

Hierarchical clusters of vegetation types

C. S. Wallace¹ and M. B. Dale²

¹ *Department of Computer Science and Software Engineering, Monash University, Clayton, Victoria, Australia.
Email: Chris.Wallace@infotech.monash.edu.au*

² *Australian School of Environmental Studies, Griffith University, Nathan, Qld 4111, Australia.
Corresponding Author. Phone: +61 7 33714414, email: m.dale@mailbox.gu.edu.au*

Keywords: Cluster, Comprehensibility, Hierarchy, Minimum Message Length, Modifiable unit area problem.

Abstract: In this paper, we examine possible sources of hierarchical (nested) structure in vegetation data. We then use the Minimum Message length principle to provide a rational means of comparing hierarchical and non-hierarchical clustering. The results indicate that, with the data used, a hierarchical solution was not as efficient as a nonhierarchical one. However, the hierarchical solution seems to provide a more comprehensible solution, separating first isolated types, probably caused from unusual contingent events, then subdividing the more diverse areas before finally subdividing the less diverse. By presenting this in 3 stages, the complexity of the non-hierarchical result is avoided. The result also suggests that a hierarchical analysis may be useful in determining 'homogeneous' areas.

Abbreviations: MML - Minimum Message Length; MUAP - Modifiable unit area problem.

Introduction

Although other methods of organising data have been used¹, unsupervised clustering has been widely employed in analysing vegetation data. Most such analyses have used hierarchical clustering methods; for example, the widespread Braun-Blanquet method (Westhoff and van der Maarel 1978) is formidably hierarchical in its approach. Whatever the *a priori* likelihood that vegetation falls neatly into the nested clusters demanded by such a model, it is surely more appropriate to test if a hierarchy *does* provide a better model of the data than alternatives, such as Galois lattices (Rodin et al. 1998) or Gaussian response curves (ter Braak and Prentice 1988). This, of course, requires that we have some means of measuring the quality of a model. The Minimum Message Length (MML) principle (Wallace and Dowe 2000) provides just such a measure; the shorter the message length, the higher the prior probability of the model.

In a previous paper, Dale (2002) examined the use of MML criteria for estimating the number of clusters re-

quired in unsupervised clustering of vegetation data. The clustering program used was Wallace and Dowe's (2000) SNOB. This provides a general regionalisation² using a nonhierarchical clustering. A variant of the program incorporates possible spatial correlations (Wallace 1998) and thereby encourages spatial contiguity of cluster members but this was not used. The program does not provide a segmentation (cf. Oliver et al. 1998) with crisp boundaries for clusters. Instead it employs a fuzzy assignment of things to clusters; such fuzziness is necessary to obtain consistent estimates of cluster parameters.

Boulton and Wallace (1973a,b) presented a method (HSNOB) using MML estimation for hierarchical clustering. This means that we can actually make a comparison of non-hierarchical and hierarchical analyses based on the message lengths. In this paper we propose to first examine the concept of hierarchy, and then consider possible reasons why vegetation might have such structure. We shall then examine the application of HSNOB and compare the hierarchical and non-hierarchical solutions to determine

1 See, for example, Austen (2002) and Dansereau et al. (1966) for alternatives.

2 From a strictly geographic point of view, regionalisation may be either special or general (Abler et al. 1971). Each region of a special regionalisation is geographically contiguous and considered unique. A general regionalisation has 'region-types', each of which may consist of several spatially disjunct areas of the same type. Wallace (1998) proposed a variant of SNOB that incorporates possible spatial correlation between plots. this also appears in the text!!

which provides a preferable model of the variation we have observed.

Hierarchies and ecology

Ecologically hierarchies have attracted much attention, with studies like those of Allen and Starr (1982) and Ahl and Allen (1996), but they have been most employed in providing classifications of vegetation, often for bureaucratic purposes. As an **example** of one such classification we use an **example** from the California Vegetation Classification (California Department of Fish and Game, 2003). A single vegetation community is described by a code such as [6 1 . 3 1 1 . 0 2]. Here there are 4 levels in the hierarchy and the key to the code is presented in Table 1. The reason for this example is to illustrate that, in practice, ecological hierarchies may be a composite of habitat, physiognomic and floristic attributes. In contrast, our study will use only floristic data.

If we cluster both species and sites we can arrange our data as a two-way table. The intersections of the 2 cluster sets can be used in a nodal analysis (Lambert and Williams 1962), and this makes any hierarchical structure much easier to identify visually. A node represents a block where a subset of species is represented strongly in a subset of sites. Quantitatively the definition can be more subtle, depending on the permissible form of variation within clusters (cf. Dale and Anderson 1973). An idealised nested structure is presented in Table 2

But do real data display such patterns? In Table 3, we show structures abstracted from Doherty and Coops (1995) concerning *Eucalyptus* species in South-eastern New South Wales. These were sorted using a numerical agglomerative clustering and a Bray-Curtis similarity measure. The concentration of entries in each nodal call was visually assessed. The results from qualitative and quantitative data differ, and neither shows any marked nested structure. Both do show a general gradient pattern although in neither case is a single gradient sufficient.

A more complex situation is shown in Table 4, which is adapted from Dale and Quadraccia (1973) after Lang (1970). The organisation here was obtained by manual sorting and encompasses all species, not just the trees. While there is certainly some nesting, the patterns of nodes are clearly quite complex, with a reticulate pattern more likely than any binary tree structure.

The final example is taken from Webb et al. (1967) and records the vegetation in 10 plots for 12 years (Table 5). As with all the other examples, it is possible to obtain a hierarchy from the relationships between the groups. However, it is not obvious that the nesting could be easily

Table 1. Explanation of Californian vegetation coding.

Denotation	Coding
general physiognomic and physical location (riparian and bottomland habitat)	6
general habitat (riparian forest and woodland)	1
floristic vegetation alliance (Sycamore alliance)	311
Association California Sycamore/Soft Chess)	02

Table 2. Hierarchy of nodes: idealised pattern. o indicates strong concentration, ... indicates weak concentration.

Clusters	Sites				
Species		A	B	C	D
Ubiquitous	•	•	•	•	•
A	•	•	•		
B		•			
C			•		
D				•	•
E				•	
F					•

Table 3. South-east Australian forest - tree species.

a. Qualitative analysis

Clusters	Sites				
Sites	B	A	C	D	E
c			•		...
a			•		
d		...	•	•	
b		•	•		
f	•	•	...		
e	•	...			

b. Quantitative analysis

Clusters	Species				
	C	A	B	E	D
f				...	•
a			•		...
e	•	•	...
c	...	•	•	•	...
d		•	•	•	...
b	•		•	•	...

described monothetically, and in several cases a common substructure would be repeated, complicating the tree structure.

These examples show us that, while a hierarchy can be used to organise data, it will often suffer from defects as a representation of the nodal pattern. One particular problem is that of duplication where a similar subtree recurs at several places in the tree. In supervised clustering this has been addressed by introducing decision graphs, but such a reticulate pattern has not been commonly sought in unsupervised clustering. Similarly, insistence on a hierarchical structure could result in fragmentation of the data into numerous small clusters. It seems that vegetation **hovers tantalising** between hierarchy and other structure, so that a means of determining which is the 'better' model is highly desirable. It is this task that the MML principle permits us to accomplish.

Hierarchy and level

Mathematically, a hierarchy is a partially ordered set, a collection of parts with ordered asymmetric relation-

Table 4. Brindabella forest vegetation - all species. o indicates strong concentration; ... indicates weak concentration.

Clusters	Sites											
	A	B	C	D	E	F	G	H	K	J	L	
Species	a	•	•	•	•	•	•	•	•	•		
	b	•		•		•	...					
	c	•	...		•		•					
	d	•										
	e		•	•					
	f		•									...
	g				•	•		...	•			
	h			•	•		•	•	•			
	j			...			•	•	•			
	k						•	•				
	l									•	•	
	m	•	•
	n									•		
	o

Table 5. Nodal cells, Mt Glorious data, Webb et al. (1975).

