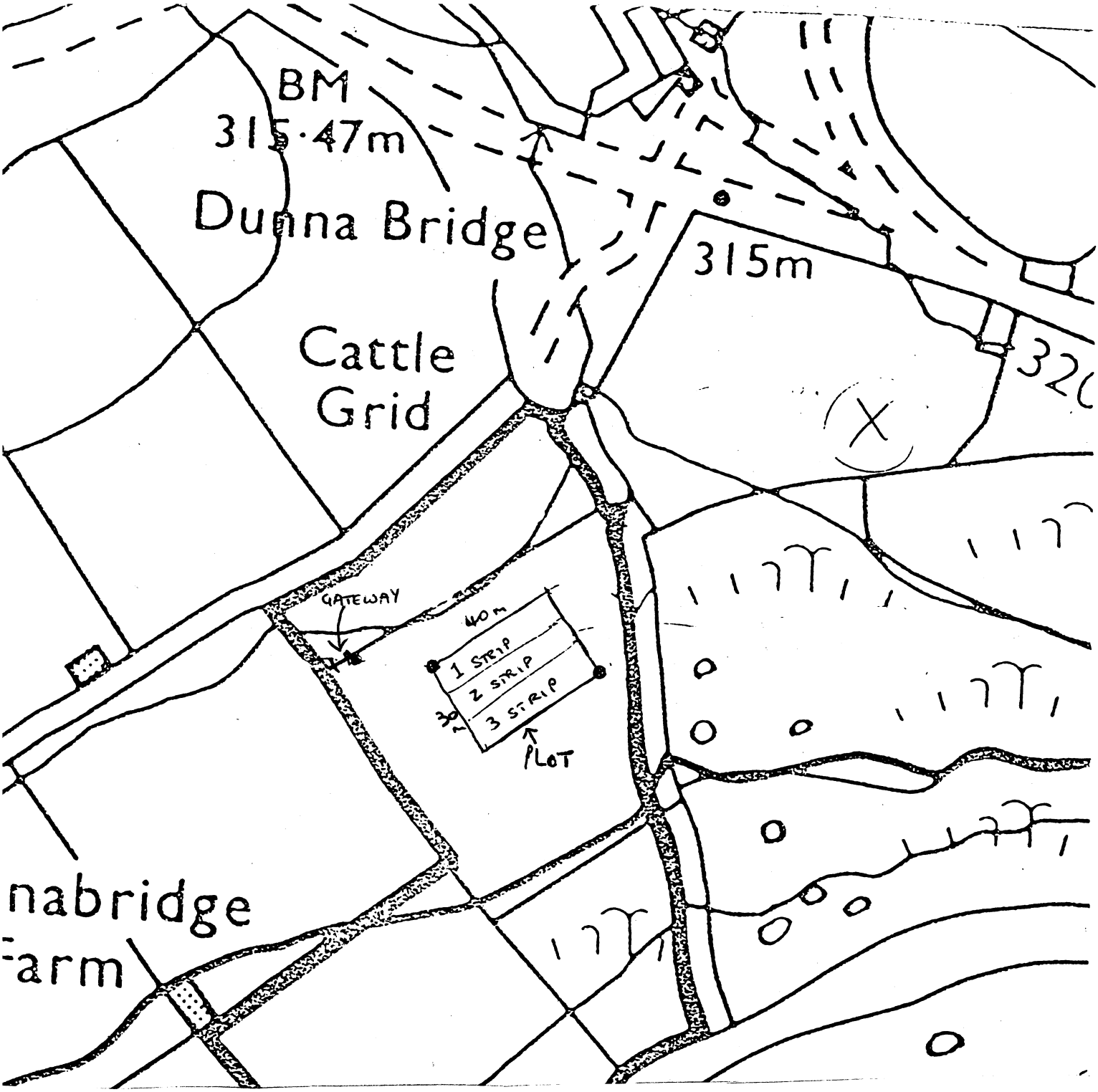
Another option might be to try to analyse the change in pattern of particular species. Curves produced by graphing increasing sub-division size (log possibly) against cumulative number of occurrences in each sub-division might be compared between years.

Questions

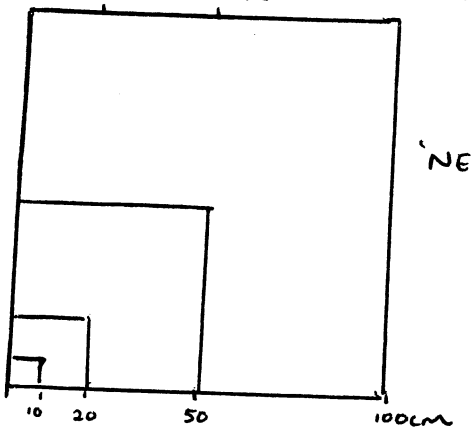
1. Are the two proposed approaches to analysis acceptable for statistical treatment and what methods of analysis would be suitable?

2. Are there ways of deciding beforehand if a completely randomized design versus a restricted randomized design would be more 'efficient' apart from using informed judgement. A restricted randomized design could have been used at Scratchy Bottom for instance, to take account of potential variation down/upslope.

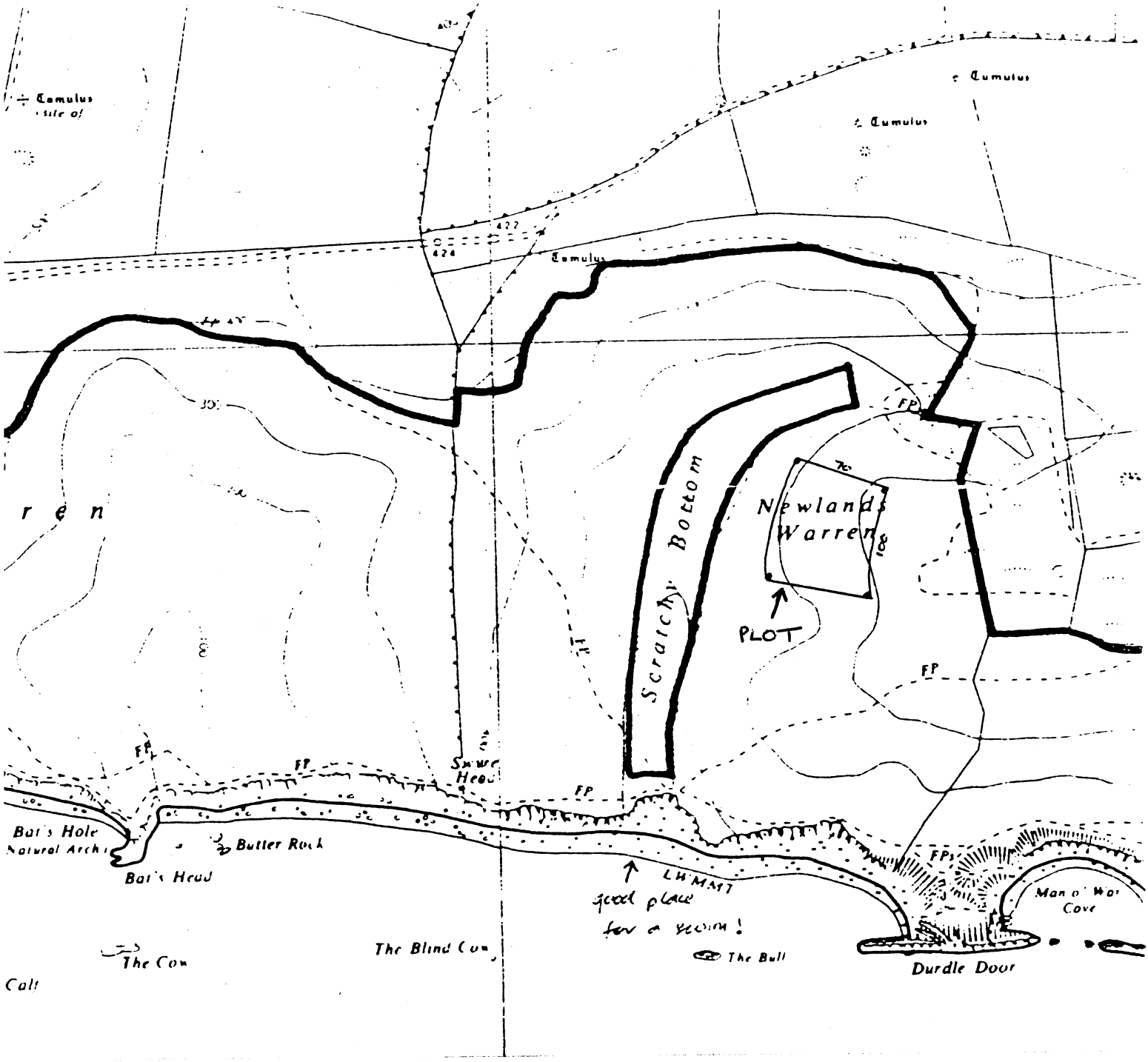
3. Can anything be inferred from the second monitoring scheme at Dunnabridge?



LOCATION OF PLOT AT DUNNABRIDGE, SHOWING ARRANGEMENT OF STRIPS.



'NESTED' QUADRAT



LOCATION OF PLOT AT SCRATCHY BOTTOM.

Frequency of species within a sample area

Recorders: SIMON LEACH

Site: DUNNABRIDGE MEADOW Date: 20/6/91

Sample area: STRIP 1

Quadrat number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
Species / Random numbers	19 8	25 5	15 8	7 3	18 8	21 2	22 1	24 5	14 6	23 9											
RUMEX ACETOSA	1	1	2	1	2	2	2	1	1	1											
RANUNCULUS ACRIS	1	1	2	1	1	1	1	5	2	2											
TRIFOLIUM DUBIUM	2	1	5	1	1	2	5	5	1	2											
HYPOCHAERIS RADICATA	1	2	2	2	2	9	5		5	1											
CERASTIUM FONTANUM	9	2	5	2	9	2	9	9	5	2											
RANUNCULUS BULBOSUS	1	1	1	1	1	2	2	2	1	1											
CREPIS CAPILLARIS	2	1	5	1	2	1	1	2	5	5											
STELLARIA GRAMINEA			5							5											
TRIFOLIUM PRATENSE	2	1	5	9	2	2	1	1	1	2											
ACHILLEA MILLEFOLIUM				9					9	5											
RHINANTHUS MINOR	1	1	1	9	1	2	1	1	2	1											
PLANTAGO LANCEOLATA	1	1	1	5	1	1	1	1	1	1											
EUPHRASIA SPP.			9	9																	
PLATANTHERA CHLORANTHA	5				2		2	9	9	2											
CENTAUREA NIGRA			2			9	9	1	2												
PRUNELLA VULGARIS		9																			
LOTUS CORNICULATUS							2		9	1											
HERACLEUM SPHONDYLIUM									9												

KEY:

1 = PRESENT IN 10cm x 10cm QUADRAT

2 = " " 20 x 20cm "

5 = " " 50 x 50cm "

9 = " " 100 x 100cm "

DUNNABRIDGE MEADOW

Part of Tabulation of Species Occurences in Nested Quadrats (10 nested quadrats in each strip, each quadrat with 4 nested subdivisions).

Strip	1				2				3				All Strips			
Species \ Subdivision size in cms	10	20	50	100	10	20	50	100	10	20	50	100	10	20	50	100
RANUNCULUS BULBOSUS	7	10	10	10	6	10	10	10	9	10	10	10	22	30	30	30
RUMEX ACETOSA	6	10	10	10	7	10	10	10	8	10	10	10	21	30	30	30
RANUNCULUS ACRIS	6	9	9	10	4	7	9	10	4	7	10	10	14	23	28	30
CERASTIUM FONTANUM	0	4	6	10	1	3	5	9	1	1	3	8	2	8	14	27
TRIFOLIUM DUBIUM	4	7	10	10	3	7	10	10	6	8	8	10	13	22	28	30
TRIFOLIUM PRATENSE	4	8	9	10	7	9	10	10	5	10	10	10	16	27	29	30
PLANTAGO LANCEOLATA	9	9	10	10	10	10	10	10	9	10	10	10	28	29	30	30
RHINANTHUS MINOR	7	9	9	10	6	9	10	10	6	9	10	10	19	27	29	30
CREPIS CAPILLARIS	4	7	10	10	3	5	7	10	0	4	6	9	7	16	23	29
HYPOCHAERIS RADICATA	2	6	8	9	1	2	7	8	3	6	7	10	6	14	22	27
PLATANTHERA CHLORANTHA	0	3	4	6	0	3	4	6	0	1	2	5	0	7	10	17
LOTUS CORNICULATUS	1	2	2	3	1	1	1	3	0	0	0	1	2	3	3	7

Discussion Data Set 6

Dunnabridge and Scratchy Bottom Grasslands

Background

The study aims to assess the impact on vegetation of changes in grassland management, from continuous grassing by cattle to hay cutting and aftermath grazing (Dunnabridge) or return to cattle grazing (Scratchy Bottom). The sampling area shows potential systematic trends which may be important.

1. *Proposed statistical analysis*

The data recorded are presence within subdivisions of nested quadrats, labelled 1, 2, 5 and 9. This scheme is used to save time, and 9 is the rarest category (expect 0 for absence).

For a given species the mean presence per square metre can be estimated approximately by the formula

$$\frac{100n_1 + 25n_2 + 4n_5 + n_9}{(n_1 + n_2 + n_5 + n_9 + n_0)}$$

where n_0 is the number of plots on which the species is absent. Use of this score to compare strips and years for each species would seem to be preferable to the choice of a particular grid size for each species.

The probability of observing significant changes in the proportions of 10 samples is not very great, so method 1 (choice of appropriate size) is presumably intended to use that part of the scale where change is most apparent. In any one year therefore, significant differences between two strips (r/10 compared with s/10) at the 95% level only occur if r and s are as separate as in the following table (taken from Siegel's 'Non parametric Statistics', Table 1)

r	0	1	2	3	4	5	6
s	4	6	7	8	9	10	10

and even greater ranges are necessary to demonstrate differences between three strips. Since the numbers are small, exact distributions are required to assess significance, and the computation for three or more strips are lengthy (expensive software is available to work it out, but this hardly seems justified in the present context).

The alternative proposal to graph cumulative number of occurrences against log size does not seem to be very useful, compared with the composite score suggested above. The data are not independent, and it would not be particularly practicable to make the suggested comparison between years.

2. *Design considerations*

There is scope for stratifying the sampling, since there are a *priori* reasons for choice of site and aspect, or relation to the gate at Dunnabridge. Block differences can then be analysed, and if necessary trends associated with gradient or distance can be incorporated into the analysis.

The argument for blocking versus complete randomisation is that when blocks are suitably chosen, the loss of degrees of freedom in estimating the sampling error is compensated for by the increase in efficiency due to the systematic component being separately estimated, leaving a smaller residual variance. If blocks are poorly chosen and are not significant, it may be valid to abandon the block component and to pool the two sources of error estimate.

3. *The second monitoring scheme at Dunnabridge*

The 'over the shoulder' technique of randomisation has been criticised as liable to produce non-random distributions of samples. This does not make the data totally valueless, particularly if it is historical data that cannot be reproduced. But inferences for statistical tests will have to be qualified by the statement that the samples were not strictly randomly distributed, and if the terrain is known to be very heterogeneous it would be necessary to check whether the samples were disproportionately in one part of the site.

DATA SET 7

NG TERM MONITORING OF CALCAREOUS GRASSLANDS

Background

English Nature has access to data collected on its National Nature Reserves for a project which is aimed at monitoring long term change over a wide range of calcareous grasslands across Britain. The method and examples of data collected are attached.

Method p.1-5
Example of data p.6-10
MATCH coefficients through time, Castor Hanglands NNR p.11
Calculation of MATCH coefficients p.12.

The method follows that described in Smartt, P.F.M. and Grainger, J.E.A. (1974) "Sampling for Vegetation Survey : Some aspects of the behaviour of unrestricted, restricted and stratified techniques" Journal of Biogeography Vol. pp 193-206.

Objectives

To assess whether individual species change significantly in their frequency in a plot over time and also whether community type, as defined by the NVC, changes over time.

Analysis

To analyse individual species frequency changes some kind of t-test or G-test has been suggested, for example a frequency of Briza media of 50% in 1991 (18 quadrats out of 36) compared to 75% in 1995 (27 quadrats out of 36).

To analyse NVC type changes, the calculation of similarity coefficients at different dates is proposed. MATCH is a computer program which compares sample data with NVC data and calculates a similarity coefficient for the sample compared to all NVC communities (see page 12). The top 10 similarity coefficients are then listed by the programme. If the species composition of the site changes sufficiently over a period it will come to resemble another NVC type rather than the original one. Page 11 shows the similarity coefficients of 3 plots with 3, 4 or 5 NVC communities at Castor Hanglands NNR.

Questions

1. What tests would be appropriate if any for assessing whether individual species frequencies have changed?

2. Is some kind of statistical analysis possible of trends in MATCH coefficients or the significance of their increases/decreases from one time to another?

RECORDING METHOD

100CC PLOTS, or, 100 years of Climate Change Plots

As far as is practicable, areas with homogeneous vegetation representative of calcicolous grasslands as described in *British Plant Communities* (NVC), with security of ownership and stability of management (especially grazing regimes) are selected. Plots are also selected for ease of relocation, to avoid areas being used for other studies and so that they do not require special attention (eg scrub removal). Plots are also located in context with other plots on the site to give a range of aspects, altitude, etc.

Plots are generally located near, but not immediately adjacent to, obvious geographical features such as walls, boulders, etc, so that they can be relocated relatively rapidly. Photographs, sketch maps and measurements are taken to aid relocation of the plot markers. A 12 x 12 m grid (rarely 6 x 24 m) is laid out with tapes so that the quadrats can be positioned rapidly. Grids are consistently orientated up-down slopes or north-south on the flat. The position of the grid is marked permanently with seven loops of insulated copper wire, buried 5-15 cm deep in the soil; these can be readily relocated using a metal detector. Copper loops are preferred to other metal markers because they give strong signals to current inducing metal-detectors, they are small and relatively unintrusive (roots can grow through them), and are less susceptible to frost heave than plates. The insulation should provide protection from corrosion for at least 20 years. The markers are placed in a consistent pattern at unique distances so that the plot can be repositioned even if up to five markers are lost.

Small soil samples are taken from the points at which the copper marker loops are buried to avoid disturbance elsewhere, and the profiles briefly described. pH measurements are made in the field using a pH meter, and will be backed up with more standardised laboratory measurements.

In simple terms, the vegetation is recorded using 36 0.5 x 0.5 m quadrats (0.25 m^2)(Figure A). In more detail, each plot is divided into four 6 x 6 m blocks, to allow flexibility of plot shape if needed. Each block is then subdivided into nine 2 x 2 m units, each of which is sampled using a 0.5 x 0.5 m quadrat. The nine quadrats in each block are positioned using a stratified systematic unaligned arrangement; the unalignment was initially determined using randomly selected coordinates. The same sampling pattern is used for each plot. The 36 quadrats are adequate replication for statistical analysis, and include an allowance for loss of individual quadrats due to one-off disturbance events such as mole hills.

The grid layout and quadrat frame are designed to allow rapid location of the quadrats. The grid is laid out using tapes marked at 0.5 m intervals, the tapes

lying along at least one border of each block. The quadrat frame is oblong, 1 x 0.5 m, and is divided into two 0.5 x 0.5 m quadrats. The quadrat to be sampled in any unit can be located by positioning the frame at the appropriate mark on the tape, or, as in 50% of the cases, by flipping the frame over once from the nearest mark (Figure B shows this for the first two quadrats block 1). The accuracy of relocation of individual quadrats, including the error associated with positioning of the grid, is about ± 5 cm (less accurate on uneven ground).

For each quadrat, the height and % cover of the vegetation is estimated and other features such as ant-hills, rabbit scrapes, bare rock, etc, noted as appropriate. All vascular plants, bryophytes and macrolichens are recorded as qualitatively present if aerial parts occur within the quadrat (shoot frequency). Inflorescences, etc, knocked down by the quadrat frame, and purely saxicolous lichens, are ignored. Species initially noted as present are separated into three groups; grasses, other herbs, and mosses, liverworts and lichens. These groups are listed in approximate alphabetical order leaving gaps for some new taxa to be added as they are found. Grasses are recorded first, as either vegetative (V) or flowering (F) to help with subsequent interpretation of data (eg vegetative *Trisetum* can be very difficult to spot and an apparent increase of flowering *Trisetum* may be due to recording bias). Other herbs are recorded next, followed finally by the lower plants. Some taxa are lumped if they cannot be practically or reliably recorded (eg dried-up *Barbula* spp. or glabrous plants of *Galium saxatile/sterneri*). Small vouchers of difficult taxa, and usually most bryophytes and lichens, are taken for verification later. Cover is noted for specific dominants such as *Bromus erectus*, but not in general due to the constraints of time. It often helps to examine the quadrats from more than one side when recording. Brief notes are made on particular recording problems to aid subsequent interpretation. These procedures gives much more consistent results than a less structured routine. In particular, the use of the form as a check-list improves the quality of the recording.

The time taken to record the plot depends on the average number of species present in the quadrats, but varies from 3 to 12 hours. It takes approximately 45 minutes to mark, lay out, and describe the position of the grid. Most plots can be recorded in one day.

The data can be handled using VESPAN II (Malloch 1988).

The technique provides a quick, efficient, accurate, repeatable technique for monitoring vegetation with minimum impact on the site. It is at a scale sufficient to account for pattern in the vegetation, and the use of frequency means it is relatively insensitive to short term fluctuations in weather and management. The frequency of species in the plot can be used to calculate

constancy classes for comparison with profiles given in *British Plant Communities*. Constancies for the less frequent species tend to be under-represented compared to the standard NVC technique (recording at least five 2 x 2 or 4 x 4 m quadrats) but preliminary tests using MATCH (Malloch 1990) show that the data agree very satisfactorily with the NVC profiles. The main limitations of the sampling technique are the general lack of cover estimates, potential loss of markers from treasure-hunters with metal detectors, and that the technique is unique and not directly comparable with other data. The technique would be widely applicable elsewhere!

Figure A. Layout of the quadrats in a plot. Each 12 x 12 m plot is subdivided into four 6 x 6 m blocks, each block into nine 2 x 2 m units. Each unit is sampled using a 0.5 x 0.5 m quadrat positioned using a stratified systematic unaligned arrangement. The solid black lines show the positioning of the tapes used to mark the plot for random sampling, and the circles indicate the positions of the permanent copper marker loops.

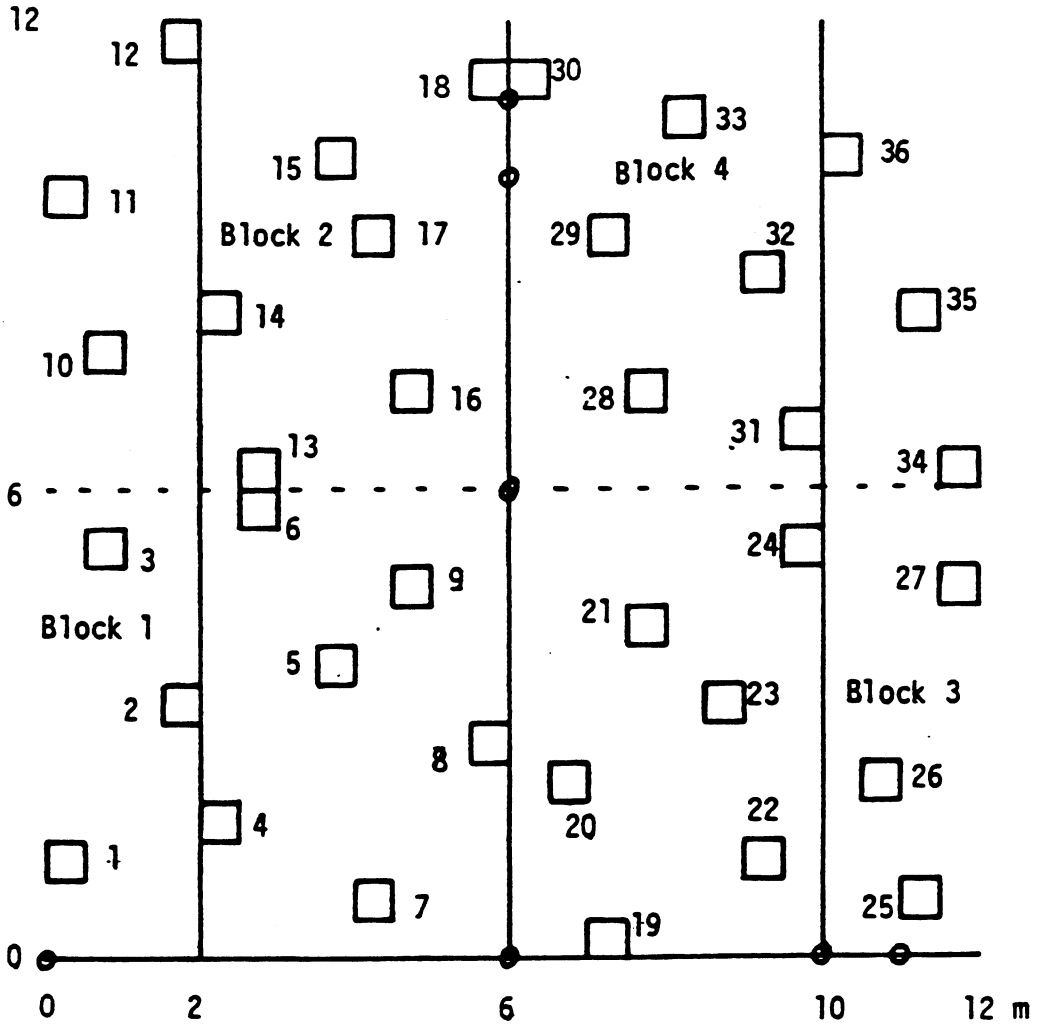
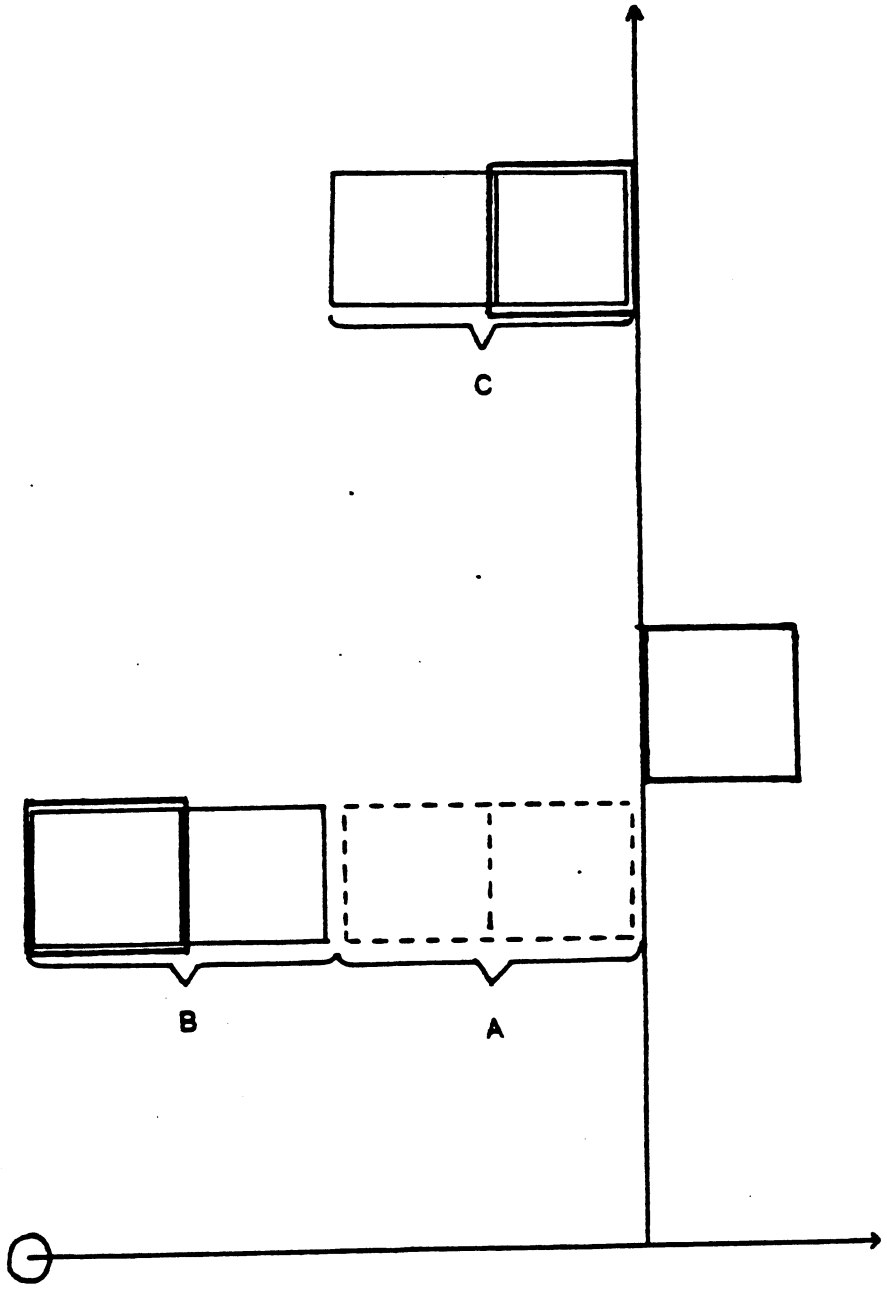


Figure B. Locating the first two quadrats in block 1. The first quadrat can be located by placing the oblong quadrat frame in position A, and then flipping it over into position B. The second quadrat can be located directly by placing the quadrat in position C.



example of data (see tables)

PLOT 83: PEWSEY DOWNS NNR, KNAP HILL, SOUTH SIDE

Site description

Pewsey Downs NNR occupies about 4 miles of the south-facing, chalk escarpment of the Marlborough Downs. The NNR has a long history of grazing and a very rich flora, with some excellent examples of calcicolous vegetation. Towards the west end of the reserve, Milk Hill Combe is a semi-circular valley with a range of aspects. Knap Hill has very rich grassland and a range of aspects.

Reasons for selection of site and plot

This site selected as part of Dorset-Breckland climate gradient, and as towards the northern limit of the CG2b subcommunity. The steep south-facing scarp will receive maximum insolation. There is a useful range of vegetation types and aspects.

This plot on a south-facing slope to contrast with plot 84 on the north-facing side, and selected as a prime example of CG2b *Festuca-Avenula* grasssland, *Succisa-Leucanthemum* subcommunity. Put in this precise spot to make it as easy as possible to relocate.

Other plots on site: Three in Milk Hill Combe, one on Sommer Down, one north-facing on Knap Hill.

Owners: NCC lease east end from New College, Oxford, and own Milk Hill Combe.

Contacts:

1) NCC South Region
   Foxhold House,
   Thornford Road,
   Crookham Common,
   Newbury,
   Berkshire RG15 8EL
2) Keith Payne, Warden
   'Greywethers'
   Bath Road,
   West Overton,
   Marlborough SN8 1QE
   Wilts (067286) 647

Management

East end farmed under a tenancy agreement from the College by Mrs S. A. Carson.

Site with a mixture of sheep and cattle grazing, this quite heavily grazed.

Plenty of rabbits, some badgers and foxes too (sets damaging grassland, especially on Knap Hill).

Access and directions for refinding plot

Park in green lane at top of road between Walkers Hill and Knap Hill. From gate on west side of hill, follow path round south side on contour, checking bearings to gate below. There is an obvious kink in the path when it faces south (magnetic, not true) and plot is immediately to the west of this point.

Miscellaneous notes
1. Species list from Keith in file. Keith has the distribution of most species plotted out to the nearest inch!!
2. Keith is monitoring five species in detail concerned with management; *Bromus erectus, Daucus carota, Linum catharticum, Trifolium repens, Senecio jacobea.*
3. NVC mapped by Keith in detail (see Local Reports, copies in file).
4. Note crop circle in distance on photo F!

PLOT RECORDING DETAILS

Grid reference: 41(SU)/121.635
Date: 3/7/1991
Time spent on survey: 7 hours
Recorder: T.C.G. Rich
Sampling: Standard, sample area 12 x 12 m
Slope: 25°
Aspect: 180 °, south facing
Weather: sunny + cool wind
Altitude: 230 m

Vegetation description
Very nice, fine, species-rich turf with abundant dicots (but bryophytes virtually absent), fairly homogenous although bottom RH corner is poorer and more disturbed by cattle poaching than rest of plot. Some species confined to path (eg *Arenaria, Odontites, Poa, Phleum*). *Avenula pubescens* and *Bromus erectus* present but heavily grazed and at very low cover.
A very nice example of CG2b *Festuca-Avenula* grasssland, *Succisa-Leucanthemum* subcommunity.

Botanical notes
Thesium also present.
Polygala seem to be largely if not exclusively *calcarea*, but possible some *vulgaris* also present
Some puzzling entire-leaved *Centaureas* which seem to be *scabiosa* rather than *nigra*
Cirsium is all good *acaule*, though hybrids with *tuberosum* not too far away.
Gentianellas all as seedlings, and presumed to be *amarella*.

Geology and soil.
Chalk.
Mixed, shallow (8cm), brown calcareous earth, with lots of chalk fragments. Field pH 7.6.

```
**********************************************
**                                          **
**        UNIT OF VEGETATION SCIENCE        **
**         UNIVERSITY OF LANCASTER          **
**                                          **
**   MATCH version 1.1, (c) 24 October 1990 **
**              copy no. 00002              **
**       supplied for the sole use of       **
**              Dr T.C. Rich,               **
**        Unit of Vegetation Science,       **
**                I.E.B.S.,                 **
**           University of Lancaster.       **
**********************************************
```

Data read from file match.83
Matching of data with diagnoses for: Calcicolous grasslands
The matching procedures have produced the following results for Plot 83. Knap Hill, south side, Pewsey Downs NNR

Community code	co-efficient	
CG 2	72.2	4 subcommunities.
CG 5	64.1	2 subcommunities.
CG 3	63.4	4 subcommunities.
CG 8	60.5	3 subcommunities.
CG 1	52.7	6 subcommunities.
CG 4	52.4	3 subcommunities.
CG 6	51.3	2 subcommunities.
CG 7	47.7	4 subcommunities.
CG 9	43.3	5 subcommunities.
CG13	32.3	2 subcommunities.

Matches against sub-communities.

Community code	co-efficient
CG 2b	75.6
CG 2a	75.0
CG 2	72.2
CG 3a	68.2
CG 2d	67.4
CG 2c	67.2
CG 5	64.1
CG 5b	63.4
CG 3	63.4
CG 5a	63.0

Plot 83. Knap Hill, south side, Pewsey Downs NNR

		quadrat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	0
11	Herb height (centimetres)		3	3	3	3	3	3	4	3	5	3	3	3	3	4	3	3	3	3	3	4	3	3	3	3	3	3	2	3	2	3	3	3	3	3	3	3	
15	Herb cover (%)		98	98	100	100	100	100	99	98	100	100	98	98	100	98	100	98	85	99	98	98	99	95	98	98	99	100	100	99	80	98	100	100	99	99	99	97	
95	Path															10		20													100									
251	Briza media		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
323	Carex flacca		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
574	Festuca ovina		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
654	Helianthemum nummularium		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
655	Avenula pratensis		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
748	Koeleria macrantha		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
1053	Sanguisorba minor		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
288	Campanula rotundifolia		1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
1244	Serratula tinctoria		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	V
1423	Viola hirta		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	V
800	Lotus corniculatus		1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	V
769	Leontodon hispidus		1	1	1	1	1	1	1	1	1	1	1	1	1		1	1		1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	V
786	Linum catharticum			1	1		1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
1333	Thymus praecox arcticus		1	1	1	1	1	1	1	1	1		1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
2687	Cirsium acaule		1		1	1		1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	V
310	Carex caryophyllea		1	1		1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1		1	V
372	Centaurea scabiosa		1		1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1			1	1	1	1		1	1	1	1	V
3200	Phyteuma orbiculare		1	1	1	1			1	1			1	1			1	1		1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	V
202	Asperula cynanchica		1	1		1	1		1	1	1		1	1	1		1	1			1	1	1			1	1	1	1	1	1	1	1	1	1	1	1	V
973	Plantago lanceolata				1		1	1		1		1	1	1	1		1		1	1		1	1	1	1	1	1	1		1		1	1		1	1	1	1	IV
988	Pimpinella saxifraga		1	1	1	1			1	1	1					1	1			1			1	1	1		1		1	1		1	1	1	1	1		1	IV
1305	Succisa pratensis		1	1	1		1	1	1	1	1				1			1			1	1	1		1	1	1	1	1			1	1			1	1		IV
3205	Polygala calcarea		1		1			1	1		1	1				1	1	1			1			1	1				1			1	1	1	1	1		1	III
403	Leucanthemum vulgare				1			1		1	1	1	1					1		1		1	1	1		1		1	1							1	1		III
576	Festuca rubra		1	1	1	1	1	1			1	1	1	1	1	1			1			1		1						1									III
1086	Ranunculus bulbosus		1			1				1		1	1		1	1	1					1	1	1						1				1	1		1		III
1205	Scabiosa columbaria						1	1				1		1			1				1	1					1	1	1			1	1				1	1	III
1239	Senecio jacobaea		1		1	1	1	1	1		1			1	1					1		1				1			1			1			1			III
284	Campanula glomerata			1				1			1		1	1				1	1	1	1		1						1	1	1						1	II
1056	Primula veris						1			1	1			1	1		1		1	1		1	1			1			1					1			1		II
964	Picris hieracioides		1	1	1			1	1							1		1		1	1		1		1		1								1			II
256	Bromus erectus		1		1		1						1				1	1	1		1						1								1		1	1	II
465	Dactylis glomerata		1	1	1	1				1	1	1			1	1	1							1															II
976	Plantago media				1	1	1			1			1						1						1			1	1			1				1			II
2971	Senecio integrifolius		1	1					1	1		1	1	1								1			1				1			1							II
234	Stachys officinalis				1			1				1	1		1	1		1								1			1							1				II
568	Euphrasia officinalis agg				1			1				1					1			1	1			1			1				1							1	II
965	Hieracium pilosella group							1					1		1					1								1	1					1	1				II
475	Daucus carota					1													1		1	1		1													1	I
572	Festuca arundinacea				1											1										1	1		1				1			1		I
877	Hippocrepis comosa				1																		1				1	1		1							1	I
1688	Fissidens cristatus		1			1	1		1	1												1																	I
122	Agrostis stolonifera				1											1			1						1					1									I
619	Gentianella amarella																1			1			1			1													I
656	Avenula pubescens			1		1			1			1															1		1										I
1059	Prunella vulgaris												1								1	1		1												1			I
174	Anthyllis vulneraria			1	1					1																												I
1349	Trifolium pratense			1																																	1	I
959	Phleum pratense bertolonii															1														1				1				I
194	Arenaria serpyllifolia																													1								I
445	Crataegus monogyna (s)																																		1			I
607	Galium mollugo											1																										I
613	Galium verum																																1					I
844	Medicago lupulina															1																						I
907	Odontites verna																													1								I
988	Poa pratensis																													1								I
1081	Ranunculus acris																		1																				I
	Number of species per sample		20	27	34	27	27	30	27	31	27	28	30	25	27	23	29	25	25	28	28	30	29	26	25	24	28	28	29	25	24	24	27	24	23	32	28	28	0