Cluster	Sites											
	A	B	C	D	E	F	G	H	J	K	L	
Species	a	•	•					
	b	•	•	•	•			...	•
	c			•	•	•
	d	...		•								
	e								
	f					•		•				
	g					•						
	h								...			
	j										•	
	k						•	...				

ships inside a whole. That is to say, upper levels are above lower levels, and the relationship upwards is asymmetric with the relationships downwards. Bunge (1969) notes that hierarchies are usually employed where there is some concept of power relationships between nodes at different levels. Such concepts are not necessarily relevant to vegetation where a better description might be what Bunge (1969; cf. Dale 1976) calls a 'level structure'. This implies a nested structure but without any implication that higher levels have any 'power' with respect to subordinate groups. It is doubtful that the nomenclature will cause difficulties and we shall use hierarchy throughout.

Most numerical methods for clustering, both agglomerative or divisive, have used binary trees, whereas the human products very often the 'arity' is much greater. One method that does find higher 'arity' branches is Goodall's probabilistic method (see Clifford and Goodall 1967; Goodall and Feoli 1988) although paradoxically it does so by separating sub-clusters one at a time. Visual methods have not been thus restricted; cf. the Braun-Blanquet

method described in Westhoff and van der Maarel (1978). The HSNOB program is not restricted to binary splits. Methodologically, any tree can be simulated by a binary tree but the higher arity splits can be useful for comprehension.

There are several criteria whereby some levels reside above lower levels. These criteria often run in parallel, but sometimes only one or a few of them apply. Upper levels are above lower levels by virtue of:

- being the context of lower levels (contextual);
- offering constraint to lower levels (power);
- behaving more slowly at a lower frequency than lower levels (dynamic);
- being populated by entities with greater integrity and higher bond strength than lower levels (organisational);

- containing and being made of lower levels (anatomical).

We need to examine mechanisms producing pattern in vegetation that might also demonstrate one or other of these level-differentiating properties. But we cannot ignore the possibility of interactions between processes also contributing to the structuring. That the interactions can be complex is illustrated by the work of Herben et al. (1993). They found that, although there was considerable change locally, such changes did not result in corresponding variation over larger areas. Hogeweg (2002) comments that “processes do not, in biotic systems, operate in isolation and the existence of entanglement at different time and space scales does not need explanation, being there by default. Ignoring it by segregating time and space scales is simply a modelling artefact”. Entanglement in time and space means that any causal relationships may be difficult to separate, nor need any single process lead to hierarchical structure. But such structure may be a reasonable initial starting point for examining such entanglement. So we must look at processes forming patterns in vegetation and determine which, if any, might individually or in combination produce hierarchical structure.

Concepts and comprehension

Analyses are performed to increase human comprehension and to this end hierarchies are a common means of dealing with complex structure. Whether vegetation is regarded as a continuum or has some form of discrete structure, we do not expect all species to be equally common and organising them into blocks of approximately the same abundance or diversity may well provide a convenience for human comprehension. This is independent of any properties of the data themselves. In principle, it is possible to determine whether a continuum model or a cluster model is to be preferred; indeed, this is a question we are presently examining. However, Dale and Anderson (1973) suggest a two-parameter analysis combining elements of both continuous and discrete variation, which should also be considered and there are other models.

It may still remain true that the less preferred is easier to understand! Pattern is, after all, ‘pattern is for an agent’ (MacKay 1969) and Kodratoff’s (1986) principle of explicability should apply; i.e., the user should be able to interpret the results easily even if they are less efficient. This principle is really the basis for the ‘logical AI school’ as opposed to the statistical school. Statisticians seek efficiency above explication.

It is difficult to suggest measures that would assess such subjective properties numerically. Such measures

would have to include a concept of the fruitfulness of some particular choice over some period into the future as well as accommodating the value of predictions made (and tested?).

Modifiable unit area problem

In the study of vegetation there is a major problem: the area sampled to identify each plot or ‘thing’ is modifiable. There is no obvious means of delimiting an ‘individual’ thing and the investigator has to simply decide on some specific size and shape. This decision seems to depend on the sizes of the individual plants and on the recognition of mosaic areas. In addition, there may be conflict suggesting that more than one size is needed. As a result any plot can contain elements of more than one vegetation type. (Dale 1988). Fuzzy assignment can obviously cope with this by partial assignment but this is not available to the hierarchical analysis at the moment.

There are several methods that seek to determine some optimal size (and shape) but most of these assume **regularity of pattern** and this is unlikely (Kershaw 1970; Goodall 1974). Methods avoiding the use of areal samples have been developed seeking to avoid this difficulty, but in most cases they have other problems (Williams et al. 1969, De Jong et al. 1980). A spatial segmentation would probably be a useful means of studying this; e.g., a two-dimensional segmentation possibly based on Markov random fields.

Even when a decision has been taken about the nature of the individual plot to be described, the investigator still has to specify the pattern of distribution in space. Is this to be random, stratified, regular or specially selected by the user. The choice will also interact, in Bayesian methodology, with the prior probabilities assigned to models. These will possibly become context sensitive. Another possibility is to identify ‘objects’ spatially which need not have uniform size and shape, and use these as sampling units. Ciazlo and Ferer (2000) provide an example from image analysis. However, some limited attempts (by MBD) to identify ‘objects’ in vegetation data have not proved very successful, largely because closed areas were not obtained. But what is critical is that *all* structures identified in vegetation have to be viewed against the ambiguities of scale interacting with sample size, shape and distribution (Huston 1999). Hierarchy is one way of allowing this.

Hierarchical structure in vegetation

In order to create a hierarchic structure, some species must be relatively widespread, while others must be re-

Table 6. Potential sources of hierarchical organisation in vegetation.

Extrinsic		Intrinsic
Abiotic	Plant-based	Measurement-based
Environmental	Morphodynamic Patches	Idiotaxonomic
Contingent Events	Behavioural and Dispersal Patches	Scale of Performance Measure
Disturbance and Gap Phase Patches	Non-equilibrium Status: succession	

stricted within the larger ranges. Such a situation can be produced in many ways and can also apply recursively to provide several levels in the hierarchy. We shall use the major categories shown in Table 6.

Environmental

This is probably the most obvious source of pattern in vegetation, associated with variables such as soil type, water availability and radiation impact. One major problem is that effects are often dynamic, so the active effect may be the probability (intensity and recurrence interval) of some extreme events; eg. ice storms, frost, flooding, drought, cyclones and so on. If these events occur too frequently then the plants may not succeed in reproducing, and thereby maintaining the population, although they may be able to survive as adults and in especially good times invade. In some cases recurrent events such as fire are a necessary component of regeneration, as in some *Eucalypt* forests or *Banksia* heaths. This cannot then be regarded as a contingent event though it is unpredictable. The ‘good times’ may of course be an historical ‘golden age’ which climate change has banished.

In some cases the environment may lead directly to hierarchic structure, as for example in the Gilgai structures of arid Australia or of the polygons resulting from cryoturbation. Many environmental factors can be regarded as continuously varying, although there may be critical values where marked changes occur; for example the effects on aluminium and manganese ionisation in soil around pH 4.5 or freezing point for temperature. Lomolino (1996) Wright et al. (1998) and Huston (1999), among others, have examined nested properties of diversity gradients, most notably the latitudinal gradient, where temperature and radiation properties are varying. Other studies, for example Brown (2001), have examined altitude gradients. Wright and Reeves (1992) describe various measures of ‘nestedness’ but these seem mostly

to assume something like Dale and Anderson’s (1973) two-parameter model and not a hierarchy³.

More generally a hierarchical structure will form if different species have differential *and* nested responses. So, niche differentiation and competitive or facilitative effects can lead to hierarchy. Another possible source is that the generating processes produce self-similar fractals.