NVC species code

Constancy class

V = 80-100%

IV = 60-80

III = 40-60

II = 20-40

I = 1-20

% of samples

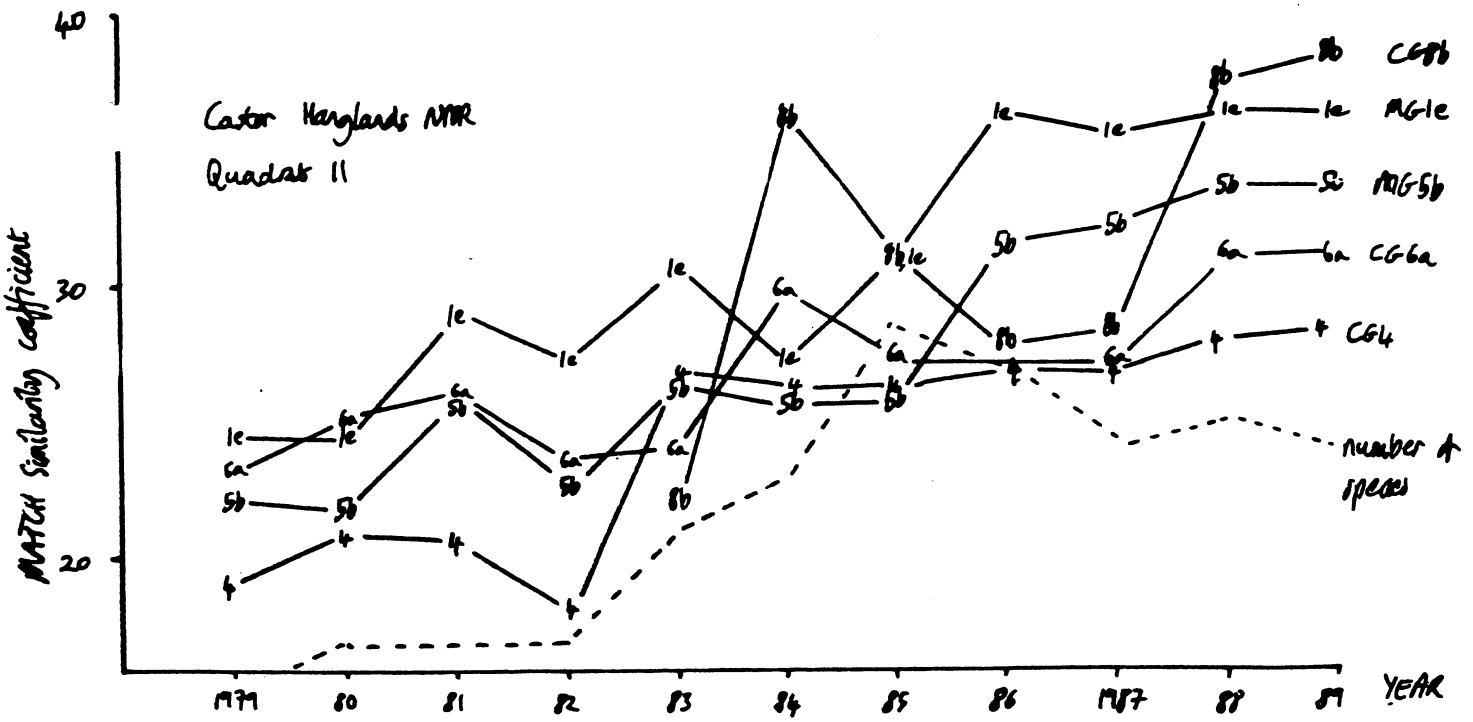
Mean number of species per releve = 27.14; standard error of the mean = .426

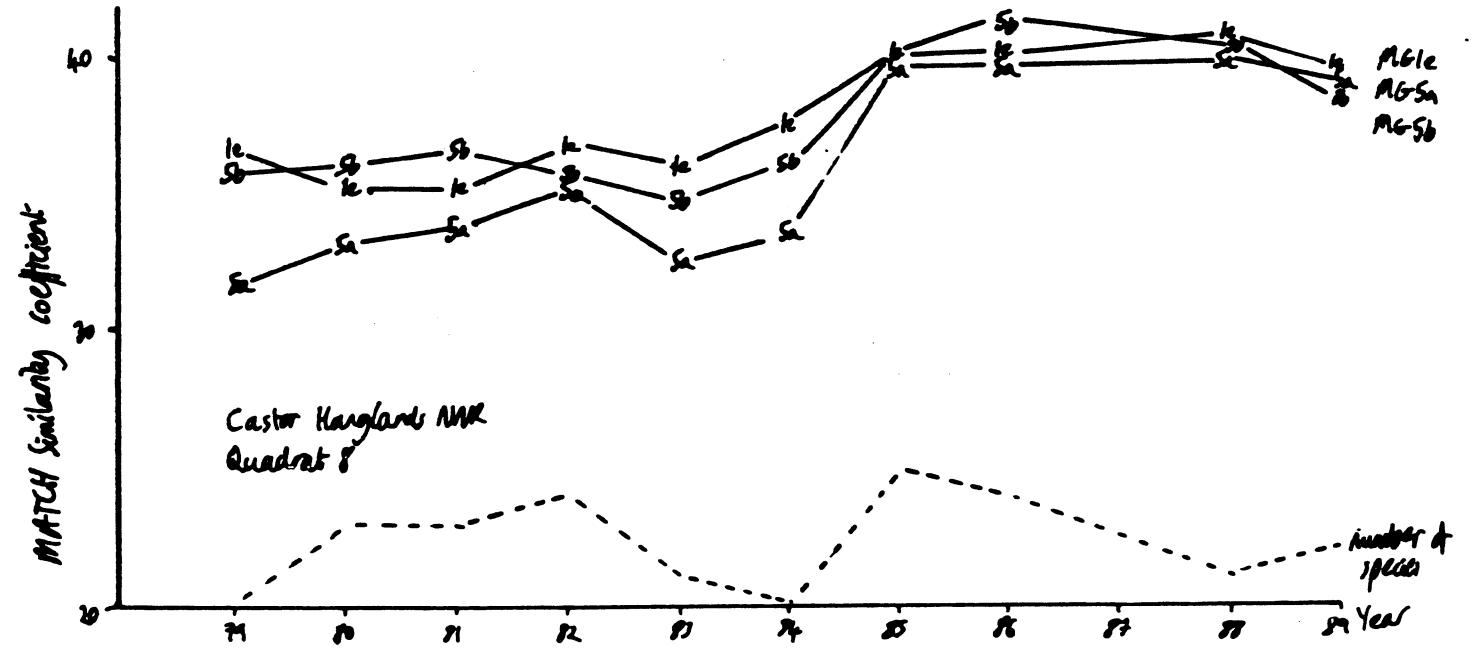
Variable number and name	Mean	St.dev.	S.E.M.	Min.	Max.	N
11 Herb height (centimetres)	3.08	.50	.08	2.0	5.0	36
15 Herb cover (%)	97.83	4.02	.67	80.0	100.0	36

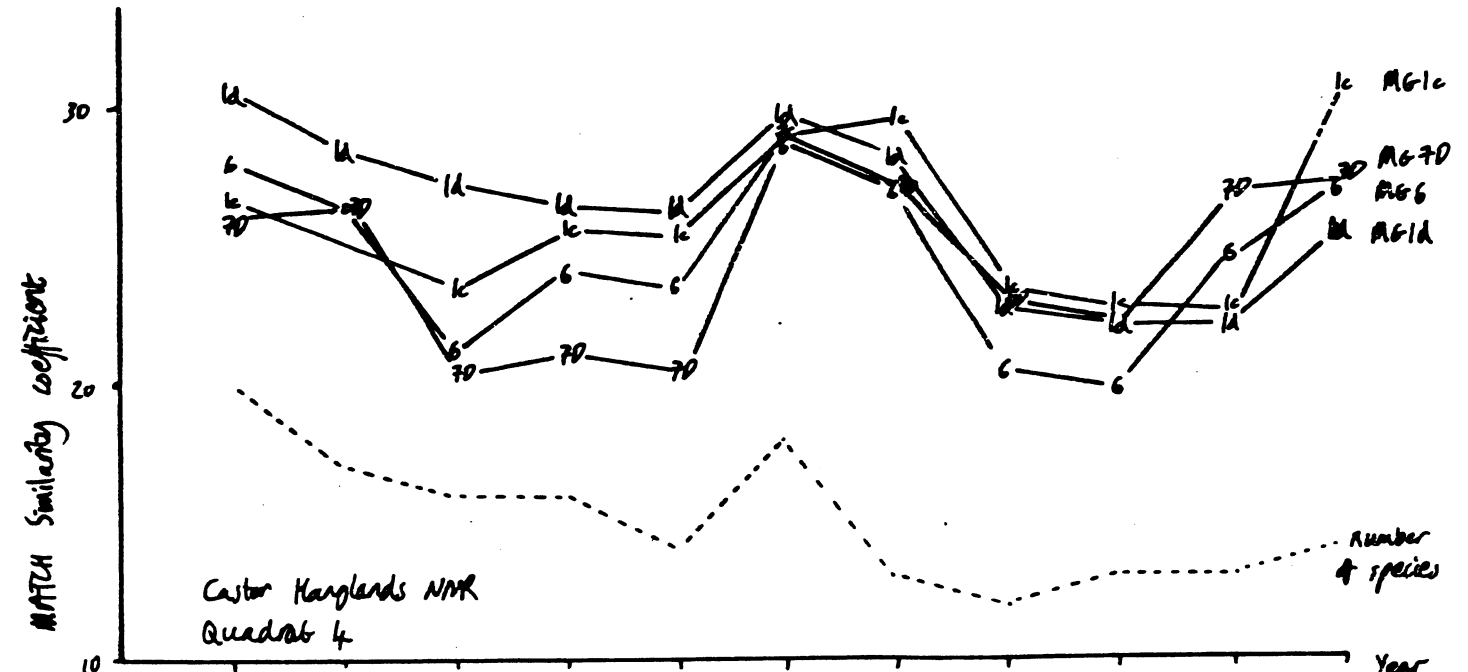
NVC species number

Species number and name	% Const	Mean	Min	Max	St.dev.	S.E.M.	N (no. quadrats)
251 Briza media	100.00	1.0	1	1	.00	.00	36
323 Carex flacca	100.00	1.0	1	1	.00	.00	36
574 Festuca ovina	100.00	1.0	1	1	.00	.00	36
654 Helianthemum nummularium	100.00	1.0	1	1	.00	.00	36
655 Avenula pratensis	100.00	1.0	1	1	.00	.00	36
746 Koeleria macrantha	100.00	1.0	1	1	.00	.00	36
1053 Sanguisorba minor	100.00	1.0	1	1	.00	.00	36
288 Campanula rotundifolia	97.22	1.0	1	1	.17	.03	35
1244 Serratula tinctoria	97.22	1.0	1	1	.17	.03	35
1423 Viola hirta	97.22	1.0	1	1	.17	.03	35
800 Lotus corniculatus	94.44	.9	1	1	.23	.04	34
769 Leontodon hispidus	91.67	.9	1	1	.28	.05	33
786 Linum catharticum	88.89	.9	1	1	.32	.05	32
1333 Thymus praecox arcticus	88.89	.9	1	1	.32	.05	32
2687 Cirsium acaule	86.11	.9	1	1	.35	.06	31
310 Carex caryophyllea	83.33	.8	1	1	.38	.06	30
372 Centaurea scabiosa	83.33	.8	1	1	.38	.06	30
3200 Phyteuma orbiculare	80.56	.8	1	1	.40	.07	29
202 Asperula cynanchica	75.00	.8	1	1	.44	.07	27
973 Plantago lanceolata	66.67	.7	1	1	.48	.08	24
968 Pimpinella saxifraga	61.11	.6	1	1	.49	.08	22
1305 Succisa pratensis	61.11	.6	1	1	.49	.08	22
3205 Polygala calcarea	58.33	.6	1	1	.50	.08	21
403 Leucanthemum vulgare	44.44	.4	1	1	.50	.08	16
576 Festuca rubra	44.44	.4	1	1	.50	.08	16
1086 Ranunculus bulbosus	41.67	.4	1	1	.50	.08	15
1205 Scabiosa columbaria	41.67	.4	1	1	.50	.08	15
1239 Senecio jacobaea	41.67	.4	1	1	.50	.08	15
284 Campanula glomerata	38.89	.4	1	1	.49	.08	14
1056 Primula veris	38.89	.4	1	1	.49	.08	14
964 Picris hieracioides	36.11	.4	1	1	.49	.08	13
256 Bromus erectus	33.33	.3	1	1	.48	.08	12
465 Dactylis glomerata	30.56	.3	1	1	.47	.08	11
976 Plantago media	30.56	.3	1	1	.47	.08	11
2971 Senecio integrifolius	30.56	.3	1	1	.47	.08	11
234 Stachys officinalis	27.78	.3	1	1	.45	.08	10
568 Euphrasia officinalis agg	27.78	.3	1	1	.45	.08	10
965 Hieracium pilosella group	22.22	.2	1	1	.42	.07	8
475 Daucus carota	19.44	.2	1	1	.40	.07	7
572 Festuca arundinacea	19.44	.2	1	1	.40	.07	7
677 Hippocrepis comosa	16.67	.2	1	1	.38	.06	6
1688 Fissidens cristatus	16.67	.2	1	1	.38	.06	6
122 Agrostis stolonifera	13.89	.1	1	1	.35	.06	5
619 Gentianella amarella	13.89	.1	1	1	.35	.06	5
656 Avenula pubescens	13.89	.1	1	1	.35	.06	5
1059 Prunella vulgaris	13.89	.1	1	1	.35	.06	5
174 Anthyllis vulneraria	8.33	.1	1	1	.28	.05	3
1349 Trifolium pratense	8.33	.1	1	1	.28	.05	3
959 Phleum pratense bertolonii	5.56	.1	1	1	.23	.04	2
194 Arenaria serpyllifolia	2.78	.0	1	1	.17	.03	1
445 Crataegus monogyna (s)	2.78	.0	1	1	.17	.03	1
607 Galium mollugo	2.78	.0	1	1	.17	.03	1
513 Galium verum	2.78	.0	1	1	.17	.03	1
844 Medicago lupulina	2.78	.0	1	1	.17	.03	1
907 Odontites verna	2.78	.0	1	1	.17	.03	1
988 Poa pratensis	2.78	.0	1	1	.17	.03	1
1081 Ranunculus acris	2.78	.0	1	1	.17	.03	1

TCR 23/4/91 Data supplied M Massey (single downs quadrats)







acteristic of the particular community featured in the diagnosis, but nevertheless found in relatively few of the samples making up the diagnosis. There is, of course, a complete spectrum of species behaviour reflecting the different tolerance ranges of the different species and the essentially "continuum" nature of vegetation.

An example of a community diagnosis is given in table 1.

How the matching works.

The data you supply (e.g. table 2) are turned into a table of constancies in the manner of diagnoses so that each species will be assigned to a constancy class according to its frequency within the samples and its maximum cover-abundance value (in the Domin scale) found. If you only have one sample, for instance, every species will have a constancy of V; as the number of samples increases, so the species constancies change to give a spectrum of different values. The constancy values from your data are then compared with the constancy profile of the communities recognised in "British Plant Communities", using the Czekanowski co-efficient:

$$C = \frac{200\ \Sigma \min(x_j, y_j)}{\Sigma x_j + \Sigma y_j}$$

where x_j is the constancy (on a scale of 1 to 5) for species j in sample x and y_j is the constancy of the same species in sample y; $\min(x_j,y_j)$ is the lesser of the two values x_j and y_j. The value of C may vary between 0 for complete dissimilarity and 100 for complete similarity.

One minor modification has been made in the case of species of constancy class I in the diagnostic data. Should these species (which are usually very numerous) not occur in the data being matched then, instead of being given the value 1, they can be assigned a lower value (e.g. 0.25) at the operator's choice. This allows more weight to be given to species of higher constancy.

The program stores the ten highest values of C (the best matches) which are displayed in order of decreasing value, together with the appropriate community codes.

Data Set 7

Long Term Monitoring of Calcareous Grassland

Background

The project involves sampling the same location exactly at regular intervals over a long period of time. 36 quadrats arranged in 4 blocks are laid out in a standard pattern, originally chosen at random but subsequently repeated at the same and different locations according to a published plan. Presence-absence is recorded for each species.

1. *Analysis of species frequency changes*

The basic data for a given species is the total frequency (out of 36) in each site and year. The scores can be plotted against time, and since many species are perennial it is to be expected that they reappear in the same quadrat over a run of years. It is therefore difficult to accept that the comparison of adjacent years can use the test for two independent binomial samples. However if the binomial test is used it will tend to be conservative, that is, to treat significant changes as non significant. Over longer periods of time it becomes easier to accept that the samples are independent.

The two sample binomial test follows the 2 × 2 table procedure (with Yates' correction) described under data set 1. Significant differences are observed roughly between the following pairs

0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19
4	6	8	10	11	13	14	15	16	18	19	20	21	23	24	25	26	27	28

21	22	23	25	26	28	30	32
29	30	31	32	33	34	35	36

so that the example of 18 and 27 is just significant.

The G-test is probably an ad hoc name for the likelihood ratio test, an alternative form of the χ^2 test which gives similar results except when the proportions are close to zero or 100 per cent. The term G-test is not widely used in standard statistical literature.

An alternative test if the assumption of independence is not justified is to treat the plot as matched pairs, and to perform a non-parametric test, as described by Siegel, Non

parametric Statistics, 1956. For example the plots in which the presence/absence status changes can be counted, and tested for significance. This form of analysis is usually tedious to organise in a large data set, unless there is a computer procedure to do it automatically.

2. *The analysis of MATCH coefficients*

The MATCH program used to compare samples with NVC standard sets is based on the Czekanowski coefficient, in this case using presence-absence only. There seems to be no particular reason for preferring this to the Jaccard coefficient, but the two are related and give the same rank ordering of coefficients, so that we would expect the same types to be chosen. The relationship is not quite linear, so mean similarities on the two scales will not correspond exactly.

(The explanation of the MATCH coefficient is not quite exact because it refers to values on a 1 to 5 scale rather than a 0 to 5 scale).

The problem of analysing similarity coefficients statistically is that there is no agreed basis for postulating a probability distribution. The coefficient is a mean value of matching scores of a wide variety of species, some common, some very rare, some local, some widespread. Without therefore any theoretical basis we can only use the observed sequences and treat either the similarities themselves, or the differences between similarities for two NVC types, as a time series in which we may look for the possibility of trends.

If for example there appears, as in the plots for Castor Hanglands, that the Match with certain types is increasing with time, we can test for a significant trend against the background of residual fluctuation. And if the plot is moving from one preferred type to another, that may show up in the differences analysed for trend. In a long sequence it might also be possible to observe periodic behaviour using periodogram analysis, although many years would be needed to establish real cycles.

The correlation between the coefficients and the total number of species no doubt reflects the fact that type profile remains fixed, and the sample profile varies with time, so that as new species are included, provided they are also in the type list, the coefficient will increase. If on the other hand the new species are not in the list then the coefficients will decline instead.

DATASET 8.

Saltmarsh Monitoring

The dataset is described in the attached papers.

Questions

The analysis appears to be potentially quite complex as the recovery of the marsh through time is also influenced by a spatial zonation from the land to the sea. Should the data be treated as a three-way classification e.g. time x disturbance x zonation and if so will the results be readily interpretable? If not, what kind of analysis is recommended.

Introduction

The site consists of a gently sloping saltmarsh about 230m in length from sea wall to bare mud. The saltmarsh showed typical vegetation zonations which were interrupted by deep creeks in places. The area had been traversed by an outfall pipeline, the laying of which had disturbed a corridor approximately 50 m wide.

Aims

To record the vegetation in disturbed and undisturbed areas in order to assess the success of recolonisation and re-establishment of the saltmarsh vegetation.

Methods

Wessex Water had provided us with aerial photographs of the site which proved very helpful in locating the extent of the disturbed corridor on the ground. The aerial photographs also show the persistence of vehicular tracks in the saltmarsh.

The saltmarsh was surveyed on 20/8/91.

The disturbed corridor was marked on the ground and a series of quadrats recorded across the area. Five 2x2 m quadrats were located at each of 4 levels in the marsh in 'control' undisturbed areas and five (sometimes six) recorded within the pipeline corridor. Each quadrat was scored for species presence and each species scored for cover-abundance using the Domin system:-

10	% cover----------91-100%
9	% cover----------76-90
8	% cover----------51-75
7	% cover----------34-50
6	% cover----------26-33
5	% cover----------11-25
4	% cover----------4-10
3	cover < 4 % many individuals
2	cover < 4 % several individuals
1	cover < 4 % few individuals

The four different levels in the saltmarsh were selected from inspection on the ground and of the aerial photographs. The four levels recorded were;

Zone 1; along the seaward face of the sea wall
Zone 2; about 20-25 m from the base of the sea wall
Zone 3; about 100 m from the base of the sea wall
Zone 4; about 200 m from the base of the sea wall

The position of the 4 areas investigated are shown on Fig. 1.

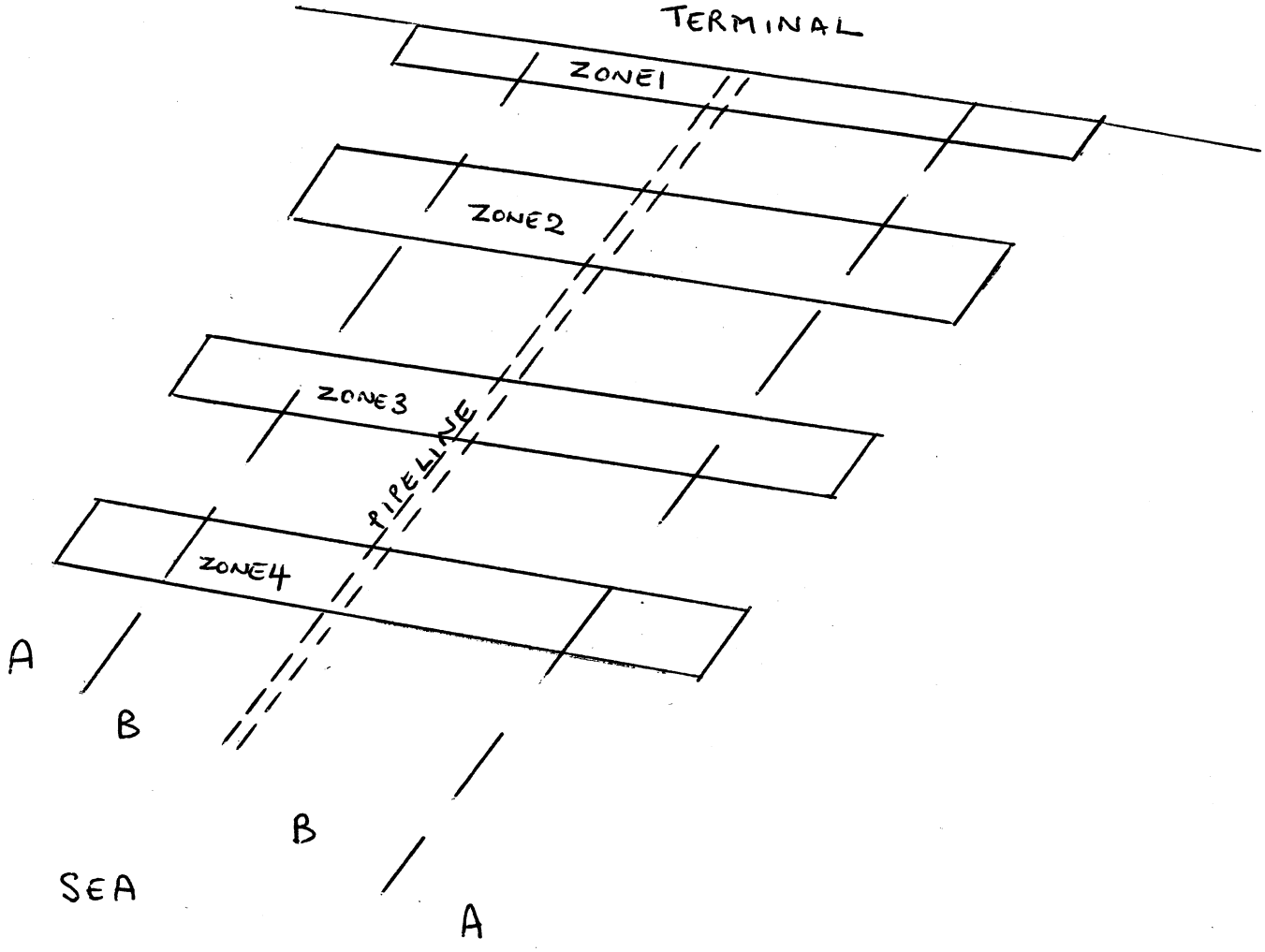


FIG 1 : SKETCH OF LAYOUT OF ZONES

A= UNDISTURBED "CONTROL" AREAS

B= DISTURBED PIPELINE CORRIDOR

Table 1.
Zone 1. Control quadrats.

Quadrat number		Domin cover/abundance 1	2	3	4	5	Constancy 0
117	*Elymus pycnanthus*	10	8	8	7	9	V
118	*Elymus repens*			3	5	4	III
122	*Agrostis stolonifera*	3	3	4	6	2	V
197	*Arrhenatherum elatius*			2			I
217	*Atriplex prostrata*		1			3	II
475	*Daucus carota*					1	I
576	*Festuca rubra*	4	5	6	7	4	V
685	*Hordeum secalinum*			2			I
988	*Poa pratensis*			2			I
1043	*Potentilla anserina*				5	2	II
2982	*Taraxacum* seedling/sp			1			II
Number of species per sample		3	4	8	5	7	0

Table 2.
Zone 1. Disturbed area.

Quadrat number		Domin cover/abundance 1	2	3	4	5	6	Constancy 0
117	*Elymus pycnanthus*	5		3		3	5	IV
118	*Elymus repens*	4	5		5	4		IV
122	*Agrostis stolonifera*	5	5	6	7	7	8	VI
156	*Alopecurus geniculatus*			1				I
217	*Atriplex prostrata*	4	4	5	2		4	V
258	*Bromus hordeaceus hordeaceus*	3	2	4			1	IV
384	*Cerastium fontanum triviale*		2	1		1		III
415	*Cirsium arvense*				1	4		II
419	*Cirsium vulgare*	1	1		1		1	IV
433	*Convolvulus arvensis*						2	I
460	*Cynosurus cristatus*		1		2			II
465	*Dactylis glomerata*						1	I
521	*Epilobium hirsutum*					3		I
526	*Epilobium parviflorum*					1		I
575	*Festuca pratensis*		1				2	II
576	*Festuca rubra*	6	7	8	3	4	5	VI
680	*Holcus lanatus*		1	2	2	4	4	V
	Hordeum jubatum			1				I
685	*Hordeum secalinum*	3	3	2	3		4	V
796	*Lolium perenne*	4	5	4	6		5	V
959	*Phleum pratense bertolonii*				2			I
981	*Poa annua*					2		I
988	*Poa pratensis*	3	3	3	3	4	3	VI
1043	*Potentilla anserina*	2		1	1			III
1086	*Ranunculus bulbosus*			2	1	3		III
1142	*Rumex conglomeratus*			1				I
1143	*Rumex crispus*	1	1					II
1271	*Sonchus arvensis*				1			I
1350	*Trifolium repens*	2	3	3	2	1		V
1368	*Urtica dioica*						1	I
Number of species per sample		12	15	16	15	13	14	0

Data Set 8

Saltmarsh Monitoring

The data are cover scores (Domin Scale) on random quadrats (not resampled on future occasions) classified by (a) disturbed-undisturbed area, and (b) zone (distance from sea or wall).

The number of species is much greater in the disturbed than in the undisturbed area. In the former cover scores of 8-10 leave little space for other species.

The three-way classification of time × disturbance × zone may be analysed statistically in a conventional analysis of variance or analysis of deviance, in which the variables are either percent presence or mean cover (on a de-transformed Domin scale, for example) for each species. The analysis of cover scores may use the individual subplots as units, but the analysis of percent presence (transformed to angles) can only use the combined information from the five or six samples within a category.

The analysis of variance (or deviance, if the GLIM approach is used) would be as follows

Source of variation	**d.f**	
Between years	Y - 1	
Disturbed-undisturbed	1	
Between zones	3	
Zones × disturbance	3	
Years × disturbance	Y - 1	
Years × zones	3(Y - 1)	
Years × disturbance × zones	3(Y - 1)	
Replication within sets	8Y(r - 1)	(for cover scores only)

For the analysis of percentages the three factor interaction would be used to estimate the error variance. The zones component may be partitioned for linear regression on distance. This analysis assumes only one error stratum. If there is evidence of greater variation between years than within years it will be necessary to analyse it in the manner of a split unit design, in which variation between years is assigned to the main stratum, and variation within years to the lower stratum. There will then be very few degrees of freedom for testing the main effect of years and the two factor interactions cannot be

treated separately because they have to be used to estimate the error variance of the main stratum. These problems are discussed in Cochran and Cox, 'Experimental Designs'.

The analysis of all species simultaneously using cluster analysis and ordination should allow the main contrasts to be displayed. If units are labelled by year, zone and disturbance, we might expect to see a large separation due to disturbance, smaller separation due to zones, and a gradual shift of position with time. Interaction between zone and disturbance are shown by non-parallel displacements. Interactions with time are indicated by differential rates of displacement between zone or disturbance groups. We would no doubt expect that the disturbed plots will become more similar to the undisturbed plots with time, so that the ordinations should display a tendency to cover, while the clusters might contain mixtures of early undisturbed and late disturbed plots. The interaction between zone and disturbance might show that disturbance is more important in zone 1 than in zone 4, or vice versa.

A further possibility is to take say the two main principal components (which represent contrasting groups of species) and to perform an analysis of variance on these to show the extent to which these contrasts are associated with disturbance, zone or time.

DATASET SET 9.

MONITORING POPULATIONS OF GRASSLAND PLANTS

Background

English Nature frequently conserves sites to protect populations of rare or uncommon species as part of an overall objective of conserving habitats and their characteristic fauna and flora. Two examples described below indicate the type of data collected to assess whether populations are being maintained, either under current conditions or after a change in management of a site. One example, on a steep valley side, is a grassland in Yorkshire with mountain avens, Dryas octapatela, as well as more common species such as Rockrose, Helianthemum nummularia. The site is grazed by sheep at moderate levels compared to similar land in the area. The second example is a hay meadow in Suffolk which has a population of Snake's Head Fritillary, Fritillaria meleagris as well as several orchid species. Hay is cut in July and then the site is grazed during the autumn.

Objectives

Under current management are the populations of uncommon species being maintained? Specifically, can significant trends be identified from simple counts that would suggest populations are declining and therefore should be subject to more detailed investigation, for instance in relation to management regime?

Sampling Design and Information Collected

At the Dryas site, 16 transects were laid out, 3 metres apart and perpendicular to the slope. Each transect was 20 metres long. Every 2 metres along each transect, the presence or absence of Dryas octapetala and a characteristic calcicole Helianthemum nummularia, was recorded in 50cm by 50cm quadrats so that a total of 10 quadrats per transect and 160 quadrats overall were inspected. The attached data sheet illustrates the results for 1991. Recording will be repeated at yearly intervals.

In the haymeadow, the whole field was divided up into a grid 5 metres by 5 metres. In every subdivision the number of flowering individuals of Fritillaria melaegris was recorded by members of the Suffolk Wildlife Trust. Several orchid species were recorded in the same way, including Green-winged orchid, Orchis morio. Recording began in 1979 and has continued every year to 1991 and will be repeated yearly in the future. Examples of data collected and graphs of population numbers over time are attached.

Questions

1. For the systematic sample of the Dryas population can a mean of the frequencies from the 10 transects in one year be compared to a mean from other years?

2. For the total count of Fritillaria and other species is some kind of trend analysis possible and likely to be helpful given that 13 years of data are available? One year (1986) is missing for Orchis morio, will this affect the analyses that can be done? Are longer runs of data needed before trends can be distinguished with some confidence from cycles and 'noise'?

THE GILL 5/7/91 : BM/PW/JC/RW

SKY Transects A-P Presence/Absence in c. 0.5 m x 0.5 m Quadrat

✓ = Present; 0 = Absent; - = Nil

Dryas octopetala

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-
2	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-
3	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-
4	✓	0	0	✓	0	0	0	0	-	-	-	-	-	-	-	-
5	0	0	0	0	✓	0	0	0	-	-	-	-	-	-	-	-
6	✓	0	0	✓	0	0	0	0	-	-	-	-	-	-	-	-
7	0	0	0	0	0	0	0	✓	-	-	-	-	-	-	-	-
8	✓	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-
9	0	✓	0	0	0	✓	0	0	-	-	-	-	-	-	-	-
10	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-

TOTAL: 9

Helianthemum nummularium

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	0	✓	✓	0	✓	✓	✓	✓	✓	✓
4	0	✓	✓	0	✓	✓	✓	0	✓	✓	✓	✓	0	✓	✓	✓
5	0	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	✓
6	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
7	0	✓	✓	✓	✓	✓	✓	✓	✓	0	✓	✓	✓	✓	0	✓
8	0	0	0	0	✓	✓	✓	✓	✓	✓	0	0	0	0	✓	✓
9	✓	✓	✓	✓	✓	✓	✓	0	✓	✓	0	0	✓	✓	✓	✓
10	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	0	0	✓	✓	✓	✓

TOTAL: 129

MARTINS' MEADOWS MONEWDEN FIRST CHURCH MEADOW

DATE OF COUNT: 14.04.91. SPECIES Fritillaria meleagris (Total nos) COUNTERS: D.Noble P.Chapman The Dawsons

32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	64	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	7	2	0	0	0	0	0	0	2	5	3	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	36	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	11	10	1	18	5	1	0	0
0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	3	17	18	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	36	4	0	0	37	0	68	84	0	0	0	0	0	43	28	35	8	0	0	0	1
0	0	0	0	0	0	0	0	1	15	5	121	97	5	27	116	84	129	137	43	0	0	0	0	26	32	15	13	1	0	0	0
0	0	0	0	0	0	0	0	0	44	5	7	45	8	101	173	160	198	125	151	10	10	0	0	1	28	46	37	13	0	0	0
0	1	2	0	0	0	0	1	0	2	1	0	1	31	92	81	299	56	95	54	15	15	0	14	13	1	16	80	69	23	0	0
32	**31**	**30**	**29**	**28**	**27**	**26**	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
0	1	2	0	0	19	0	1	1	61	11	164	147	44	220	407	615	458	443	248	25	25	0	27	100	110	179	179	89	24	0	1

TOTAL : 3369

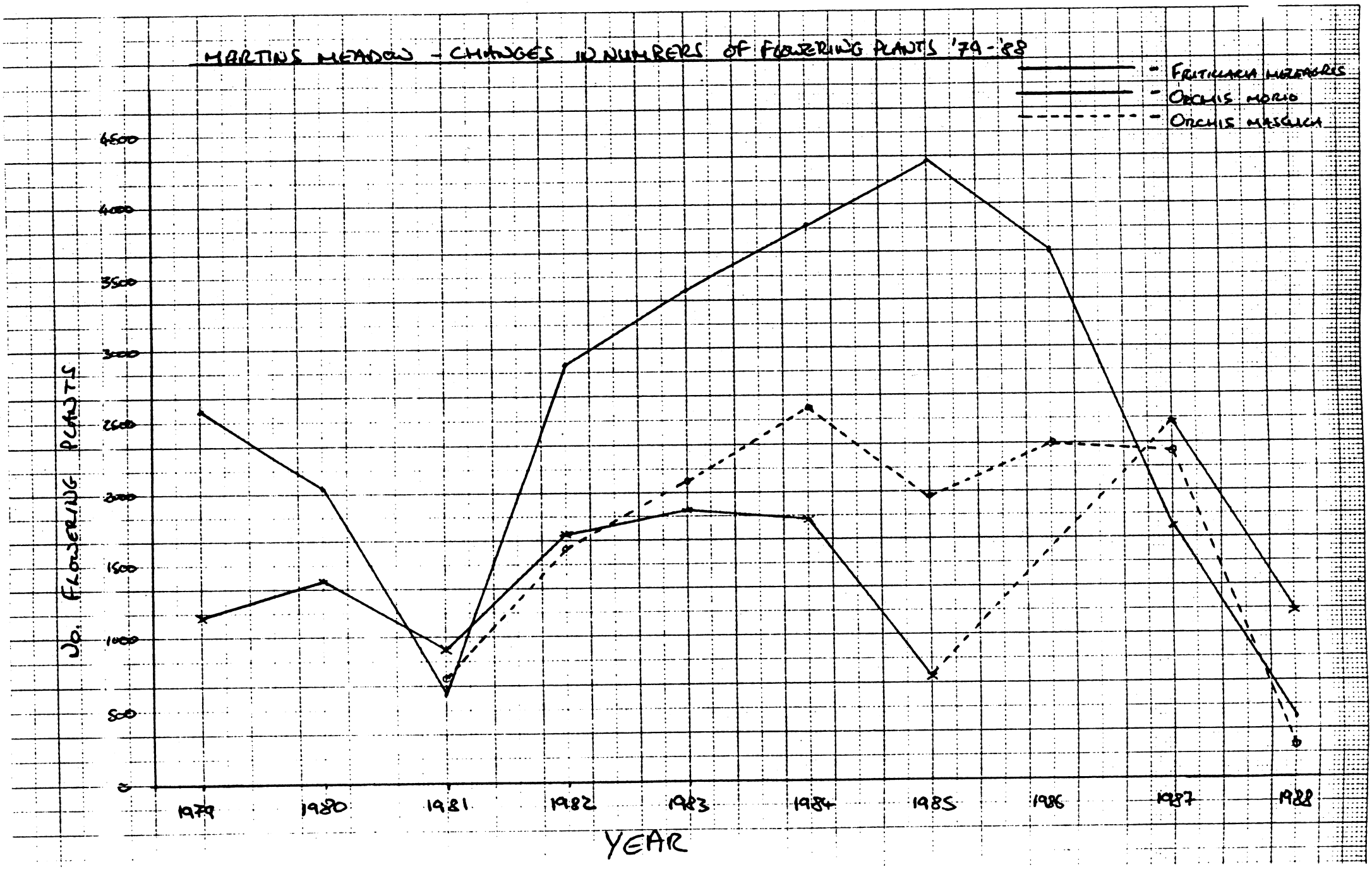
MARTINS' MEADOWS MONEWDEN FIRST CHURCH MEADOW

DATE OF COUNT: 19.05.91. SPECIES Orchis morio COUNTERS: 6th formers from Framlingham Col

32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
0	1	1	0	0	0	0	0	1	0	0	3	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
0	0	3	2	0	0	0	0	1	0	0	0	5	1	0	44	0	0	0	0	3	0	0	0	0	1	7	8	2	1	0	0
1	0	3	2	1	0	0	0	5	1	0	3	28	86	83	0	21	2	1	1	33	1	0	1	0	1	1	8	4	7	0	0
12	0	1	2	0	0	0	0	2	0	3	8	17	118	41	3	8	2	0	1	8	0	0	0	0	4	3	26	26	10	0	3
8	3	3	7	4	0	0	0	4	0	3	4	2	13	19	0	0	2	1	0	2	0	0	0	0	7	4	43	10	16	12	3
2	2	2	0	3	1	9	0	4	0	4	2	10	12	6	5	4	1	0	0	1	2	0	0	0	3	63	49	28	4	7	7
6	15	0	2	2	12	9	1	4	2	0	5	3	5	5	3	2	0	0	1	1	2	0	7	9	3	24	19	4	4	8	3
4	1	2	0	1	4	4	2	0	1	5	0	0	0	0	3	7	3	2	1	0	3	28	6	1	2	5	25	9	6	9	6
9	2	0	0	8	9	4	2	0	4	0	4	0	4	5	13	0	2	0	1	1	2	3	20	6	9	4	4	4	3	7	2
7	7	3	3	13	7	14	2	0	5	7	5	3	3	4	5	2	5	3	16	0	17	101	36	18	2	1	8	8	0	7	2
3	9	13	10	18	14	42	9	0	14	5	9	8	4	3	21	1	14	4	14	0	7	5	15	14	8	31	18	39	32	7	0
6	4	8	16	18	4	21	16	0	20	19	6	3	10	0	33	1	4	7	14	0	8	9	24	13	28	13	11	28	13	12	4
2	9	31	17	14	4	12	24	0	24	40	10	30	6	8	12	2	3	3	2	0	1	5	2	2	4	22	35	10	3	4	4
1	15	1	8	17	2	5	21	0	12	14	10	5	10	2	0	0	6	1	7	0	2	1	0	0	2	3	15	0	0	0	0
32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
61	68	71	69	99	57	120	77	21	83	100	69	114	272	176	150	48	44	22	58	49	45	152	111	63	74	181	269	173	99	73	34

TOTAL : 3102

White form found in 4th square down in row 22


MARTINS MEADOW - CHANGES IN NUMBERS OF FLOWERING PLANTS '79 - '88
- FRITILLARIA MELEAGRIS
- ORCHIS MORIO
- ORCHIS MASCULA
No. FLOWERING PLANTS
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
YEAR

Data Set 9

Monitoring Populations of Grassland Plants

There are two different types of observations described: presence-absence of species of creeping habit (Dryas, Helianthemum) and counts per quadrat of individual specimens (Fritillaria, Orchis). The former provides only binomial variables of low precision, whereas the latter allows more detailed analysis of aggregated distributions.

1. *Comparison of mean frequencies of Dryas presence*
The question refers to 10 transects, which presumably should be 10 samples on 16 transects. However, in the table 8 of the transects record 'Nil' rather than absent, which is taken to mean that the observations were not made.

Assuming all the data were available as in the second table, the sampling error of the mean number of quadrats with presence scores may be computed directly from the 16 (or 8) column totals, giving

	Mean	s.e.	s.e. (Mean)	s.e. (Binomial)
Dryas	1.125	0.991	0.313	.316
Helianthemum	8.125	1.450	0.459	.395

which may be compared with the theoretical standard error for quadrats distributed completely at random. In the case of helianthemum there is evidence that the frequencies are not the same on each strip.