Contingent variation

Variation in vegetation has a large noise component because the success of any single propagule is often related to quite specific, even unique, and unpredictable events. Such isolated, unusual events are present in almost every vegetation cover. So deposition of animal dung, local disturbance by animal burrowing or even damage by passing animals all provide opportunities or emergencies for one species or another – assuming the propagules are able to reach the place. This means that, while most of an area may be covered by one general type, a few unusual places are always to be expected. If such events become too frequent then they can be better regarded as an environmental effect although the chance component remains large. The most obvious example of this would be wind-throw in forests, which leads to gap regeneration. In some cases the isolated plots will represent particular specialised environments, such as outcrops of ultra-mafic rocks. In such a case detailed study would reveal this, but with much contingent variation it may be impossible to determine the original cause, it being lost in unrecorded history.

Patches: disturbance and imperturbability

We shall distinguish 3 different kinds of patch-forming process – gap phase, morphological and behavioural. All can lead to hierarchical structure via Maslof’s (2001) observations of limited favourability of patches.

Patches are of course related to scale. And, as Bar Yam (2002) says of hierarchical clustering, ‘Clustering is not a complete description of the process of complexity/information loss (entropy increase) implied by the formalism, but the conditions at which clustering occurs are the characteristic scales at which degrees of freedom are lost to the observer’. By examining the rate of loss it may be possible to simplify the tree structure. Notice that this view of the value of hierarchies is not associated with corresponding structure in the attributes. The trade-off be-

3 Paterson and Atmar (1986) and Wright and Reeve (1996) can serve as examples of ‘nestedness’ measures. The former identify, for each species, the least diverse sample in which it occurs and then count the number of absences in more diverse samples. This value is summed over all species. The latter identify the same critical sample but count the number of absences in less diverse samples.

tween scales of behaviour characterizes the relationship between dynamic and coherent structures at different scales in non-equilibrium systems (Bar Yam *loc cit*; cf. Herben et al. 1993)

A casual visual inspection will show that most vegetation exhibits patchiness at various local scales. Several possible causes exist for this but, although environmental differences may exist between patches, these could well be the result, rather than the cause, of the patches. Attribution of all patterns to environmental causes is unwarranted, if only because it is a universal model which cannot be tested. In any case, if there exists sociologically induced pattern then the result over time is likely to be the development of environmental differences. The existence of a difference is *not* enough to demonstrate its causal nature.

Spatially, in a uniform environment, patch structure should tend to have a regular pattern but this may not be clear in a practice where environmental changes also occur for in that case the sizes of the patches will vary, possibly over a large range. In fact, regularity of patch distribution has only been identified in a few cases, mostly where regularity in environmental conditions can also be observed.

Gap phase dynamics: carousel

Another form of dynamic pattern, similar to contingent variation, is the carousel process discussed by Van der Maarel and Sykes (1993). It is marked by the greater random component and the lack of interaction between species. The carousel model suggests that there exists space that can be occupied by any of several species for short periods; the species complement is itself in flux. Glenn and Collins (1993) argue that the gap phase constituents are a random selection. Over a longer period, it is likely that all of the species may appear temporarily in combination. Herben et al. (1993) suggest that pattern at local spatial scales is not a major cause of pattern at larger spatial scale. This would be counter to any hierarchical organization. Vandvik (2004) finds dependence on gaps increases during succession and that a large proportion of the species are gap associated (74%). Poulin and Guégan (2000) report a spatial pattern for ectoparasites which might then be nested because of host distribution.

Morphological patterns: sociability

This is pattern based on morphological properties of individuals. Many plants use vegetative reproduction that involves occupying patches of ground, leaving the interstices for other species. This effect can be negative where

other plants are excluded but can also be positive, as with the ‘nurse plant’ effect where seedlings develop in the shelter of other individuals or hepatics in a *Sphagnum* mat.

Sociability or gregariousness is an expression of horizontal pattern of species. It measures the degree of clustering (contagion) of the species. It can be directly recorded as an ordered scale from observations within a single sample plot, provided this is of sufficient extent. Thus:

- Growing solitary, singly;
- Growing in small groups of a few individuals, or in small tussocks;
- Growing in small patches, cushions or large tussocks;
- Growing in extensive patches, carpets, or broken mats;
- Growing in great crowds or extensive mats completely covering the whole plot

(from Westhoff and van der Maarel 1978).

Because of MUAP, such spatial patterning could well extend beyond a single plot and it could lead to a hierarchical pattern being generated if several such single species patterns are conjoined to form a mosaic, which is at least partially repetitive.

Behavioural patches

Extending beyond morphological patterns, the dispersal of propagules may induce local patterns maintained dynamically (cf. Dale and Hogeweg 1998; Dale 1999). This results in a mosaic of slightly different patches of varying shapes and sizes that exist for some period and then are modified or eliminated. There is no necessity for environmental variation to pre-exist for such patches to form. Rather, the existence of the patches would lead to environmental differences and evolutionary consequences. Such patterns have been discussed both by Boerlijst and Hogeweg (1991, 1995a,b) and by Rietkerk et al. (2002). They differ from morphological patterns in that the patches have an essentially fixed size through time, rather than expanding with age. Boerlijst-Hogeweg patterns require relatively strong interactions. Sociological patterns tend to have at least some components with fixed scales; indeed these are generated by the process.

Thus, hummocks in bogs might well be a result of dispersal patterning, as could many observed ‘ring’ structures commonly regarded as being due to expansion with

age. Boerlijst and Hogeweg's work would suggest that the 'rings' are really double spirals and that the whole patch might rotate, albeit slowly. The size of the 'rings' depends on interaction intensity and not on time. Hummocks we have examined do seem to have a small gap in the perimeter but long term studies to identify rotation do not seem to have been made. A mosaic will develop from such patterns without environmental influence; environmental changes are likely to disturb any regularity.

The result of both morphological and dispersal processes will be the break-up of the vegetation cover into a series of patches of irregular size, making any choice of a single sampling area for plots hazardous. The plot may include part of a single patch or parts of several patches. Each patch will differentiate from the others because it has a different flora, so that environmental differences will appear. This can lead to evolutionary consequences (Pagie and Hogeweg 1997).

Non-equilibrium – migration, expansion and contraction

Another source of variation concerns the disparity between the actual state of the vegetation and the potential state; these are rarely in equilibrium. Such a non-equilibrium state may be directed ('al) or fluctuating (chaotic attractors) and differences between timing for various species might lead to hierarchical organization. Plants are generally immobile in the adult state, which means that any moderate change in environmental conditions need not result in immediate change in the vegetation. The advent of propitious conditions does not mean an immediate colonisation, since the plants must first transport themselves from some source overcoming any barriers along the way. Some tree species appear to have only recently managed to cross the Alps after extinction north of them during the last glaciation! Conversely, while intense adverse condition may destroy a population rapidly, slow deterioration will mean that resistant individuals may be present even when the population is no longer self-sustaining; viable populations will only exist in more favoured refugial areas. Adult plants are often much more robust than seedlings so that failure of reproduction does not mean immediate extinction of perennial species. Immigration and extinction, and their hierarchical potential, have been discussed by Lomolino (1996).

Non-equilibrium means that species may be increasing, stable or decreasing dynamically and a snapshot sample cannot distinguish such temporal properties. A mixture of species results that may not all be associated with the same vegetation type, resulting in fuzziness of assignment to types. In fact the non-hierarchical clustering

showed little fuzziness and this source of variation will be largely ignored for the rest of this paper.

One special case is that of adventitious species introduced through human activity. Since these tend to be generalists they may well have native species nested in communities. Unfortunately, the data we are using does not contain such species.