The question asks whether it is valid to make comparison between years. Now if the quadrats are fixed and the plants are perennial, we would naturally expect high correlations between years, and it would therefore be preferable to record the numbers of quadrats which Change Status, as in Data Set 1. Statistically this is called a 'matched pairs' comparison. If different quadrats are chosen we can use the ordinary tests for two independent samples.

The interpretation of changes between years will depend on whether it is thought to be a temporary or permanent change of frequency, and long runs will be necessary to establish this, particularly if there are large differences in temperature and rainfall between years.

2. *Analysis of sequences of counts of Fritillaria and Orchis*

The distributions of individuals per quadrat are obviously non-random, by inspection, and contour plots of species density may be constructed, preferably using some moving average technique to smooth the plots.

The distributions of counts in the two tables fit quite well to the Negative Binomial distribution (using the Rothamsted Maximum Likelihood Program, MLP), as follows

	Fritillaria	Orchis
Mean Count	7.76	7.19
Variance	743.0	181.2
Negative Binomial 'K'	0.046	0.448
χ^2 for goodness of fit	17.9 on 13 d.f.	40.7 on 18 d.f.

These analysis indicate a high level of aggregation, which means that the variance attached to any estimate of mean count must be obtained by analysis of replicate samples rather than by assuming a random distribution of individuals.

The analysis of changes of numbers with years suggests that a time series analysis will be possible as the number of years increases. Time series analysis allows one to test for the presence of trends and cycles within data series, taking into account serial correlations. There may be external variables such as mean temperature and rainfall which are important, and with only ten years illustrated in the figure one can do little more than compute the serial correlation (+0.51 for Fritillaria) and observe that the two minima are several years apart. If there was any reason to suspect a true periodicity, many such cycles would need to be observed, and the interpretation of some lengthy time series is still the subject of controversy.

Statistical models for fluctuations in species frequencies involve many different mechanisms, such as competition, predation, selection pressure and environmental change. Statistical tests of significance may be able to establish the reality of change, but extra biological knowledge is necessary to decide which model is most reasonable in explaining the changes.

DATA SET 10 - FLOOD-PLAIN MEADOW MONITORING

Background

Gravel extraction has recently begun adjacent to a species-rich flood-plain meadow in the Thames Valley. Because of concern that water table changes could result from the extraction, monitoring of the plant species and water table in the meadow was begun. If deleterious changes become evident engineering action or other measures will be taken to try to remedy the situation.

Objectives

To assess whether the characteristic species assemblages of the meadow are being maintained and if changes can be related to changes in water table levels.

Sampling design and information collected

Two approaches have been adopted. For one method, 16 permanent plots or stands have been located in 'representative' areas of vegetation. The meadow varies patchily in wetness and species richness and samples have been located to cover this variation. Each plot is 10 by 10 metres in size. Within each plot 10 1m x 1m quadrats are defined in a regular arrangement and each is divided into 25 20cm by 20cm sub-quadrats. Presence of vascular plant species (excluding grasses) is recorded in each sub-quadrat. The layout of the quadrats and sub-quadrats is shown on the attached sheet. Most plots (13) are within 15m of boreholes where water levels are recorded while 3 are 50m from boreholes. For the second method, a transect 1 metre wide and 300 metres long was located across the meadow, more or less perpendicular to a potential gradation in water level change around the gravel pit. Boreholes were located at either end of the transect. The transect was treated as a linear series of 1m by 2m quadrats in which five 20cm by 20cm sub-quadrats were recorded for the presence of vascular plant species (excluding grasses). A total of 150 1m by 2m quadrats were therefore recorded along the transect. The layout of the sub-quadrats is shown on the attached sheet.

Data was collected in the year preceeding gravel extraction and in each subsequent year.

Analysis

Examples of the species frequency data obtained from the 2 sampling methods are attached. Analysis of changes in individual species thought to indicate change in hydrology is wanted eg Sanguisorba offinalis, Succisa pratensis and Filipeidula ulmaria which may be affected by lower water tables. Analysis of change in the assemblage of species is also important. In addition correlation of changes with changes in water level (probably represented by some kind of index) need to be examined. Along the transect it may be possible to relate vegetation change to differences in water level which have been derived from a model of the water table across the whole site as a substitute for actual water level records along the length of the transect.

Questions

1. Can the plot data be combined in any kind of analysis, for example in an ordination of samples? Can differences in location of these samples within an ordination be assessed statistically?

2. Would correlations, say of species frequency and water level have to be carried out for each plot separately?

3. Can correlations of species frequency and frequency changes and distance along the transect from gravel extraction be made? Can correlations be made with modelled water levels if this data has been considered to be acceptable?

4. What other methods would be recommended?

Figure 4.

Position of Some Stands and Transect on Pixey and Yarnton Meads

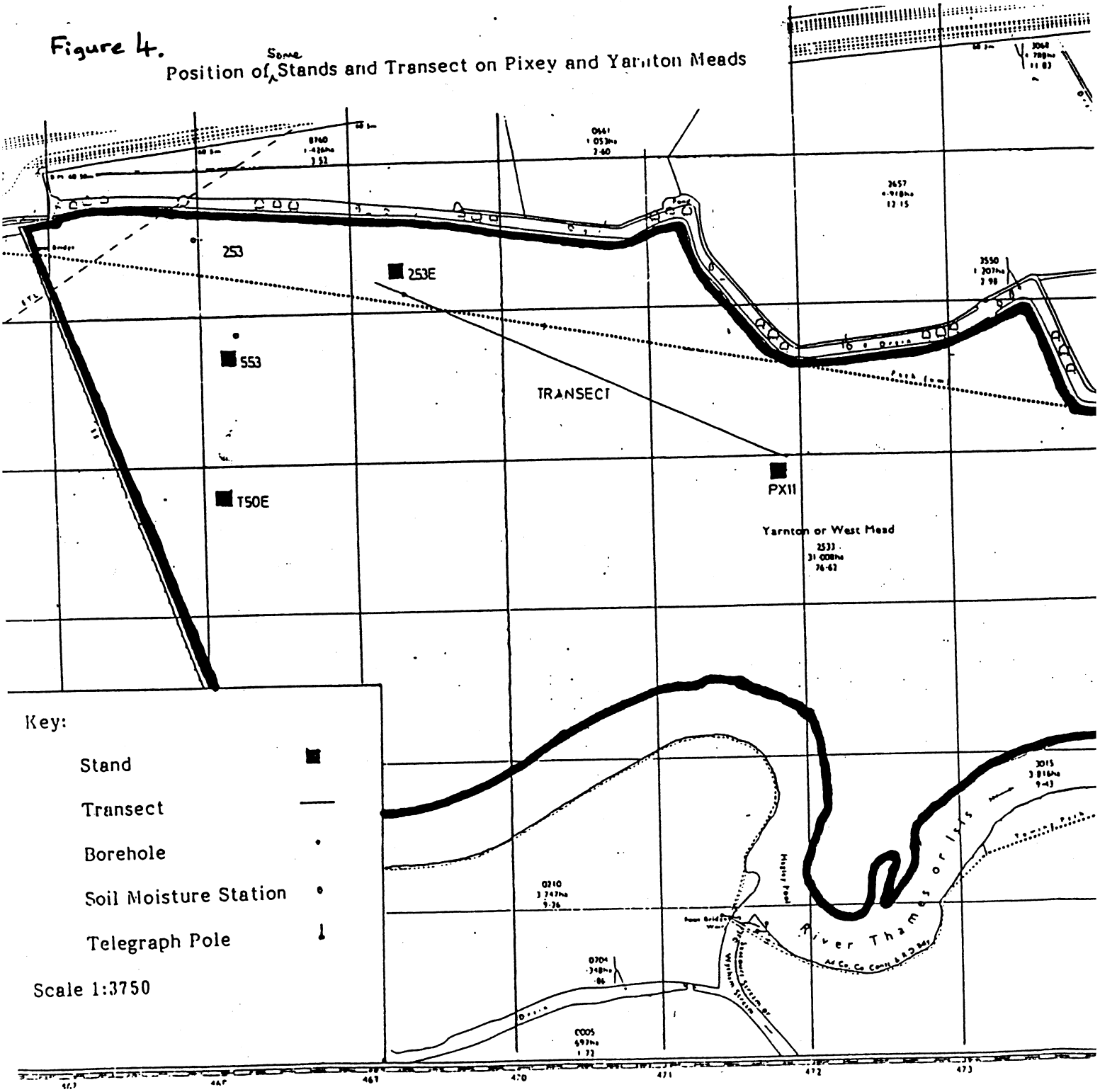
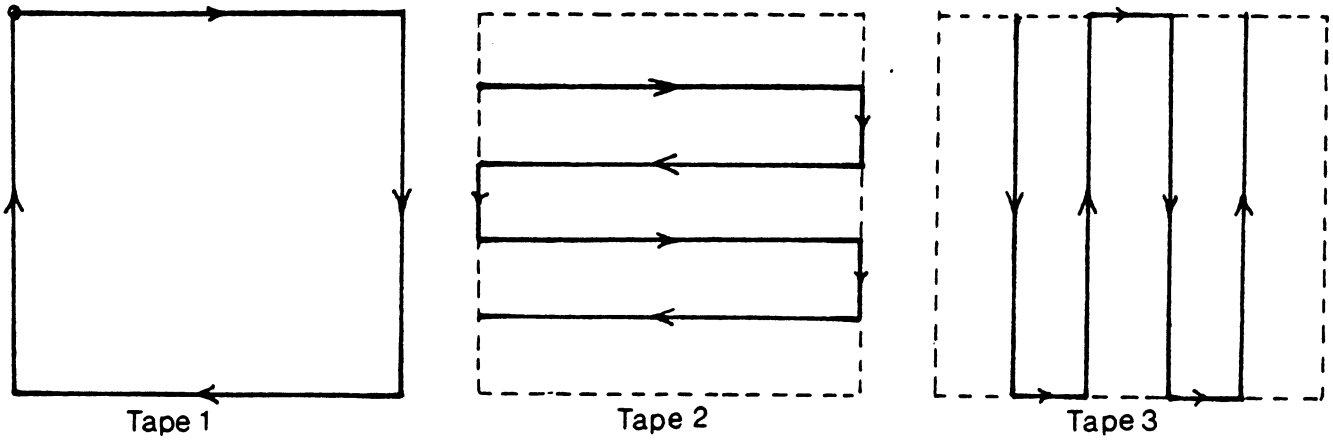
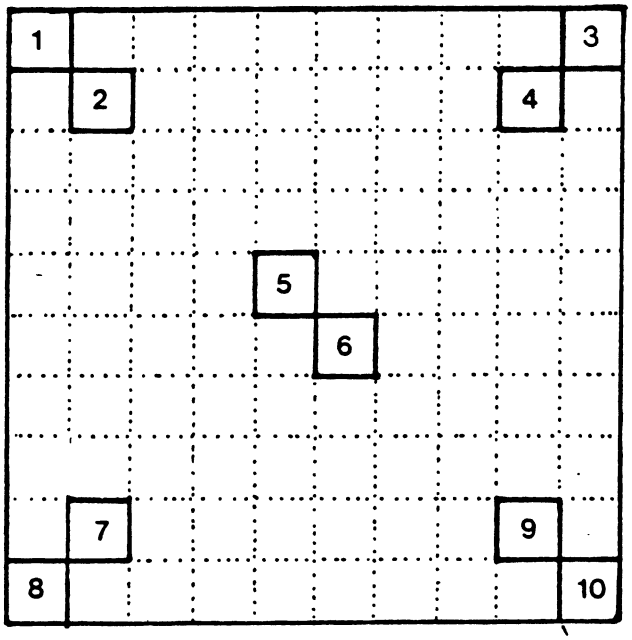


Figure 5. Detail of sample area and quadrat recording.

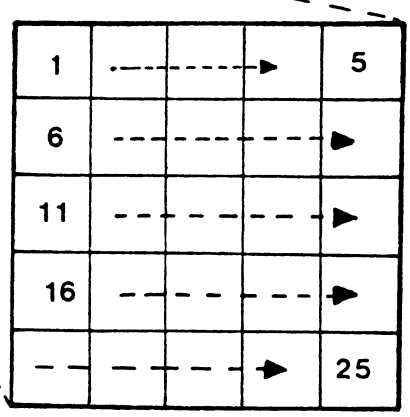
5 i) Layout of marker tapes on the 10m x 10m sample area.



5.ii) Layout of quadrats in 10m x 10m sample areas and order of sampling.



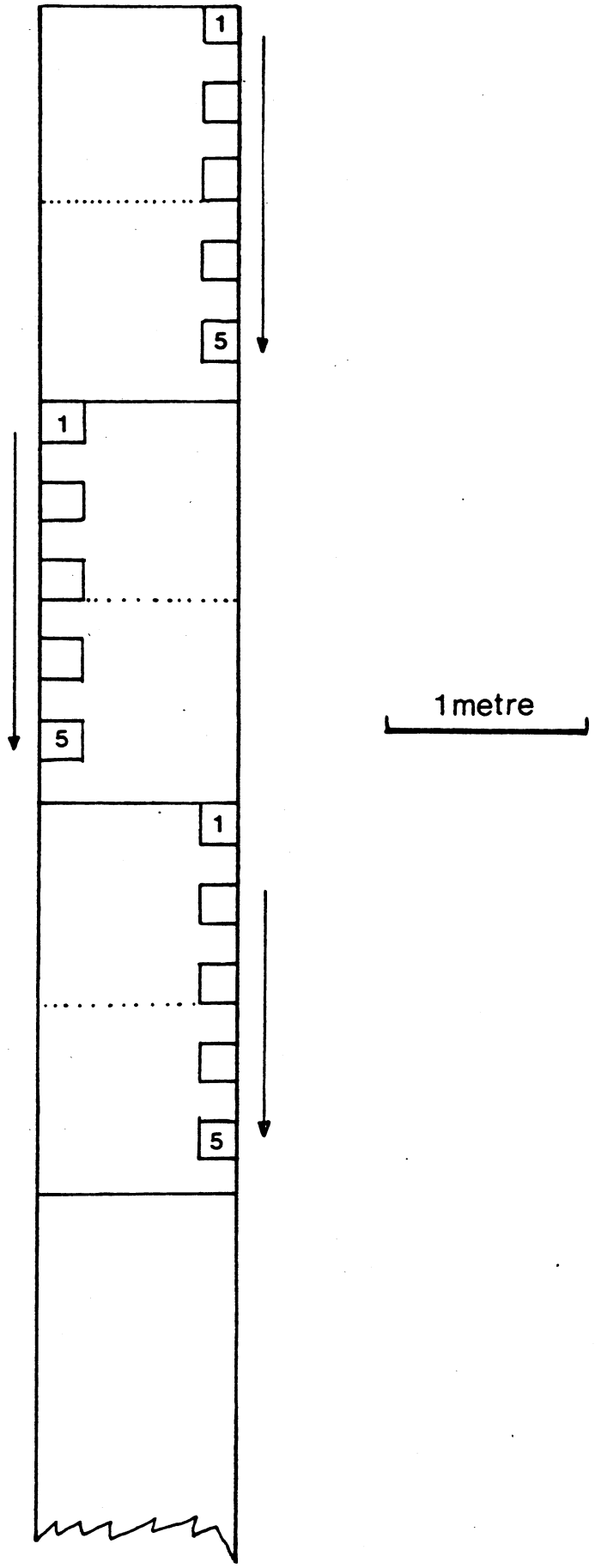
5.iii) Layout of 1m x 1m quadrats showing sub-divisions and the order of sampling.



APPENDIX D. STAND DATA - REORDERED BY EYE (FREQUENCY (F/25) OF SPECIES IN EACH QUADRAT OF EACH STAND (ie. 10x10m plot)

STAND	ERRENCE 553										253N										253E										PX11					
QUADRAT NO.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6
LEUCANTHEMUM VULGARE	7	9	15	20	20	15	12	14	10	11	1		3	7	15	18	18	14	19	15	1	2	1	15	1	2	7	11	6	12	5	2			2	5
RANUNCULUS BULBOSUS	16	19	18	22	19	18	24	23	21	18	15	15	16	13	24	23	22	23	25	25	25	25	22	24	24	19	25	23	22	20	21	20	17	24	18	24
RUMEX ACETOSA	1	1	5	2		1	3	4	1		1						2		3		1	2	1	1	3		2		9		3	6	1		1	6
SANGUISORBA OFFICINALIS	25	25	25	24	25	25	25	25	23	25	22	25	25	25	25	25	25	25	25	25	24	25	25	25	25	25	25	25	25	25	25	25	25	25	24	25
TARAXACUM OFFICINALE																								1			1	1		1	2					
TRIFOLIUM PRATENSE	16	19	22	21	24	18	21	6	14	21	8	7	13	16	15	16	17	16	18	16	13	10	14	20	18	20	25	21	19	13	10	9	13	16	17	17
BELLIS PERENIS																																				
CERASTIUM FONTANUM		2	3	1		2	2	4	1	1	4	3	3			2	3	1	3	2		3	1	1		1	1	1			4		4	2	2	1
LOTUS CORNICULATUS	11	2	7	7	3	9		3	5	7			1	1	5	2	3	4	2	7	3		1	2	2	1	1		2	4			1	3	14	
LUZULA CAMPESTRIS		2		2	2	1	2	1	3	1		3		8	4	2	2	3	3	3	1		2		1	16	2	1								
RANUNCULUS ACRIS	3				1	3	6				9	5	9	7	2	1		1		11	1	2	4	5	11	15	8	14	22	12	19	18	16	11	19	9
RHINANTHUS MINOR	25	25	25	25	25	25	25	25	25	25	2		4	6	2	2	14	13	14	6	23	25	23	24	21	25	22	23	18	18	5	8	7	5	8	5
FRAXINUS EXCELSIOR												1	3			1				1			1													
FILIPENDULA ULMARIA					1	2													1		5	1	9	3	4	11		3	15	11		5	6	1	3	4
HERACLEUM SPHONDYLIUM																																				
PRUNELLA VULGARIS	1	5	4	1	5		3	3	3	17	2	4	5	13	9	4	5	11	8	13	3	2	15	11	5	4	7	2	3	5	4	5	4	2	2	6
LATHYRUS PRATENSIS	2		3	3	1		8	2	2	1				3			2		1	3										1	1	3	2		1	1
VICCIA CRACCA		1	12	14	3		5	6	8	6		2	6	5	7	5	4	2	10	2		5	5	2	3		3	4	3	2		3	2			1
PRIMULA VERIS	16	11	13	12	18	14	17	12	18	18	7	3	3	2	11	11	2	5	8	7	1		1	7	1	1	7	5	6	4	12	14	3	3	10	1
CENTAUREA NIGRA	3	1	2	2	9	12	9	6	9	3						1		1	2	1	6		3	4	1	5				1	5	7	1	10	10	12
CIRSIUM ARVENSE																																				
LEONTODON HISPIDUS																								2			1	7	2	12	1			1	1	
PLANTAGO LANCEOLATA	1	4	5	3	5	10	8	5		4	1	1	1		1			2	2	1		1	1	13	2	3	11	15	3	9		1		1	2	1
TRIFOLIUM REPENS									18	6																										
CRATAEGUS MONOGYNA								1																												
GALIUM MOLLUGO																																				
RANUNCULUS REPENS										1																										
LYSIMACHIA NUMMULARIA																															9	10				
SUCCISA PRATENSIS	20	13	21	21	21	18	19	10	25	21	19	16	12	22	21	18	19	19	20	19	18	13	7	13	11	18	14	15	3	5	9	8	5	10	16	13
CARDAMINE PRATENSIS																																				
RUMEX CRISPUS																																				
LEONTODON AUTUMNALIS				1		1																								1						
SILAUM SILAUS	1		1	1	3	5	3	1	2	5					1								1			1				1	1				3	1
TRAGAPOGON PRATENSIS																																4				
CAREX FLACCA				7	8	13			4			1				4			3	10			3	6	2	17	2	9	12	10	2	6	8	16	8	6
EQUISETUM PALUSTRE																				1																
OPHIOGLOSSUM VULGARE													2	3										3	3		2	1				3		1	1	
CREPIS BIENNIS																																				
CAREX ACUTIFORMIS																																				
CALTHA PALUSTRIS																																				
THALICTRICUM FLAVUM																																				
POLYGONUM PERSICARIA																																				
LINUM CATHARTICUM	8	6	12	13	13	16	4	15	16	11	1			1	1	12	18	10	18	11													2	1	1	
MEDICAGO LUPULINA																																				
POTENTILLA REPTANS																																				
GALIUM APARINE																		1															1			
LEONTODON TARAXACOIDES																								1												
AJUGA REPTANS	3	2		5	4		8		7		4		1		5	4	3	3			2					3	1						22		7	1
CAREX PANICEA	12	15		8	17	1	20	17	15	21													15	6					1	1		1			17	19
QUERCUS ROBUR																																				
JUNCUS INFLEXUS																																				
CAREX NIGRA						5				1	1								1																	
GALIUM VERUM																				10																
SERRATULA TINCTORIA	1		8	4	1	1	1	2	14	6								1	1	2																
CAREX RIPARIA																																				
POLYGALA VULGARIS	12	13	5		12	10	18	17	12	19																						2				
POTENTILLA ERECTA	6	5	13	7	6	1	7	8	4	8																						3				3
FILIPENDULA VULGARIS	6	6					21	12																												
ORCHIS MORIO				1																																
LYCHNIS FLOS-CUCULI							1																													
TRIFOLIUM DUBIUM			2		1																3	3	3	1	1			1			1	2	4		3	5
JUNCUS ARTICULATUS																															4	8		1	5	11
ALLIUM VINEALE																																1				
CREPIS CAPPILARIS																																				
CIRSIUM PALUSTRE																																				
CAREX HIRTA																																				

Figure 7. The layout of sampling sub-quadrats on the transect.



APPENDIX H. TRANSECT DATA (Frequency (F/5) of species in each 1m x 2m section of the transect)

DISTANCE AL TRANSECT (METRES)	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-22	23-24	25-26	27-28	29-30	31-32	33-34
SPECIES																	
CAREX FLACCA	4	2	1	2	1	5		2		3	5	1	2	3	1		1
CAREX PANICEA	2	2				1			1	2	4	2	4	4	1	2	
LINUM CATHARTICUM	2						1			4	2	1	4	2	2	1	
LOTUS CORNICULATUS	2	2				1				1	2	4	1	3			1
POLYGALA VULGARIS	3			2	1			1			1						
PRIMULA VERIS	3	2		4	2	5	5	2	1	4	4	4	4	3	2		1
PRUNELLA VULGARIS	1		2	1	1	1	1			2	1	3	2	4	4	3	1
RANUNCULUS ACRIS	2	2	2	3	2	2	2	2	2	3	3	3	3	4	1	5	5
RANUNCULUS BULBOSUS	1	3	5	4	4	4	3	4	2	4	3			1			
RHINANTHUS MINOR	3		2	4		3	1		2	5	5	5	5	4	5	5	5
RUMEX ACETOSA	1	2	2		2		3	1	2	1				1	2	2	1
SANGUISORBA OFFICINALIS	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
SUCCISA PRATENSIS	4	1	2	4	2	4	3	2	2	5	4	5	3	4	3	2	2
CENTAUREA NIGRA		1	1			1	1			1	1	1			1		
CERASTIUM FONTANUM		1				1		1	1	2			2				
LATHYRUS PRATENSIS		1							1	1	2	1					
LEONTODON HISPIDUS		1								1	2	5	4	4			
LEUCANTHEMUM VULGARE		1	1	3	1	1	1			4	2	4	3	1			
TRAGOPOGON PRATENSIS				1												1	
TRIFOLIUM PRATENSE	3	2	1	3	4	2	3	1	2		2	1	3	5	1	1	1
TRIFOLIUM REPENS		1			4	1			1		2	2					
VICCIA CRACCA	1																
FILIPENDULA ULMARIA					1	2		2			1				1	1	2
LYSIMACHIA NUMMULARIA								2		2	1				2	1	
OPHIOGLOSSUM VULGARE					1	1			2	2				2			
SILAUM SILAUS						1		1	1	1		1	1	1	1		
TRIFOLIUM DUBIUM							1		1		1		1		2	2	1
TARAXACUM OFFICINALE											1	1	1	2	2	3	1
VALERIANA DIOICA											2			1			
LEONTODON TARAXACOIDES									1								
CAREX ACUTIFORMIS															1		1
CAREX NIGRA													1				2
GALIUM APARINE																	
JUNCUS INFLEXUS																3	
LEONTODON AUTUMNALIS													1	2		3	2
PLANTAGO LANCEOLATA													1			2	2
CAREX HIRTA																	
MEDICAGO LUPULINA																	
SENECIO AQUATICUS																	
QUERCUS ROBUR																	
EQUISETUM PALUSTRE																	
GALIUM VERUM																	
JUNCUS ARTICULATUS																	
POTENTILLA REPTANS																	
FRAXINUS EXCELSIOR																	
CARDAMINE PRATENSIS																	
RANUNCULUS REPENS																	
CRATAEGUS MONOGYNA																	
STELLARIA PALUSTRIS																	
AJUGA REPTANS																	
HERACLEUM SPHONDYLLIUM																	
CAREX SP.																	
BELLIS PERENNIS																	
MYOSOTIS SCORPIOIDES																	
LYCHNIS FLOS-CUCULI																	
THALICTRUM FLAVUM																	
TOTAL NO. OF SPECIES	15	16	11	12	14	18	13	13	16	20	23	18	20	20	18	17	17

Data Set 10

Flood Plain Meadow Monitoring

Data consists of frequencies of presence of each species in 25 (or 5) sub-quadrats. For some species such as Sanguisorba the score is often 100 per cent, indicating that the quadrat size may be too large to enable differences to be detected. If that species were of primary importance it might be preferable to use a smaller size of quadrat.

1. *Ordination of sample plot data*
Ordination and cluster analysis may be performed on these data in the same way as the German Meadow samples reproduced in Gauch's 'Multivariate Analysis in Community Ecology', illustrated in Appendix 2.

The first problem is to define an appropriate similarity or distance measure. The simplest is to use presence absence only and to use a matching coefficient. There are two forms, the Jaccard coefficient ignoring double absences, and the simple matching coefficient which includes them. These coefficients rely on the rarer species to provide the necessary contrasts.

To make use of the detailed quantitative information it is necessary to decide how to transform the scales, and how to compensate for the different maximum frequencies (the range of values formed for each species). No method is perfect, some giving more weight to the rarer species and others to the species that occur with medium frequency. The proportions out of 25 are not binomial variables but measures of abundance related to a density scale by the formula

$$\text{density} = -k \log (1 - p)$$

where p is the expected proportion of quadrats in which the species occurs. We may observe 25/25 but cannot infer certainty of observation, which would imply infinite density. The formula assumes random distribution of individuals over a small area. If we then wish to compare plots on the basis of log density, this implies use of the complementary log log transformation, which is not very dissimilar to the logit transformation. In practice this means that a similarity coefficient such as the commonly used 'city block metric' or mean absolute difference in observed scores, would not be

very misleading except in respect of the very rare or very common species.

Having chosen a similarity measure, GENSTAT or CLASP will provide cluster analysis and principal coordinates, while TWINSPAN and DECORANA will provide their own versions of the same. The details will differ but the major contrasts between the plots, and the species responsible for these contrasts will be the same, apart from the possible placing of outlier plots with a number of rare species.

The final question relates to the possibility of statistical analysis of ordination scores with respect to predefined groups (10 plots on each stand). Since the grouping is independent of the variables analysed, it is certainly possible to conduct an analysis of variance for each coordinate separately.

An alternative analysis is Linear Discriminant Analysis, which assigns weights to each species to provide an overall score which maximises the differences between each plot. This will provide a set of axes on which the data may be plotted to give maximum separation between each pre-assigned group. This is easily performed by GENSTAT.

2. *Correlations between species frequency and water level*
If water level is available for each plot, it would be possible to compute the correlations between species frequency (or the complementary log log transformation of frequency) and water level, for each plot. Graphs should be drawn of those species for which a significant relationship is found, as the scatter may not be at all linear. The success of this method depends on how much variation there is between plots in respect of water level. It should work better on the transect than for the four stands. There is of course the possibility that the correlation with water level is accidental and not to do with the water requirement of the plant.

There are probably too many species to be analysed conveniently by multiple regression, but it would be possible to use some version of canonical correlation analysis to identify the linear combination of species frequencies most correlated with water level.

3. *Correlations with transect distance and water level*
If water level is not strongly correlated with transect distance it would be possible to use canonical correlations to find out which species depend strongly on distance and which on water level. The problem with distance along a transect is that, although there may be patterns along the line, the relationship may not be linear or continuing, so that a trend may be established but later reversed.

With such small numbers (0 to 5 out of 5) there is not a great deal that can be expected from simple correlation analysis. A more sophisticated method of analysis is to relate frequency to water level or distance using maximum likelihood estimation of the slope in a generalised linear model. This can be accomplished using the programs GLIM, GENSTAT or MLP. The probability of occurrence at any one site is assumed to be related to distance or water level via the logit or complementary log log transformation. The significance of the relationship for each species may be computed, and those of interest displayed graphically.

4. *Alternative analyses*

The transect data would allow some estimate of aggregation by combining neighbouring sections (giving scores out of 10 or 20) and comparing the variances at each level of aggregation.

ΓA SET 11 - VEGETATION MONITORING IN EAST ANGLIAN FENS

Background

English Nature is concerned about the deterioration of a number of fen habitats in East Anglia that may have resulted in significant loss of fen species and scarce fen communities. Changes in hydrology and management are thought to be two important factors influencing these losses. 'Historical' data on the vegetation of a number of fens are available. Information was collected in the late 1950s by Dr David Bellamy. In 1991, W Fojt and M Harding began a re-survey of these fens.

Objectives

To assess whether significant losses of fen communities and species have occurred and whether changes in hydrology and management can be related to these losses.

Sampling design and information collected

Thelnetham Fen, a calcareous valley fen, provides an example of the way monitoring has been carried out.

In 1959 Dr Bellamy sampled one representative, temporary 10 x 10m plot from each homogenous vegetation type in the fen. Within each plot he recorded 20, ½ x ½m random quadrats. He recorded all higher plants and bryophyte species and assessed their cover-abundance using Braun-Blanquet scale. (See Appendix 1 and data sheet - only 10 of the 20 quadrats are shown.) He also gave a figure for cover/frequency for each species eg $0.33/_{33}$ (see attached data sheet). The top figure is the sum of the mid-points for the cover-abundance range represented by the Braun-Blanquet scale (eg 5 (80-100%) = 90%, 4 (60-79%) = 70% etc with x = 1%), divided by the total number of quadrats recorded. This figure is regarded as a average measure of the cover of the species in the plot. The lower figure is percentage number of quadrats in which the species occurred and is taken as a measure of the frequency of occurrence in the plot.

Fojt and Harding have repeated Bellamy's work as closely as possible. The methodology is the same and the 10 x 10m plots are located within the general area indicated by his maps - although they are not in the exact same location.

Analysis

It is hoped to study changes within communities in individual sites but also differences between sites which had the same communities in 1959 but have had different histories since then in terms of management and hydrological change. It might be possible to find several examples from one community type say which have had management change (usually cessation of management) but no known hydrological change and several that have undergone both types of change.

Questions

1. For individual sites and communities, can the data be analysed to discover if there are significant changes in species frequencies and cover from 1959 to 1991 and if so what are the magnitude of these changes? Can changes in proportions of fen species to 'other' species be analysed?

Can changes in individual communities between 1959 and 1991 be analysed using the data on the species composition, cover and frequency of component species, for example in some kind of ordination?

3. Can data from several communities in different sites be analysed together, say by using ordination. Alternatively can the data be analysed by assigning samples from the same original communities to different 'histories' such as 1) cessation of management only and 2) change in management and hydrology and subsequently analysing species and community changes?

APPENDIX 1

Braun Blanquet Scale

1959	1991 - altered to remove the slight ambiguity of the 1959 estimates
5 = 80-100%	80-100
4 = 60-80%	60-79
3 = 40-60%	40-59
2 = 20-40%	20-39
1 = 2-20%	1-19

x = small cover value

1959 DATA.

Bellamy. Area II. Thelnetham.

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	
~~Cladium mariscus~~	—										
Molinia caerulea	1	3	3	2	K	2	1	K	K	K	13.4 /100
Phragmites communis	K	K		K	1	K	·	K	2	2	9.4 /80
Juncus subnodulosus	1	K		K	K	K		K	2		6.4 /70
Eupatorium cannabinum	·		K		K	K			K	K	1 /50
Berula lacerata	K		K	K	K	K		·	·	1	2.5 /60
Schoenus nigricans	1	K	K	1	3	2	2	·	·	·	19 /70
Valeriana dioica	1	1	K	1		K		K	·	K	4.9 /70
Festuca arundinacea	K	1	1	1		1		·	1	·	8.5 /60
Calliergon cuspidatum	·	1	·	1	·	2	2	·	3	·	26 /50
Campylium stellatum	X	X	X	X	2	·	·	X	X	·	5.8 /60
Hydrocotyle vulgaris	·	·	·	·	1	·	·	·	1	1	10 /30
Mentha aquatica	·	K	·	K	K	·	K	·	·	·	1 /40
Parnassia palustris	K	·	·	K	·	K	·	·	·	·	1 /30
Succisa pratensis	K	·	K	·	K	1	·	·	·	·	3.25 /40
Drepanocladus revolvens	K	·	·	·	·	1	·	·	·	·	5.5 /20
Mnium pseudopunctatum	·	·	K	·	·	·	·	·	·	·	1 /10
Riccardia pinguis	K	·	·	·	1	·	·	1	·	·	7 /30
Alnus glutinosa	·	·	·	·	2	·	·	·	2	·	30 /20
~~Cattha palustris~~	—										
Carex lepidocarpa	·	·	·	·	·	·	·	·	X	·	1 /10
panicea	·	·	X	·	·	·	·	X	·	·	1 /20
Dactylorhiza praetermissa	·	·	·	·	·	·	·	X	·	·	1 /10
Salix palustris	·	X	·	·	·	·	·	·	·	·	1 /10
~~uliginosum~~	—										
~~Lythrum salicaria~~	—										
Sanguisorba officinalis	·	X	X	·	·	·	·	·	·	·	2 /20
~~Typha angustifolia~~	—	—	—								
Vicia cracca	·	·	·	·	·	·	·	·	·	X	1 /10
Calliergon giganteum	·	·	·	·	·	·	·	·	·	1	1 /10
Bryum pseudotriquetrum	·	·	·	X	·	·	·	·	·	·	1 /10
Campylium elodes	·	·	·	·	1	·	·	X	X	·	4 /30
Chiloscyphus polyanthos	·	·	·	1	·	·	·	·	·	X	5.5 /20
Mnium seligeri	X	·	·	·	·	·	·	·	·	·	1 /10
Pellia fabroniana	·	X	·	·	·	·	·	·	·	·	1 /10
Philonotis calcarea	·	1	·	·	·	·	X	·	·	·	5.5 /20
Riccardia multifida	·	·	·	·	X	·	·	X	·	·	2 /30
R pinguis	X	·	·	·	1	·	·	1	·	·	7 /30

Tholenham Area II

1991 DATA

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩		
Filipendula ulmaria	5	2	3	5	4	3	3	2	5	2	58	/100
Cirsium arvense	1	3	4	1	4	2	2	2	1	3	36	/100
Juncus subnodulosus	1	1	1	1	1		1		1	1	10	/80
Phragmites australis	1	1			1	1	1	1	1	1	10	/80
Sanguisorba officinalis		1	1			1	1	1		1	10	/60
~~Eupatorium Speciosum~~												
Lythrum salicaria		1		1	1						10	/30
Molinia caerulea				1				1		1	10	/30
Eupatorium cannabinum		1									10	/10
Carex acutiformis							1				10	/10
~~Juncus effusus~~												
Brachythecium rutabulum								1	1	1	10	/30
Potentilla erecta	1			1							10	/20
~~Schoenus nigricans~~												
Humulus lupulus			1								10	/10
Lotus uliginosus						1					10	/10
~~Centaurea nigra~~												
~~Succisa pratensis~~												
Fraxinus excelsior								1			10	/10
Valeriana officinalis									1		10	/10
Campylium stellatum									1		10	/10
Carex panicea									1		10	/10
Carex flacca									1		10	/10
Viburnum opulus												
~~Lathyrus pratensis~~												

Data Set 11

Vegetation Monitoring in East Anglian Fens

This data set compares the original observations of Dr David Bellamy in 1959 with recent observations on the same sites.

Bellamy's data records cover abundance of all higher plant species using the Braun-Blanquet Scale. Various vegetation types are chosen, with one plot for each, sampled at 20 random sub-quadrats.

1. *Changes at individual sites*
The main changes at each site may be displayed graphically by plotting frequency or mean cover in 1959 against the same score in 1991. Points far removed from the line of equality indicate likely candidates for significance of change, either increasing or decreasing.

To test for changes in frequency of occurrence the binomial test may be used, with the proviso that the sampling distribution may not be strictly binomial. Over such a period of years there is almost no suggestion of repeated measurements, and the sub-quadrats are in any case not the same.

To test for changes in cover percentage, one should not use Bellamy's mean score because it only refers to quadrats where the species was present. It would therefore be necessary to include the zero scores, so that for example, in the second line of the data where the mean cover is given as 9.4 in 8 plots out of 10, that would correspond to overall mean cover of 7.5 percent.

To test the differences between two mean scores the only available estimate of standard error is from the 20 replicate quadrats, and the two sample t-test may be used. For example we have for *Schoenus nigricans*

Data	1	×	×	1	3	2	2	0	0	0
Score	10	1	1	10	50	30	30	0	0	0

The changes in proportions of groups of species, in particular 'fen' species and 'others' is a different question, as it relates to different entities. The mean cover of the two groups can be calculated, but it is difficult to assign it a standard error, except by pooling the variances of the component cover scores, if they have been computed.

A simple binomial test of the proportion of fen species in each quadrat would at least give an indication of the significance of change, set out as a 2 × 2 table.

	In 1959	In 1991
Fen	——	——
Others	——	——

2. *Ordination of individual communities*
Ordination and classification of all quadrats in both years together should give a picture somewhat similar to the analysis of agricultural acreage in English counties in Appendix 1. If there has been systematic change since 1959 the later dated points will appear displaced in the same general direction, whereas if there has been random or little change, the displacements will show no particular pattern.

3. *Ordination of several communities*
Ordination of all sites and years, possibly pooled over all quadrats to reduce the amount of data, will encompass a wider range of variation than that of individual sites. The changes over time may indicate convergence of some sites which are more similar in 1991 than in 1959, or vice versa.

If sites can be sub-classified by history of management, then these sites can either be analysed on their own, or as part of the main set forming a pre-identified group. The analysis will then be similar to that of data set 10, where ordination scores may be treated as random variables in a between group analysis of variance. However it may be difficult, with only two end points and many possible causes of change, to establish what has been responsible for the changes.

TA SET 12 - MONITORING OF NORTHERN HAYMEADOWS

Background

The management of a number of species-rich northern haymeadows has been changed over the past few years, mainly through reduction in the use of inorganic fertilisers. English Nature is interested in the changes in species occurrence and abundance that may occur as a result of these reductions and whether fields that continue to receive higher levels of fertiliser continue to deteriorate over time in terms of species richness and composition.

Objective

To assess the changes if any, in species composition and relative abundance in meadows receiving two levels of inorganic fertiliser.

Sampling design and information collected

Ten fields from each type of fertiliser regime and containing the same NVC community type were selected randomly. In each field 5 1m by 1m permanent quadrats were laid out diagonally across the field at regular intervals. Plant species present and their cover, measured using the DOMIN scale, were recorded. Quadrats will be re-recorded annually, at least initially. A sample data sheet from a quadrat is attached.

Questions

1. Can species frequencies, calculated as presence in the 5 quadrats be used as a measure of the species frequency for the field?

2. Can an average of the cover values from the 5 quadrats be used as a measure of the cover of a species in the field?

3. Can these averages be used to generate mean values for the two types of fertiliser treatment at different times and can these means be compared statistically?

4. The total number fields that continue to receive higher levels of fertiliser is greater than the total number that now receive lower levels, thus the sample of 10 represents a lower proportion of these fields. Should samples be proportionate to the size of the population if this is known?