Idiotaxonomy

A possible source of hierarchical structure in vegetation lies in the taxonomic hierarchy. The nesting of taxonomic units means that structure associated with higher level categories such as super-orders, orders and families might have nested **within it structure related to genera and species**. Even if we have only species records, sub-specific units might still present some hierarchical structure. However, it is more likely that the clustering will be used to suggest such possibilities than that we shall have available suitable data to interpret any hierarchical structure in this way. Apomictic species are likely to produce patches of this kind but they will be difficult to distinguish from patches of morphological origin. A study by Dale and Clifford (1976) did not show any indication of hierarchic structure. Genera were effective one-to-one substitutes for species with no hierarchical implications. The relationship with subfamilies, families and orders was much less at this local scale and again was not clearly associated with any hierarchical structure. However, an analysis of rainforests in eastern Australia did produce a major disjunction associated with the occurrence of the epiphytic subfamily of figs (**Figs are family Moraceae Trevor to find subfamilies from Corner, reference from Specht's book**). The differences would seem to be associated with large areas. Taxonomic information was not made available to the program so we would not expect to discover such a hierarchy, even if it existed.

It might be possible to learn attribute hierarchies but this is not a feature of the present program. Xiao et al. (1993) and Hogeweg. (1998) discuss various aspects, but any such discovery procedure is beset by the problem that the hierarchy may not be unique. Thus soap is a subclass of toiletry. It is also a subclass of cleaners and of chemicals produced using sodium hydroxide; no unique hierarchy is necessarily present. This clearly adds problems because we need to learn appropriate structures. For plants we have taxonomic, life form, dispersal, age and functional relationships any of which might provide hierarchical structure for the primitive attributes.

Scale of measure

Table 7. Nesting of density clusters within presence: Mt Glorious data (see Webb et al. 1967).

Clusters		Density							
		3	4	5	10	11	12	13	14
Presence	5	0	0	0	10	17	0	0	0
	11	3	0	2	0	0	7	21	17
	12	8	17	11	0	3	0	0	4

There is in almost all vegetation a strong component of presence/absence variation. Even if we have abundance measures the measurement is conditional on presence. But, as Babad and Hoffer (1984) put it, even no data may have value. It is therefore possible that we can find structure based on the presence/absence component and nested within this further structure reflecting patterns in abundance. In fact, such nesting was observed in rain forest data by Dale (2000; see also Dale and Anand 2004) and is illustrated in Table 7. This could be directly investigated by partitioning the data into presence and abundance components (Williams and Dale 1962) and then determining if levels of hierarchical structure are associated with either measure. It is also possible that structure will be reflected in variations of diversity which will be dominated by presence effects initially. An alternative approach possible with the SNOB program is to first code presence/absence as a nominal attribute and then code the abundance values as numeric but with absences coded as missing values.

The abundance measures themselves can be further arranged in a series of increasing information; density, valence, cover, cover-abundance and biomass form such a sequence and all can be cast as probability measures. In principle the structure at each position in such a series could be nested in structure derived from preceding measures. In practice it is unusual to use more than one of these abundance measures available.

Methods

The Minimum Message Length clustering principle is described in Wallace and Dowe (2000). The principle is based on the notion that an optimal clustering solution will require a balance between the number of clusters and the fit of the model and is an implementation of Occam's razor⁴

A complex model will fit the data well, but is less likely to generalise to novel observations and will usually be less comprehensible. A simpler model will fit less well, but will be more comprehensible and hopefully more generally applicable. There is still considerable argument

concerning the exact role of simplicity in induction, which we shall here regard as a useful heuristic, at least.

Wallace and Dowe (loc cit) provided a *k*-means-like procedure implemented in SNOB, which both estimated the number of clusters and also provided a fuzzy assignment of items to clusters to obtain consistent cluster statistics. The choice of the number of clusters relies on a compromise between complexity of model, essentially its prior probability, and the adequacy of fit to the data, measured by the message length required to encode the data assuming the correctness of the model.

It is possible that the required message length can be reduced still further. Wallace and Dowe assume that attributes are uncorrelated within clusters and the introduction of within cluster correlation might reduce the message length. This is presently being studied. Agusta and Dowe (2003) have examined this possibility and it is the subject of on-going work. A similar reduction could occur by changing the acceptable distribution of variation within clusters, with a choice between Poisson, Gaussian, *t*-distribution and gamma distribution being provided within the present program. Dale (2001) has examined the effects of choosing Poisson or Gaussian distributions. A non-hierarchical clustering was obtained using 120 random starts of the SNOB program (Wallace and Dowe 2000). This program uses the minimum message length criterion to determine the number of groups and provides a probabilistic assignment of all things to all clusters. It also uses MML estimates for cluster parameters.

Another method of potentially reducing message length is to organise the clusters hierarchically. Hierarchical clustering is not uncommon in ecological studies but assumes a somewhat different aspect when combined with MML clustering. The idea here is that it may be possible to describe a subset of clusters partially at a higher level in the hierarchy, thus reducing the required message length for the individual terminal clusters. This is the approach adopted in this paper. There are some difficulties about how the search for clusters should be performed. With a divisive search strategy, things may be misallocated at high levels and will never be placed in their

4 'Entia non sunt multiplicanda praeter necessitatem' or, more likely, 'Pluralitas non est ponenda sine necessitate'.

proper final (low level) clusters. However, assignment of new things is relatively easy. Agglomerative approaches do not suffer from the misassignment problem, but they do not provide any means of making assignments.

The hierarchical clustering was obtained using a program, HSNOB, developed by Wallace from earlier proposals of Boulton and Wallace (1973a). Details are presented in Appendix 1. It is worth noting here that the program does not employ the same prior probabilities as the non-hierarchical program and this renders exact comparison of message lengths difficult. In addition the hierarchical analysis does not use fuzzy allocation to clusters, which will tend to increase the message length and also lead to inconsistent estimation of the cluster parameters.

Having recovered exemplars of the several clusters from the data, we need to convert these, by induction, into defining rule sets so that assignment of new sites is possible. Traditionally, emphasis has been on monothetically defined clusters, which capture classical concepts in a few, simple rules which are both necessary and sufficient. Both SNOB and HSNOB are polythetic procedures so the concepts they recover will generally be prototypic; that is they are defined nondeterministically by some number of a set of rules, individually neither necessary nor sufficient⁵.

Objectives and analyses

The first question to be addressed concerns the separation of these sources through the use of an hierarchical clustering procedure. Such a procedure may distinguish various levels of nesting in vegetation typology some of which hopefully will reflect these three sources. Whether the hierarchy is to be preferred to a non-hierarchical representation, where no nesting is assumed, is also of interest. The preference may be determined empirically but can also be affected by questions of comprehensibility.

One difficulty here is that both environmental and morpho-dynamic structure will show spatial aggregation, which is not a characteristic of the other sources. **Environmental patches are likely to be larger and probably some zonation.** Morpho-dynamic patches will generally be smaller and more irregular in size, but may also show regularity

The second question to be addressed here concerns the nature of the data used. Recently much attention has been paid to correspondence analysis, which implies an underlying chi-square metric whereas most clustering has retained an underlying Euclidean metric. Legendre and Gal-

agher (2001) proposed a transformation that allows the user to modify the data so that the underlying metric is a chi-square metric. While this is likely to emphasise rare species, it is of interest to examine the effects such a transformation might have on clustering results. The well-known TWINSpan method (Hill et al. 1975) also employs a chi-square metric but this method has no secure estimate of the number of clusters and due to the hierarchical nature of the process the actual metric changes with the subgroup being considered, which may be an advantage of course.

What TWINSpan does provide is an oligothetic approximation to the polythetic divisions. This leads to interesting questions concerning the nature of the concepts that we are seeking. The initial stage of class definition will usually employ an exemplar model; i.e., an explicit list of things which belong to the class. Classically we then seek to define a few simple rules which permit other things to be assigned to the class; typically these would be monothetic conjunctions. However, in many cases, we may actually need a prototypic definition and this is exactly the kind of definition that TWINSpan provides. (There are other ways of providing appropriate definitions.)