SITE MONITORING : GRASSLAND RUNNING RECORD (DOMIN VALUES FOR 1X1m QUADRATS)

SITE NAME : LITTLE ASBY INRAKES AND OUTRAKES QUADRAT CODE : LA1

SPECIES: \ DATE:	13.8.87	31.7.89						
Carex panicea	7	5						
Molinia caerulea	7	5						
Festuca rubra }	5	4						
Festuca ovina }								
Succisa pratensis	5	5						
Carex pulicaris	4	2						
Primula farinosa	4	4						
Carex sp. 89: hostiana	3	3						
Juncus articulatus	3	1						
Leontodon hispidus	3	4						
Parnassia palustris	3	3						
Ranunculus acris	3	2						
Serratula tinctoria	3	4						
Trifolium pratense	3	2						
Valeriana dioica	3	2						
Briza media	2	2						
Carex flacca	2	1						
Cirsium dissectum	2	2						
Potentilla erecta	2	2						
Prunella vulgaris	2	2						
Sanguisorba officinalis	2	2						
Trifolium repens	2	1						
Euphrasia sp.	1	1						
Lotus corniculatus	1	1						
Plantago lanceolata	1	1						
Juncus gentiflorus		3						
Cirsium palustre seedling		1						
MEAN VEG HT. (cm)	15	15 - 40						
GRAZING	LIGHT - MEDIUM	Little						
TOTAL VEG COVER (%)	100	100						
OTHER (SPECIFY)								

Data Set 12

Monitoring of Northern Haymeadows

The experiments described here resemble the early history of the Rothamsted Park Grass Experiment, where species composition changes associated with fertilizer dressings have been monitored since 1856.

1. *Measuring Species Frequency*
If by species frequency is here meant the relative abundance of each species in a sample, then the question relates to the probability of observing a given species in 0 to 5 of the 5 quadrats. If the question simply refers to species richness, the total number of species observed in the field, that is a different matter.

The relationship between the total number of species observed and the area sampled was studied by C.B. Williams (*Patterns in the Balance of Nature*). He found empirically that the number of species recorded increased linearly with the logarithm of the area sampled and the extra number of species for a given size increase was used as an index of diversity. This relationship can be justified theoretically if the sample counts (of individually countable species) follow a logarithmic series distribution.

If it is meant that the number of quadrats (out of 5) is to be used to estimate species density for a given species, then there are only six outcomes, of which 0 and 5 give little information. Even if no specimen is observed in any of the 5 sample quadrats that does not mean the species is absent entirely. Indeed, if the species consists of individual specimens with a negative binomial distribution with mean m and dispersion parameter k, then the probability of observing the specimen in each of 5 samples is (1 + m/k)**(5 - k). For example, with the Fritillaries in data set 9, m is 7.76 and k is 0.046, and the probability of observing no specimen is 31%.

In either case it is clear that presence in at least one quadrat is a minimum measure of presence in the field, and the total number of less frequent species may well be much larger, depending on the species diversity and on the degree of aggregation of each species.

2. *Measurement of cover value*

The mean cover, as represented by central values on the DOMIN scale, is the best estimate available of mean cover, although the quadrat size may be more suited to some species than to others. The variance of the estimate can be obtained directly, although with only 5 quadrats and a discrete scale there is a possibility that all the scores will be identical and no variance can be estimated.

3. *Analysis of mean cover scores between treatments over time*

Treating mean cover score as a random variable with unknown distribution, plots of mean cover against time for the two treatments may be made. Because the quadrats are permanent there will be serial correlations in time, but the time trends, fertilizer effects and interactions (variations in fertilizer effect with time) will be displayed.

The statistical analysis of all ten fields simultaneously, treated as a three factor system (fields, treatments and years) is hampered by the lack of internal replication by which the local error variance may be estimated. Although some general statement may be made about the magnitude of the fertilizer effect and how it changes with time, it is not possible to make reliable significance tests if the fields are widely separated, because the between-field variance will be dominant.

4. *Should samples be in proportion to population?*

If we are making comparisons between two groups of treatments there are advantages in having equal numbers in each group, so that the differences between treatments are measured with greatest efficiency.

The idea of sampling in proportion to population values is only relevant in sample surveys aimed at estimating, say, the overall productivity of farms. Even then it is more important to guarantee a minimum sample size for each group than to achieve proportionality.

ENGLISH NATURE Contract No. F72-07-56

Appendix 1

Analysis of changes in agricultural land use

The following data set is used to illustrate how changes in species cover with time may be displayed, and how contrasting sites may be subjected to multivariate analysis in order to summarise the main observed changes.

The source of the data is the series of agricultural censuses recording the acreages devoted to different crops in different counties. The years chosen are 1945, 1955, 1965 and 1975 (there is too little detail in the 1985 census), and the counties are chosen partly because they display minimal boundary changes during the period, and partly to provide major contrasts. The crops are grouped in such a way as to avoid major changes in definition, although maize was not recorded before 1975.

The acreage for each crop in any year is expressed as the percentage of the total agricultural acreage for that county. Each year for each county is treated as a separate data unit with 21 variables.

The 40 sets of records, ten counties in four years, were subjected to several different multivariate analyses. The distance measure used standardised ranges for each variable, so that the 21 variables were given equal status, in spite of the large differences of total acreages for each crop. This meant that contrasts due to minor crops were not lost from the analysis.

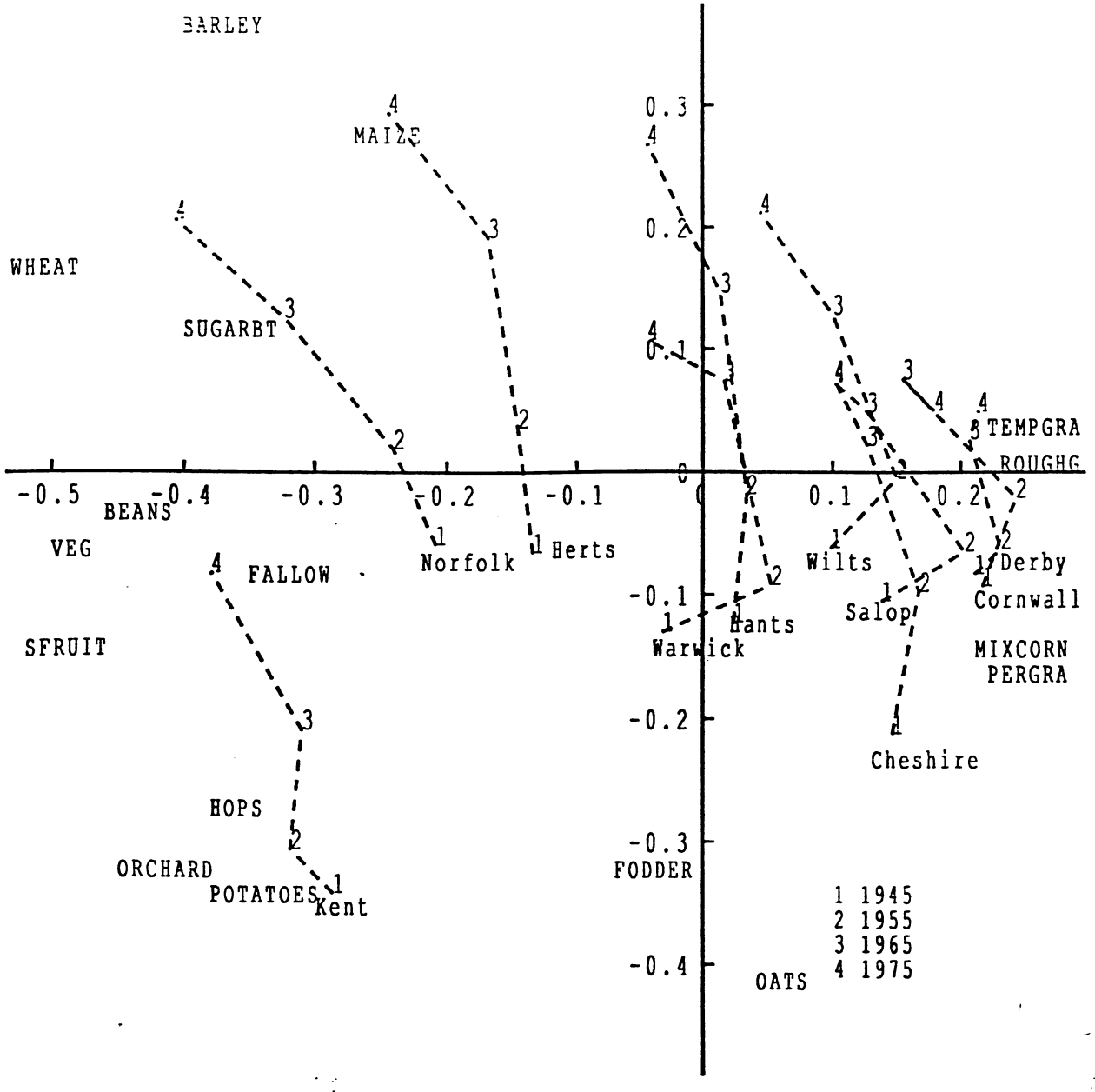
The data are here presented as a two-way ordination of the first two principal co-ordinate axes (by a technique broadly similar to correspondence analysis). Corresponding points for each county are joined by a dotted line, to show how the changes progressed with time. The directions of increase of each crop are indicated (for those crops associated with the first two axes).

Changes in Agricultural Land Use

Data from Agricultural Censuses, 1945, 1955, 1965 and 1975

Ten selected counties: Kent, Norfolk, Herts, Hants, Warwick Wilts, Salop, Cheshire, Derby, Cornwall.

21 classes of crop: Wheat, barley, oats, rye, maize, mixed corn, beans, potato, fodder roots, sugar beet, rape, hops, orchards, small fruit, vegetables, flowers and glasshouse, fallow, temporary grass and clover, permanent grass, rough grazing.



The diagram illustrates several facts simultaneously:

1) The major contrast between the arable counties and the livestock counties.

2) The minor contrast between Kent and other counties.

3) The crops associated with the major contrasts.

4) The general change towards more arable farming, and replacement of oats by barley growing, and increases in maize and sugar beet production.

5) The change between 1955 and 1965 being greater than at other periods.

6) The change being more pronounced in the arable counties.

Other changes are revealed when minor axes are studied, such as the increase in rape production and the decline in glasshouse crops.

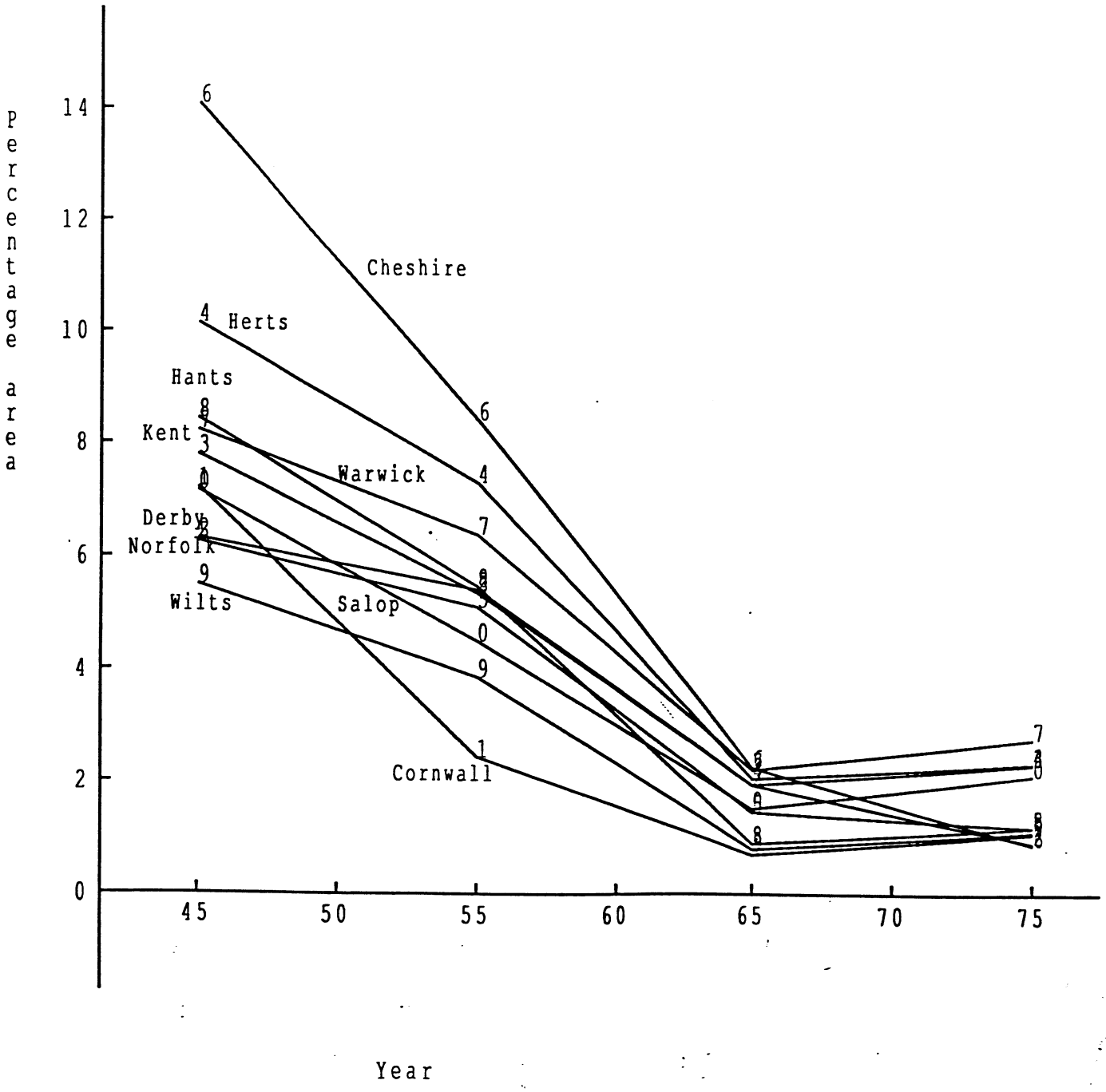
A detailed study of particular crops or pairs of crops can now be undertaken. The decline in oats and its replacement by barley are illustrated in the next two figures. Different patterns are observed in particular counties, which may suggest further possible studies.

This example is included in the Report in order to illustrate how changes of vegetation with time may be revealed and studied.

Changes in Agricultural Land Use

Changes in time for a particular crop.

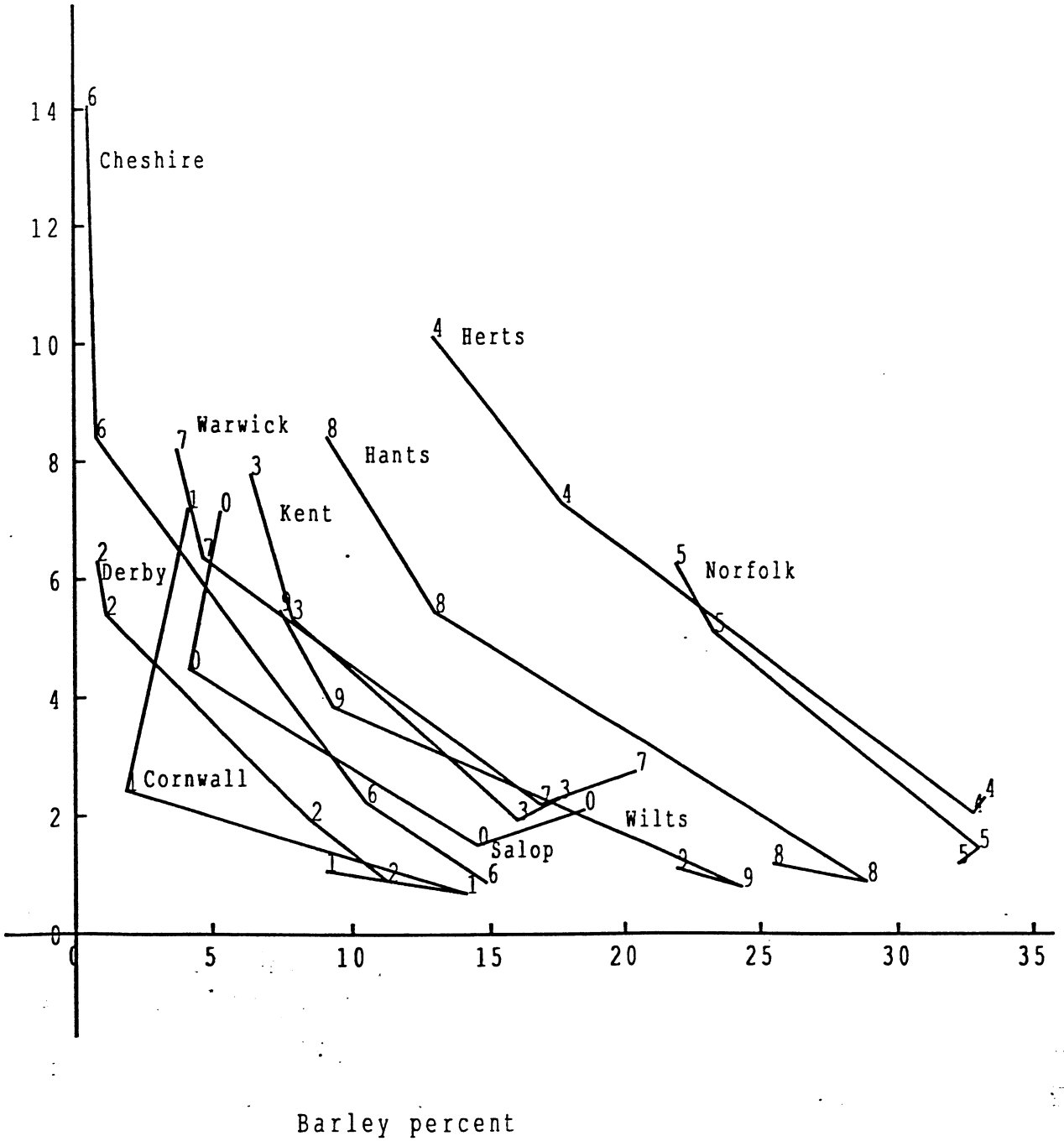
Decline of acreage of Oats between 1945 and 1975.



Changes in Agricultural Land Use

Changes in time for two competing crops.

Replacement of Oats by Barley between 1945 and 1975.



Appendix 2

Analysis of German Meadow Samples

The data set given in 'Multivariate Community Ecology' by H.G. Gauch (Ref. 40) (Table 1.4 etc.) was studied because it allowed comparison of TWINSPAN and DECORANA analyses with the more general clustering techniques available in GENSTAT and CLASP (Refs. 54 and 56) with which I am more familiar.

The data, consisting of 25 plots and 55 species measured on a cover scale, bears resemblance to several of the data sets discussed in this report.

Three main types of analysis are presented by Gauch:

1) Cluster analysis of the plots, broadly confirming the a priori classification into three major groups and one outlier.

2) Ordination in two dimensions, displaying both plots and species in relation to two axes. An environmental variable (wet-dry) is imputed to be associated with the first axis.

3) Reordering the data matrix to bring together species associated with each group.

The first problem is which similarity to use. Data are Braun-Blanquet cover scores, and it appears that these have been treated by Gauch as if they were a linear scale for the purposes of comparison between samples. Zero matches have been included in the similarity scores, and I have therefore used a City-Block metric, proposed by J.C. Gower, which includes zero matches. The Braun-Blanquet scale is not very different from a logarithmic abundance scale that I have used in similar circumstances.

Reordering of the rows and columns of the data matrix on the basis of the first principal co-ordinate axis, and the association of each species with that axis, gives a data matrix broadly similar to that presented by Gauch.

Table 1.4. *German meadow samples arranged by ranked detrended correspondence analysis ordination scores*

Species	Sample[a] B B B B B B B B G G G G G C G C C C C C C C C C O 1 1 2 1 2 1 2 1 1 2 2 1 1 1 1 1 2 2 1 1 4 0 9 3 5 2 4 6 0 2 8 5 3 2 3 4 7 6 7 8 1 1 5 9
Koeleria pyramidata	4 3 4 4
Thymus serpyllum	2 1 3 1 1
Festuca ovina	3 3 2 2
Bromus erectus	8 9 7 8 8 8 6
Scabiosa columbaria	1 2 1 1 3
Viola hirta	4 3 1 1 1 1
Salvia pratensis	5 3 4 2 4
Linum catharticum	1 1 1 1 1
Briza media	2 2 3 2 3
Campanula glomerata	1 2 2 2 1 1 2
Silene inflata	1 3 1
Festuca rubra	6 3 4 4 5 3 3 3 1 1 2 3 1 3 1
Achillea millefolium	5 2 5 3 4 3 1 4 5 5 2 6 3 1 4 4 6 1 2 1 1
Bellis perennis	2 1 1 1 2 1 1 1 1 1 1
Chrysanthemum leucanthemum	2 1 5 3 4 2 1 5 2 4 4 1 2 2 2 3 2 4 3 2 1 1 1
Carex panicea	3
Poa pratensis	4 5 5 6 6 6 9 7 6 5 7 6 4 6 6 3 5 5 3 5 4 5 2 6 2
Centaurea jacea	1 5 1 2 1 3 3 2 1 4 3 3 4 1
Helictotrichon pubescens	2 1 4 2 2 6 7 2 4 7 3 4 4 5 4 1
Leontodon hispidus	1 2 2 4
Euphrasia odontites	1 1
Taraxacum officinale	1 1 1 1 1 1 1 1 2 4 4 1 3 1 1 2 4 2 1
Pastinaca sativa	4 2 1
Tragopogon pratensis	1 2 1 3 2 2 1
Anthriscus silvestris	1 3 1 1 1
Arrhenatherum elatius	1 3 5 5 3 6 2 7 6 7 6 7 4 7 7 6 6 6 4 4 6 6 7 8 7
Crepis biennis	1 1 1 2 2 2 3 5 2 1 1 2 5 1 1
Phleum pratense	1 2
Sanguisorba officinale	1
Cardamine pratensis	1 1
Glechoma hederacea	1 1 1 1 1 1
Angelica sylvestris	2 1 2 2 3 1
Filipendula ulmaria	1 1 4 3
Alopecurus pratensis	3 3 6 4 5 6 5 6 2
Pimpinella magna	2 1 1 2
Lysimachia nummularia	1 1 1 1 2 1 1 1
Cirsium oleraceum	1 1 2 1 1 7 3 6 7 7 4 3 4 1
Melandrium diurnum	2 1 1 1 2 1 3 2 1 4 1
Geum rivale	2 1 1 1 1 1 1 3 2 5 4 2 1 2
Holcus lanatus	2 1 2 3 3 2 2 3 3 3 6 1
Lychnis flos-cuculi	2 1 1 1 1 1 1
Deschampsia caespitosa	2 6 6 3 7 5 3

The plot of the first two principal co-ordinates show some resemblance to the DECORANA plots, with the following differences:

1) DECORANA deliberately removes the 'horseshoe' effect, by which is meant the tendency for the endmost plots in a series to be plotted closer together. This effect is shown in the principal coordinate plot. The reason is that all plots with little in common have roughly the same distance measure, and the plots aim to fit the given distances as well as possible.

Many people are happy to accept 'horseshoes' because other features of the data are not distorted.

2) The differences between the two measures affects the outlier plot 19, which is assigned to the third axis in the principal coordinate plot. Plot 19 contains a number of species not represented elsewhere, and these are given greater weight by DECORANA.

The cluster analyses in Gauch are confirmed by a variety of clustering methods, but as the clusters are fairly clearly defined it is not surprising that the main features are revealed by most methods. These are not therefore discussed further.

German Meadow Samples

DECORANA plot (reproduced from H.G. Gauch)

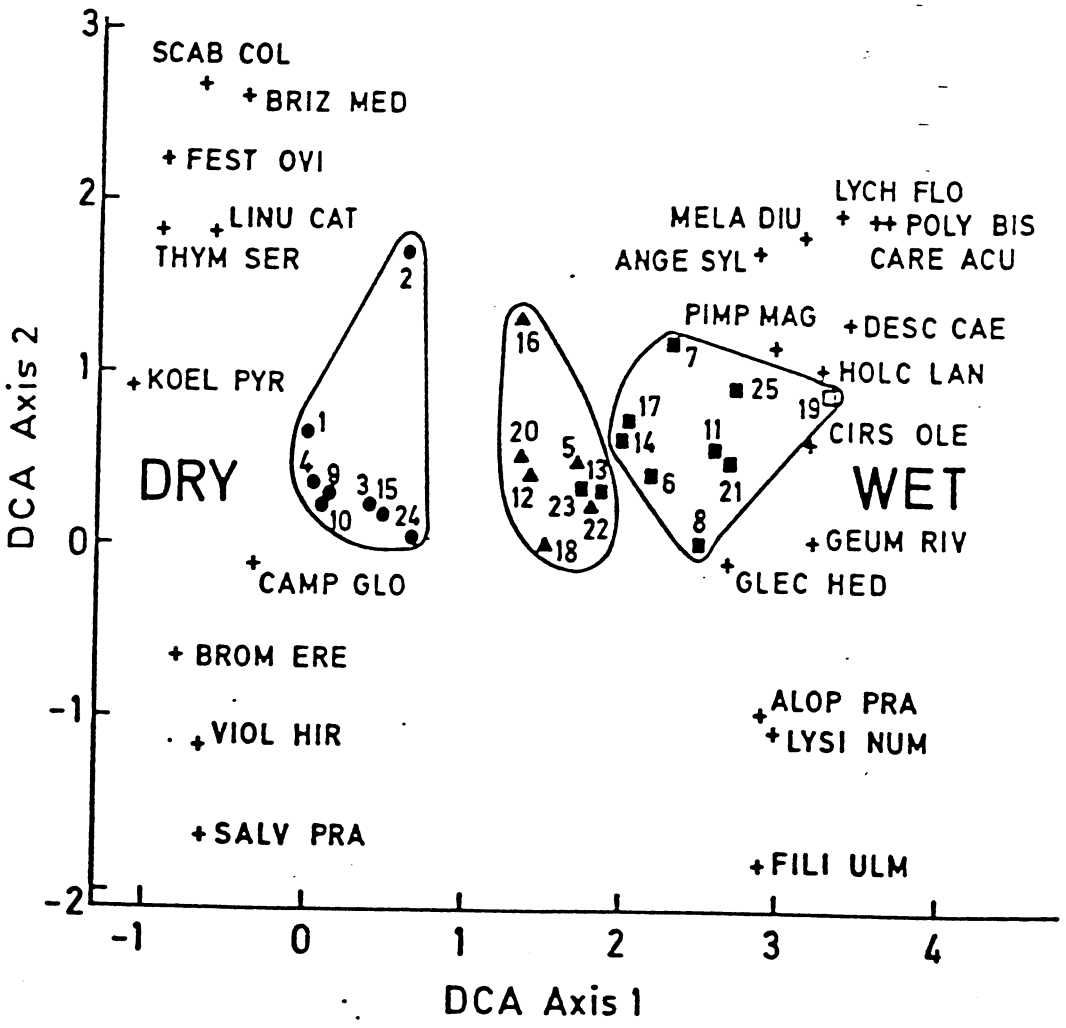
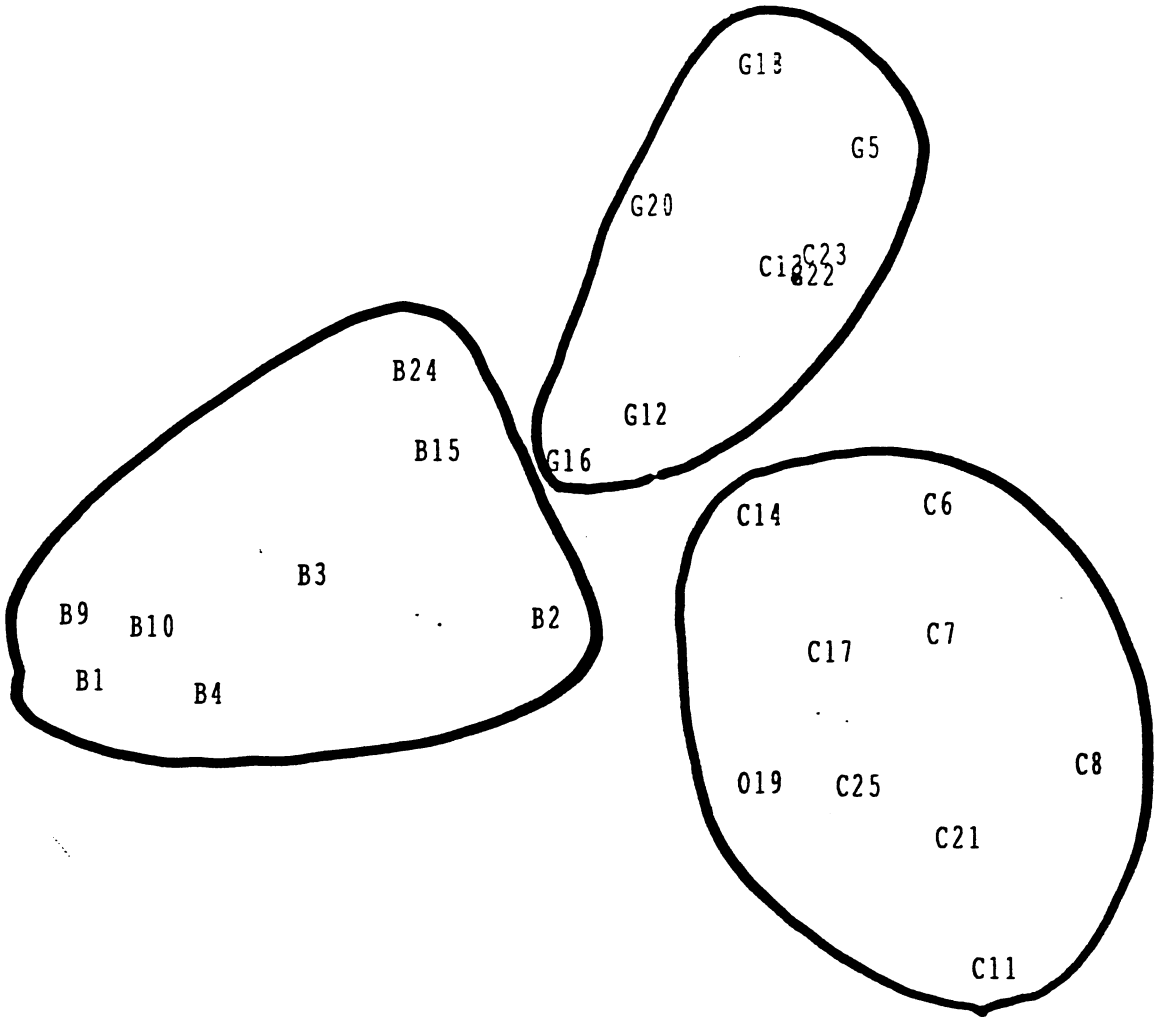


Figure 1.5. Detrended correspondence analysis (DCA) ordination of German meadow samples and species. The samples were previously classified by Braun-Blanquet analysis into three community types: *Bromus–Arrhenatherum* (•), *Geum–Arrhenatherum* (▲), and *Cirsium–Arrhenatherum* (■) with one outlier (□). An environmental interpretation of the first DCA axis is offered: a dry-to-wet soil moisture gradient. Also indicated are three sample clusters resulting from composite clustering, a nonhierarchical clustering technique. (Data tabulated in Table 1.2; also see Table 1.4)

German Meadow Samples

Ordination by Principal Coordinate Analysis

First two vectors account for 21.7 % of the variation.



Bibliography

This bibliography is compiled from the Statistics Department Library at Rothamsted Experimental Station. Some books may be out of print, or more recent editions may be available. The list is selective, and the level of difficulty is variable.

The series in Statistical Ecology published by International Co-operative Publishing House, 1971-9, is a set of conference proceedings, containing several papers of particular relevance to the present study.

General Statistical Texts

1. Goulden, C. (1952 and later edns)
Methods of statistical analysis
Wiley

2. Kotz, S. and Johnson, N. (1983-1988)
Encyclopedia of Statistical Science (12 volumes)
Wiley

3. Snedecor, G.W. and Cochran, W.G. (1952 and later edns)
Statistical Methods
Iowa State

Sampling Theory

4. Cochran, W.G. (1962)
Sampling Techniques
Wiley

5. Barnett, V. (1991)
Sample Survey, Principles and Methods
Edward Arnold

6. Cormack, R.C., Patil, G.P. and Robson, D.S. Eds (1979)
Sampling biological populations
ICPH Statistical Ecology Series

7. Patil, G.P., Pielou, E.C. and Waters, W.E. Eds (1971)
Sampling and modelling biological populations
ICPH Statistical Ecology Series

Frequency Distributions

8. Johnson, N.L and Kotz, S. (1972)
Distributions in Statistics (4 volumes)
Wiley

9. Ord, J.K, Patil, G.P. and Taillie, C (1979)
Statistical Distributions in Ecological Work
ICPH Statistical Ecology Series

10. Patil, G.P., Pielou, E.C. and Waters, W.E. (1971)
Spatial patterns, statistical distributions
ICPH Statistical Ecology Series

Regression and Curve Fitting

11. Draper, N.R. and Smith, H. (1966 and later edns)
Applied regression analysis
Wiley

12. McCullagh, P. and Nelder, J.A. (1982, 1990)
Generalized Linear Models
Chapmna and Hall

13. Ratkowsky, D.A. and Owen, D.B. (1990)
Handbook of nonlinear regression models
Dekker

14. Ross, G.J.S. (1991)
Nonlinear estimation
Springer

15. Sprent, P. (1969)
Models in regression and related topics
Methuen

Population Dynamics

16. Bartlett, M.S. (1960)
Stochastic Population Models
Methuen

17. Bartlett, M.S. and Hiorns, R.W. (1973)
Mathematical Theory of dynamics of biological populations
Academic Press

18. Hassell, M.P., and May, R.M. (1990)
Population regulation and dynamics
Royal Society of London

19. Nisbet, R.M. and Gurney, W.S.C. (1982)
Modelling fluctuating populations
Wiley

20. Patil, G.P., Pielou, E.C. and Waters, W.E. (1971)
Many species populations, ecosystems and systems analysis
ICPH Statistical Ecology Series

Ecological Diversity

21. Grassle, J.F., Patil G.P., Smith, W.K. and Taillie,C. (1979)
Ecological diversity in theory and practice
ICPH Statistical Ecology Series

22. Hastings, A. and Levin, S. (1986)
Community ecology
Springer

23. Pielou, E.C. (1977)
Ecological Diversity
Wiley

24. Williams, C.B. (1964)
Patterns in the balance of nature
Academic press

Theoretical Ecology

25. Jeffries, C, Lucas, W.F. and Thompson, M. (1988)
Mathematical modeling in ecology
Birkhauser

26. Legendre, L. and Legendre, P. (1983)
Numerical ecology; developments in environmental modelling
Elsevier

27. May, R.M. (1976)
Theoretical Ecology
Blackwell

28. Pielou, E.C. (1977)
Introduction to mathematical ecology
Wiley

Multivariate analysis

29. Barnett, V. (1982)
Interpreting multivariate data
Wiley

30. Chatfield, C. and Collins, A.J. (1980)
Introduction to multivariate analysis
Chapman and Hall

31. Hand, D.J. (1981)
Discrimination and classification
Wiley

32. Krzanowski, W.J. and Copas, J.B. (1988)
Principles of multivariate analysis: a user's perspective
Clarendon

Cluster Analysis

33. Gordon, A.D. (1982)
Classification
Chapman and Hall

34. Sokal, R.R. and Sneath, P.H.A. (1963)
Principles of Numerical Taxonomy
Freeman

35. Van Ryzin, J. (1977)
Classification and Clustering
Academic Press

Ordination methods

36. Everitt, B.S. (1978)
Graphical techniques for multivariate analysis
Heinemann

37. Greenacre, M.J. (1984)
Theory and application of correspondence analysis
Academic Press

38. Van Rijckevorsel, J.L.A. and De Leeuw,J. (1988)
Component and correspondence analysis
Wiley

Ecological Classification

39. Digby, P.G.N. and Kempton, R.A. (1987)
Multivariate analysis of ecological communities
Chapman and Hall

40. Gauch, H.G. (1982)
Multivariate analysis in community ecology
Cambridge University Press

41. Orloci, L., Rao, C.R. and Stiteler, W.M. (1979)
Multivariate methods in ecological work
ICPH Statistical Ecology Series

42. Seal, H.L. (1964)
Multivariate statistical analysis for biologists
Methuen

Spatial Analysis

43. Cormack, R.M. and Ord, J.K. (1979)
Spatial and temporal analysis in biology
ICPH Statistical Ecology Series

44. Webster, R. and Oliver, M.A. (1990)
Statistical methods in soil and land resource survey.
Oxford University Press

Time Series

45. Box, G.E.P. and Jenkins, G.M. (1970)
Time Series Analysis
Holden-Day

46. Chatfield, C. (1980)
Analysis of Time Series: an introduction
Chapman and Hall

47. Diggle, P.J. (1990)
Time Series: a biostatistical introduction
Clarendon

Repeated Measures

48. Crowder, M.J. and Hand, D.J. (1990)
Analysis of repeated measures
Chapman and Hall

49. Hand, D.J. and Taylor, C.C. (1987)
Multivariate analysis of variance and repeated measures
Chapman and Hall

Non parametric methods

50. Siegel, S. (1956)
Nonparametric statistics for the behavioural sciences
McGraw-Hill

Statistical Computing

51. GLIM
Reference Manual
NAG Ltd, Oxford

52. Healy, M.J.R. (1988)
GLIM: an introduction
Oxford

53. Lane, P.W., Galwey, N, and Alvey, N.G. (1987)
GENSTAT 5, an introduction
Oxford

54. Payne, R.W. and others (1987)
Genstat 5 reference manual
Oxford University Press

55. Ross, G.J.S. (1987)
Maximum Likelihood Program, MLP
NAG Ltd, Oxford

56. Ross, G.J.S. (1980)
CLASP User's Guide
Rothamsted Experimental Station

TABLE 1. WYTHAM TYPE OF DATA. Presence/absence for

	Plot	452074		452079		452081		452089		453076		----> 164 plots altogether
	Year	1974	1991	1974	1991	1974	1991	1974	1991	1974	1991	
Angelica sylvestris		✓	✓									
Brachypodium sylvaticum		✓	✓	✓	✓	✓	✓		✓	✓	✓	
Circaea lutetiana		✓	✓	✓	✓	✓	✓	✓		✓		
Deschampsia cespitosa		✓	✓	✓	✓		✓				✓	
Endymion non-scriptus		✓	✓			✓	✓	✓	✓		✓	
Galium aparine		✓	✓				✓	✓		✓		
Geum urbanum		✓				✓			✓		✓	
Glechoma hederacea		✓	✓	✓	✓	✓	✓			✓		
Mercurialis perennis		✓	✓	✓		✓	✓	✓	✓	✓	✓	
Poa Trivialis		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Rubus fruticosus		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Urtica dioica		✓	✓			✓	✓	✓	✓			
Viola riviniana		✓				✓	✓					
Arum maculatum			✓		✓				✓		✓	
Hypericum hirsutum			✓					✓	✓			
Rumex sanguineus			✓									
Dryopteris filix mas			✓	✓			✓			✓		
Carex sylvatica			✓	✓	✓							
Filipendula ulmaria			✓									
Scrophularia nodosa			✓									
Taraxacum officinale			✓									
Ajuga reptans				✓								
Bromus erectus				✓								
Carex acutiformis				✓								
Carex pendula				✓	✓	✓						
Galeobdolon luteum				✓		✓	✓				✓	
Rosa arvensis/spp				✓	✓	✓		✓	✓	✓	✓	
Tamus communis				✓						✓	✓	
Juncus effusus							✓					
Ribes sylvestris							✓					
Chamaenerion angustifolium								✓				
Clematis vitalba								✓	✓			
Epilobium montanum								✓	✓			
" tetra.								✓				
Humulus lupulus								✓				
Poa pratensis								✓				
Pteridium aquilinum								✓	✓			
Rubus caesius								✓				
Solanum dulcamara								✓				
~~Arum maculatum~~									✓			
Arctium minus										✓		
Dryopteris dilatata										✓		
Sonchus oleraceus											✓	

TABLE 2. Comparing overall species frequency (no of plots % in which species occurred)

Species	Total no of plots	1974	1991
		159	159
Arum maculatum		84	120
Desch. cesp.		70	100
Brach. sylv.		66	88
.		.	.
.		.	.
.		.	.
Rubus fruticosus		150	150
Endymion non.		96	100
Poa trivialis		86	80
Epipactis helleborine		6	8
Daphne laureola		4	0
.		.	.
.		.	.
.		.	.
Glechoma hederacea		80	60
Geum urbanum		75	50
Circaea lutetiana		100	60

↑ species showing increased frequency in 1991

species showing little relative 'change'

↓ species showing decreased frequency in 1991

Possible questions

? What are these common characteristics?