Data

The data used are the Mallee Data from Goodall's (1953) pioneering study on automated clustering. This comprises 256 stratified random plots ('things') described by 61 plant forms (mostly species but some life forms). Of these 61 only 32 are common, the remainder being present in only 1 or 2 plots. In each plot, the percentage cover of all species present was recorded using point quadrats for the estimation. The location of each plot is known so that the clusters can be mapped. This allows visual assessment of spatial contiguity. Formal methods for such assessment are possible, based on the method of Krishnalyer (1949) as extended by Critchlow (1985)

Goodall reported a dune-swale structure with some asymmetry across the area making the left and right sides dissimilar. The environmental reasons for this differentiation have not been explored but could be due to burning patterns, grazing

Results

General

5 For a more extended discussion of types of concept, see Laurence and Margolis (1999).

Table 8. General results.

Analysis	Initial MML	Number of Clusters	Cluster MML	Change in MML	% Structure Recovery
Non-hierarchical	45550.4	16	20652.2	24898.2	56.2
Hierarchical	45550.4	5 14 18	22093.4	23456.8	51.5

Table 9. Correspondence of the hierarchical and non-hierarchical clusters.

Level - Number Of Clusters	χ^2	Degrees of Freedom	Probability	R ²
Low - 18	372.6	272	<0.0001	0.26
Middle - 14	280.665	208	<0.0001	0.20
High - 5	61.959	80	0.4061	0.04

In Table 8, we show the overall characteristics of the MML analyses. While the non-hierarchical clustering provides 16 clusters, the hierarchical clustering can be interpreted at 5, 14 or 18 clusters; these will be discussed in more detail later. The immediate conclusion, though, is that the non-hierarchic result is to be preferred because of its shorter message length, resulting in greater capture of structure. The non-hierarchical structure is either less complex or better fitting (or both).

ining the assignment of sites to clusters in non-hierarchical and hierarchical results. From Table 9, it can be seen that the 5-cluster level does *not* show any marked correspondence with the non-hierarchical result but that the 14 and 18 cluster levels do show a reasonable levels of correspondence. (The χ^2 values are somewhat suspect as the expected values in some cells are very small.) So, the hierarchical analysis is apparently finding some structure that is not identified by the non-hierarchical. Whether this additional structure is valuable remains to be seen. The simultaneous recognition and superposition of the clus-

The two analyses can be directly compared by exam-

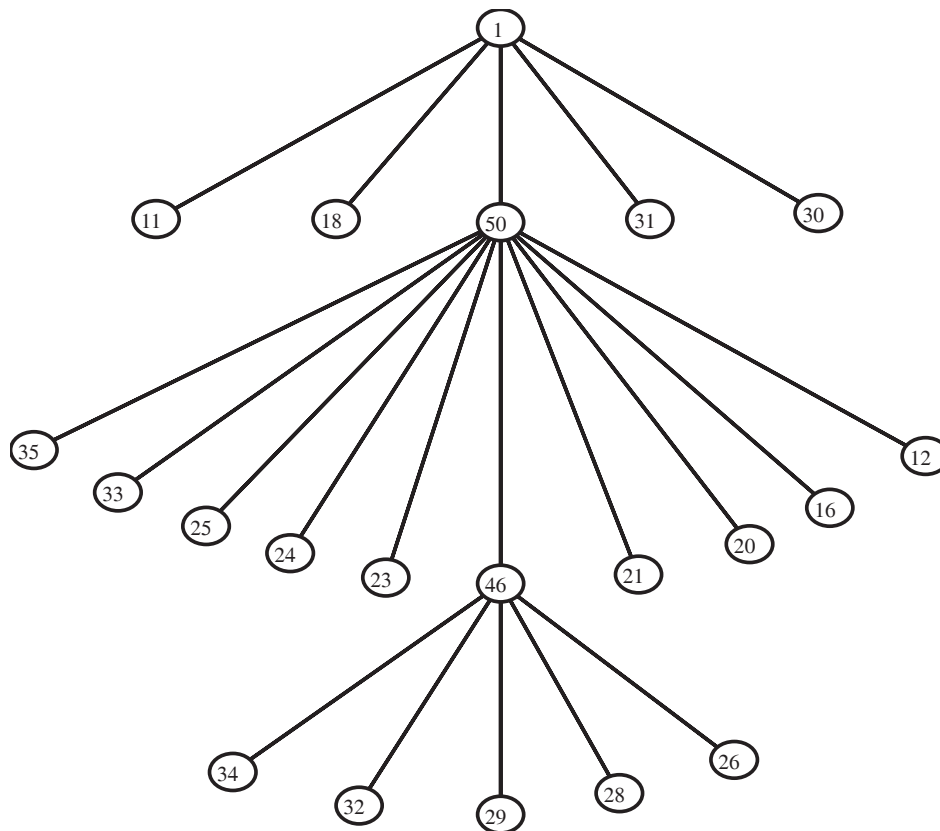


Figure 1. The recovered hierarchy.

Table 10. Size of clusters formed at each level. Cluster 1 is the root cluster comprising the population.

Group Label	Size	Parent	Group Label	Size	Parent	Level Label
11	8	1	18	6	1	
30	8	1	31	4	1	High
50	230	1				
12	8	50	16	4	50	
20	18	50	21	7	50	
23	12	50	24	63	50	Middle
25	9	50	33	20	50	
35	5	50	46	86	50	
26	5	46	28	48	46	
29	7	46	32	16	46	Low
34	10	46				

ters in space does complicate interpretation and some loss of efficiency to gain comprehensibility is perhaps desirable.

Non-hierarchical analysis

Results are presented as a map of the distribution of clusters (Fig. 1). Although the dune-swale structure and the asymmetry can be discerned, the interpretation is complicated by the superposition of the many patterns. The dune structure is accidentally emphasised by the choice of circles to indicate those clusters.

Hierarchical analysis

The hierarchical analysis resulted in a 3-level structure shown in fig 1. The initial division is into 5 clusters (the high level), one of which then splits into 10 clusters (the middle level) and one of these further splits into 5 sub-clusters (the low level). The sizes of these clusters are shown in Table 10, which reveals that many of them are quite small.

High level clusters. At the high level, there are 5 clusters, 4 of which are numerically small comprising in total 25 things only. The clusters are shown in Fig 2. I believe the small clusters represent aberrant samples, possibly due to contingent variation or perhaps very localised and disparate environmental conditions. The aberrant clusters are spread over the area with little concentration spatially for the small clusters. The bulk of the plots are assigned to a single cluster 50. It would be interesting to have a fuzzy assignment here to see if the 4 small clusters show any tendency to overlap with the large cluster.

Middle level clusters. At the middle level, there are 14 groups (4 from the high level and 10 from the middle) of which 9 are numerically small. Since the 4 remaining high level clusters have already been examined, only the sub-clusters of cluster 50 are shown in Fig. 3. There is an obvious linear structure associated with a single large sub-cluster (46) but little spatial aggregation for the others. These latter are found in the swales between the dunes but show a patchy structure. The swales are rather more diverse than the dunes but it is not clear that there is a strong environmental control of the other sub-clusters. They could well represent 'fuzzy' edges to the major clusters; ie. they could represent ecotonal types that exist where boundaries are gradual. Alternatively they may represent patch dynamics of some kind. However some of cluster 24 is sufficiently large to suggest environmental control and both cluster 20 and 33 could also be so defined.

One possibility is that there exists some kind of gradient leading to correlation between species. The analysis does not consider correlation within clusters, and this could result in the definition of extra clusters. A change in the model to incorporate within-cluster correlation would clarify this.

Low level clusters. At the low level there are 18 clusters (4 high, + 9 middle +5 low, with 11 of these numerically small). The blurring of the linear structure resembles that for the non-hierarchical result – it is rather too detailed for convenience.

In Figure 4, we show the sub-clusters of cluster 46 only, which resolves structure on the dunes. Some of the clusters certainly show spatial aggregation and it is again possible that there are local environmental differences underlying, or correlation between species. We could for-

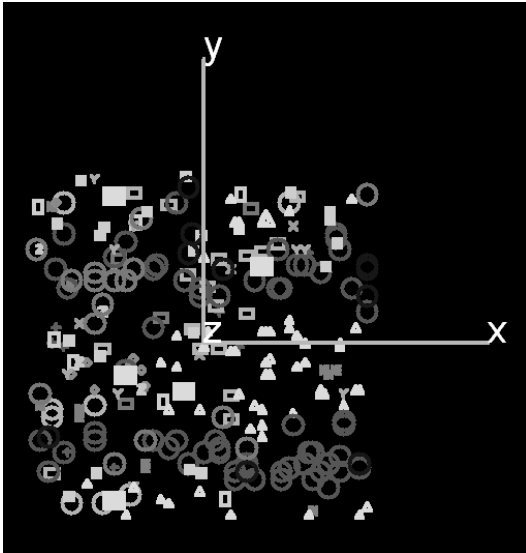


Figure 2. Spatial distribution of non-hierarchical clusters.

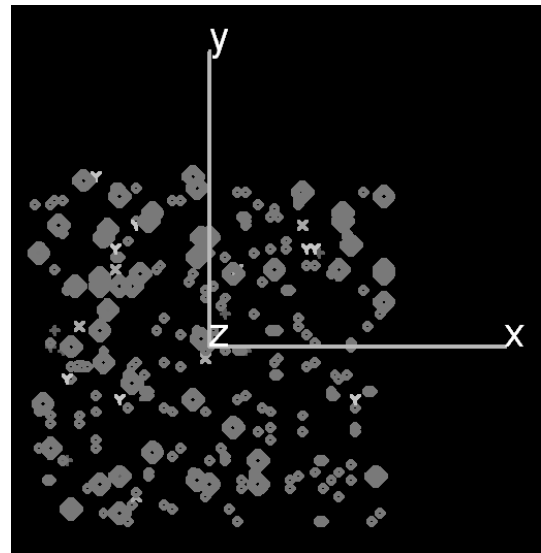


Figure 3. Spatial distribution of clusters - high level.

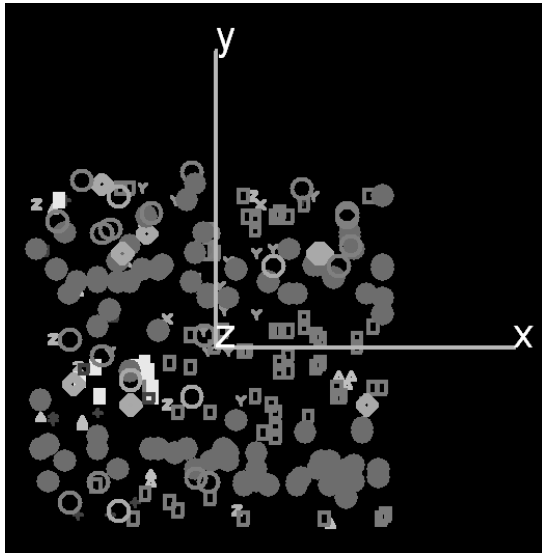


Figure 4. Sub-clusters of 50 - middle level.

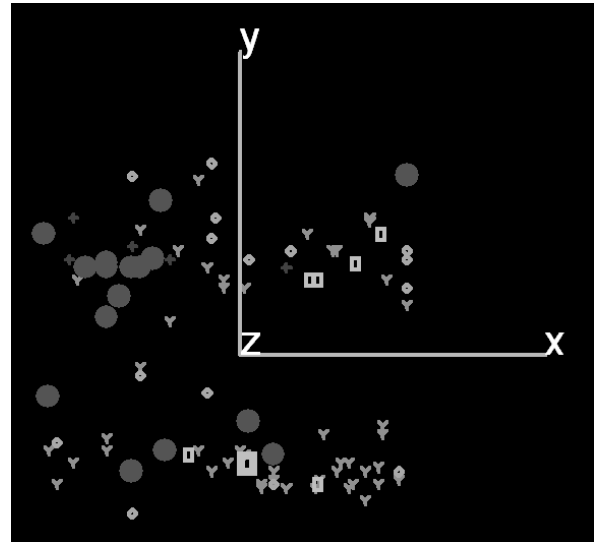


Figure 5. Sub-clusters of 46 - low level.

mally assess the spatial aggregation using Critchlow's (1985) extension of the Sandland and Young (1979) test or Krishna-Iyer's (1949) test. However, spatial aggregation is also a result of morpho-dynamic patch formation and is not a prima facie case for an environmental cause.

Comparison of clusters. Hierarchically, we find 21 classes (18 leaves) for a message length of 22093.4 with structure captured = 51.5%. The difference between non-hierarchical and hierarchical = 1441.2. As noted earlier, there is a difference caused by the use of different prior probabilities in the 2 analyses and also the use of fuzzy assignment in the non-hierarchical clustering. Whether these sources are sufficient to account for what is a large

difference is doubtful; the odds in favour of the shorter non-hierarchical result are approximately $e^{-1441.2}$. To remain faithful to the MML criterion, we would have to accept the non-hierarchical as the preferred clustering. However, there remains an outstanding question concerning the comprehensibility of the result, which we shall return to later.

Discriminating species

In Table 11, we show the species positively associated with the various clusters at all levels. In all cases there was a marked preponderance of negative associations and cluster 46 show several cases of species being entirely ab-

Table 11. Association of species and clusters at the several levels. # indicates no test possible as species is everywhere absent; o indicates no significant association with any group at that level; (•) indicates marginally significant association. ; + indicates species occurs significantly more than expected; - indicates species occurs significantly less than expected. (Not all absences are included).

Species	Cluster Level		
	High	Middle (cluster 50 only)	Low (cluster 46 only)
Acacia regens	+50	+12	#
Bassia parvifolia	•	+33	+26
Bassia uniflora	•	+46	•
Byeria opaca	+30	+23	#
Callitris verrucosa	(+11)	+21 (+12)	#
Cassia eremophila	•	+46	+32
prostrate Chenopodium pseudomicrophyllum	•	•	+26
erect Chenopodium pseudomicrophyllum	•	+20	-34
Danthonia semiannularis	+30	+20	•
Dodonea bursarifolia	+50	+25	+34
Dodonea stenozyga	•	+46	+34 -28
Enchylaena tomentosa	•	+33	+32
Eucalyptus calycogyna	•	+23 +33	•
Eucalyptus dumosa	•	+12	-28 (-26)
Eucalyptus oleosa	•	+20 +46 -24 -25	•
Grevillea huegelli	•	•	+32
Kochia pentaphylla	+11	•	+34 -28
Lepidospermum viscidum	+50	+12 +25 -46	#
Lomandra leucophala	+18	+16	#
Melaleuca pubescens	•	+23 -46	#
Melaleuca uncinata	•	+24 (+16) (+33)	#
Micromyrtus ciliatus	+50	+12 +21	#
Myoporum platycarpum	+31	+25	#
Olearia muelleri	+31	(+46)	+26 -28
Santalum murrayanum	•	+21	•
Stipa elegantissima	•	+35	•
Stipa mollis	•	+12 +20	•
Stipa variabilis	•	+20	•
Triodia irritans (live)	•	+21 +35 -46	#
Vittadina triloba	+18	+33	+29
Westringia rigida	+31	•	+32 -34
Zygophyllum apiculatum	(+11)	+46	+29 -28
Triodia irritans (dead)	•	+23 +24	#

sent from the cluster. Thus, at the high level cluster 50 is largely defined by the absence of species critical for the definition of the smaller clusters. There is no suggestion of groups characterised primarily by quantitative variation. Possibly a denser sampling network is needed to obtain such a result.

Discussion

Hierarchical analysis

The hierarchical analysis provides an interesting structuring of these data. The hierarchy does provide novel information and also presents it in a comprehensible format. Unfortunately, it seems that it is also less efficient since the non-hierarchical message length is signifi-

cantly smaller. It is possible, of course, that further search would recover an improved hierarchical solution. Also the different prior probabilities assigned in the 2 methods will also cause differences, although the message length associated with the prior probability is the smaller component of the total.

Using a fuzzy solution would reduce the message length and it is possible that, if such were available, the hierarchic solution would be preferred. However, the data do not give extremely fuzzy assignments. The best SNOB result, so far, estimates that there are 16 clusters with message length shorter than that of the hierarchical result. If we examine the fuzzy assignment, 15 things were assigned to 2 clusters and 5 of these had probabilities of 0.1 or higher of belonging to the second cluster; the worst

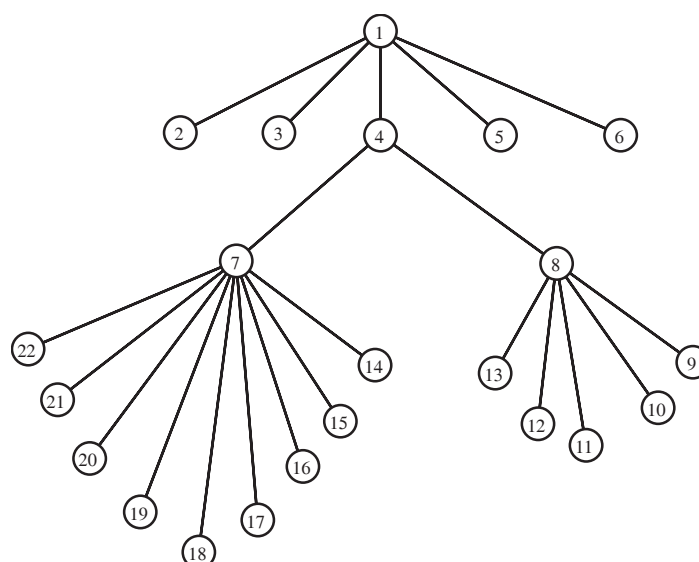


Figure 6. Alternative structure.

case had a probability of 0.33 of belonging to a second cluster. The remaining things were uniquely assigned. Thus only 0.02% of the things can be regarded as doubtfully assigned. Reduction in message length from this source would not be expected to be large.

Examined spatially, the non-hierarchical result suggests a linear arrangement across the area with some distinct types on the left side. This is in accord with the topographic features – a dune-swale system and a general trend across the area. However, the simultaneous presentation of 16 clusters tends to mask the horizontal linearity somewhat because all patterns were superposed. We think, because all 3 source of variation are being simultaneously addressed, some confusion is almost inevitable.

The hierarchical result is an interesting one because the 3 levels give an opportunity to look at the vegetation at 3 different scales. What it does not allow us to do is distinguish patchy environmental controls from morpho-dynamic or behavioural structure. Sub-clusters at the middle and low levels in some cases show spatial aggregation while in other cases they do not.

So the hierarchical result allows us first to identify (and eliminate if we wish) some contingent elements, then provides a useful general organization at the environmental level, before relapsing finally into a detailed presentation rather similar to the non-hierarchical solution in its complexity if not in the details of assignment. If we regard it as a progression through the available structure, from coarse to fine, it is rather helpful. It is possible that the diversities of the individual clusters at the middle level are similar to each other, and that the low level patterns,

internal to the dunes, reflect the larger size of that cluster, although the size variation is not extreme, as it is at the highest level. Certainly a hierarchy such as that in Figure 6, where cluster 46 is a sister group to the remainder of cluster 50 might have been expected, but the analysis suggests that this is not the case.

Clearly the results obtained do not allow us to separate the possible causal processes. While patches are present this alone is insufficient. The small high level clusters do seem to represent exceptions to the general vegetation of the area, but at middle and low levels morpho-dynamic, disturbance and behavioural processes probably interact with environmental differences.

Hierarchical relationships reflect a particular pattern of between-cluster variation. Within-cluster variation can be captured using a within-cluster ordination and more complex patterns of sharing of axes can be envisaged than that displayed by a hierarchy. Search procedures for such models have yet to be developed and will probably require objective rotation of the axes to simple structure (see Kiers 1994, 1998).

Scale of measure

This possibly does not fit well with the usual view of hierarchy in vegetation that associates different suites of species with each level. That is, we regard some species as having patterns over larger areas (smaller scales) than others. The present method of capturing hierarchical structure actually corresponds with this exactly, since it does not explicitly consider selection of subsets of fea-

tures apposite to hierarchical levels. Such subsets might, of course, appear as a *result* of the modelling.

The present method for acquiring hierarchic organisation does not address the possibility of a hierarchy based on the scale of measure of the data or the idiotaxonomic hierarchy. The species table (Table 10) suggests that presence/absence is the dominant form of variation and there is no suggestion of quantitatively defined groups within the clusters. We hope to address this in a future study.

Comprehension and the value of hierarchies

However, there is an argument to be made that the hierarchical model did provide a more immediately comprehensible picture than the non-hierarchical. The three levels do provide a comprehensible representation of three (different?) sources of variation in vegetation. While these may be present in the non-hierarchical model, they are more difficult to isolate and appreciate. The outstanding question is whether any gain in comprehensibility is worth the loss in effectiveness of the model. We need a considerable gain in comprehensibility to offset the difference in message lengths. But how can we measure comprehension?

Acknowledgements: Our thanks to Dr. David Goodall for permission to use his data.

References

- Abler, R., J. S. Adams and P. Gould. 1971. *Spatial Organisation: The Geographer's View of the World*. Constable, Edinburgh
- Agusta, Y. and D. L. Dowe 2003. Unsupervised Learning of Correlated Multivariate Gaussian Mixture Models using MML. in: Proceedings 16th Australian Joint Conference on Artificial Intelligence, Perth, Australia, 3-5 December 2003.
- Ahl, V. and T. F. H. Allen 1996. *Hierarchy Theory, a Vision, Vocabulary and Epistemology*. Columbia University Press.
- Allen, T. F. H. and T. B. Starr 1982. *Hierarchy: Perspectives for Ecological Complexity*. University of Chicago Press, Chicago.
- Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157, 101-118.
- Babad, Y. M. and Hoffer, J. A. 1984. Even no data has value. *Communications Association Computing Machines* 27: 748-756
- Bar-Yam, Y. 2002. Sum rule for multiscale representations of kinematically described systems. *Advances in Complex Systems* 5:409-431.
- Boulton, D. M. and C. S. Wallace. 1973a. An information measure for hierarchic classification. *Computer J.* 16: 254-261.
- Boulton, D. M. and C. S. Wallace 1973b. Occupancy of a rectangular array. *Computer J.* 16:57-63.
- Boulton, D. M. and C. S. Wallace. 1975. An information measure for single-link classification. *Computer J.* 18: 236-238.
- Boerlijst, M. C and P. Hogeweg. 1991. Self-structuring and selection: spiral waves as a substrate for evolution. In: C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen (eds.) *Artificial Life II*. Adison Wesley Redwood City. pp. 255-276.
- Boerlijst, M. C and P. Hogeweg. 1995a. Spatial gradients enhance persistence of hypercycles. *Physica D* 88:29-39.
- Boerlijst, M. C and P. Hogeweg. 1995b. Attractors and spatial pattern in hypercycles with negative interaction. *J. theoret. Biol.* 176:195-210.
- Brown, J. H. 2001. Mammals on mountainsides: elevational patterns of diversity. *Global Ecol. & Biogeography* 10:101-109.
- Bunge, M. 1969 Metaphysics, epistemology and methodology of levels. In: L. L. Whyte, A. G. Wilson and D. Wilson (eds.), *Hierarchic Structures*, American Elsevier, New York. pp. 17-28.
- Californian Department of Fish and Game 2003. *List of California Terrestrial Natural Communities Recognized by The California Natural Diversity Database*. Wildlife and Habitat Data Analysis Branch, The Vegetation Classification and Mapping Program.
- Ciazlo, A. and D. Ferer. 2000. eCognition Evaluation: Object oriented image analysis EOP-AGR Definiens Imaging, Munich <http://www.definiens-imaging.com/>
- Clifford, H. T. and D. W. Goodall. 1967. A numerical contribution to the classification of the Poaceae. *Australian J. Bot.* 15: 499-519.
- Critchlow, D. E. 1985. *Metric Methods for Analyzing Partially Ranked Data*. Lecture Notes in Statistics 34, Springer Verlag, Berlin.
- Dale, M. B. 1976. Hierarchy and level: prolegomena to a cladistic classification. Technical Memorandum, Division Tropical Crops and Pastures, CSIRO, Brisbane. pp. 9
- Dale, M. B. 1988. Some fuzzy approaches to phytosociology: Ideals and Instances. *Folia Geobotanica Phytotaxonomica* 23: 239-274.
- Dale, M. B. 1999. The dynamics of diversity: mixed strategy systems. *Coenoses* 13:105-113.
- Dale, M. B. 2000. Mt Glorious Revisited: Secondary succession in subtropical rainforest. *Community Ecol.* 1:181-193.
- Dale, M. B. 2001. Minimal message length clustering, environmental heterogeneity and the variable Poisson model. *Community Ecol.* 2:171-180.
- Dale, M. B. 2002. Models, measures and messages: an essay on the role for induction. *Community Ecol.* 3:191-204.
- Dale, M. B. and M. Anand. 2004. Domain knowledge, evidence, complexity and convergence. *International J. Ecol. Environmental Sci.* in press
- Dale, M. B. and D. J. Anderson. 1973. Inosculate analysis of vegetation data. *Australian J. Bot.* 21: 253-276.
- Dale, M. B. and H. T. Clifford 1976. On the effectiveness of higher taxonomic ranks for vegetation analysis. *Australian J. Ecol.* 1:37-62.
- Dale, M. B. and P. Hogeweg. 1998. The dynamics of diversity: a cellular automaton approach. *Coenoses* 13:3-15.
- Dale, M. B. and L. Quadraccia. 1973. Computer assisted tabular sorting of phytosociological data. *Vegetatio* 28:57-73.
- Dansereau, P., P. F. Buell & R. Dagon. 1966. A universal system for recording vegetation II. A methodological critique and an experiment. *Sarracenia* 10: 1-64.
- Davidson, I. 1996. Clustering using the Minimum Message Length criterion and simulated annealing. In: *Proceedings 3rd International Workshop Artificial Intelligence*, Prague.
- DeJong, P., L. W. Aarsen and R. Turkington. 1980. The analysis of contact sampling data. *Oecologia* 45:322-324.

- Doherty, M. D. and N. C. Coops. 1995. Vegetation of the Batemans Bay study area. *Report on the Batemans Bay Vegetation Dataset CD-ROM*. CSIRO Australia. Canberra.
- Glenn, S.M. & Collins, S.L. 1993. Experimental analysis of patch dynamics in tall-grass prairie plant communities. *J. Veg. Sci.* 4: 157-162.
- Goodall, D. W. 1953. Objective methods for the classification of vegetation. 1. the use of positive interspecific correlation. *Australian J. Bot.* 1: 39-63.
- Goodall, D. W. 1974. A new method for the analysis of spatial pattern by random pairing of quadrats. *Vegetatio* 29:135-146.
- Goodall, D. W. and E. Feoli. 1988. Application of probabilistic methods in the analysis of phytosociological data. *Coenoses* 1: 1-10.
- Hartigan, J.A. 1975. *Clustering Algorithms*. Wiley, New York.
- Hartigan, J.A. 1981. Consistency of Single Linkage for High-Density Clusters. *Journal American Statistical Association* 76: 388-394.
- Herben, T., F. Krahulec, V. Hadincová and H. Skálová. 1993. Small-scale variability as a mechanism for large-scale stability in mountain grasslands. *J. Veg. Sci.* 4: 163-170.
- Hill, M. O., R. G. H. Bunce and M. W. Shaw. 1975. Indicator species analysis: a divisive polythetic method of classification and its application to a survey of native pinewoods in Scotland. *J. Ecol.* 63: 597-613.
- Hogeweg, P. 1998. On searching generic properties in non-generic phenomena: an approach to bioinformatic theory formation. In: C. Adami, R. Belew, H. Kitano and C. E. Taylor (eds.) *Artificial Life VI*, MIT Press. pp. 285-294.
- Hogeweg, P. 2002. Computing an organism: on the interface between informatic and dynamic processes. *BioSystems* 64:97-109.
- Huston, M. A. 1999. Local processes and regional patterns: Appropriate scales for understanding variation in the diversity of plants and animals. *Oikos* 86: 393-401.
- Jardine, N. and R. Sibson. 1971. *Mathematical Taxonomy*. Wiley, London.
- Kershaw, K. A. 1970. An empirical approach to the estimation of pattern intensity from density and cover data. *Ecology* 51:729-734.
- Kiers, H. A. L. 1994. SIMPLIMAX: Oblique rotation to an optimal target with simple structure. *Psychometrika* 59: 567-579.
- Kiers, H. A. L. 1998. Three-way SIMPLIMAX for oblique rotation of the three-mode factor analysis core to simple structure. *Computational Statistics and Data Analysis* 28: 307-324.
- Kodratoff, Y. 1986 *Leçons d'apprentissage symbolique*. Cepadues-ed., Toulouse.
- Krishna-Iyer, P. V. 1949. The first and second moments of some probability distributions arising from points on a lattice and their application. *Biometrika* 36: 135-141.
- Lambert, J. M. and W. T. Williams. 1962. Multivariate methods in plant ecology IV: Nodal Analysis *J. Ecol.* 50:775-803.
- Lang, G. 1970. Die Vegetation der Brindabella Range bei Canberra. *Mathematisch-Naturwissenschaftlichen Klass Jahrgang 1970 - NR. 1*. Akademie der Wissenschaften und der Literatur, Mainz.
- Laurence, S. and E. Margolis. 1999. Concepts and cognitive science. In: E. Margolis and S. Laurence (eds.), *Concepts: Core readings*. MIT Press, Cambridge MA. pp. 3-83.
- Legendre, P. and E. D. Gallagher 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271-280.
- Lomolino, M. V. 1996. Investigating causality of nestedness of insular communities: selective immigrations or extinctions? *J. Biogeogr.* 23:699-703
- MacKay, D. 1969. Recognition and action. In: S. Watanabe (ed.), *Methodologies of Pattern Recognition*. Academic Press, London. pp. 409-416.
- Maslof, A.A. 2001. Spatio-temporal dynamics of populations in forest plants and testing of carousel model in the *Myrtillus*-type pine forest. *Bulletin Moscow Society Naturalists Biological Series* 106:59-65.
- Oliver, J. J., R. A. Baxter and C. S. Wallace. 1998. Minimum message length segmentation. In: X-D. Wu, K. Ramamohanarao and K. Korb. (eds.), *Research and Development in Knowledge Discovery and Data Mining*, Lecture Notes in Computer Science 1394, Springer Verlag, Berlin. pp. 222-233.
- Pagie, L. and P. Hogeweg. 1997. Evolutionary consequences of co-evolving targets. *Evolutionary Computation* 5:401-418.
- Patterson, B. D. and W. Atmar. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biol. J. Linnaean Soc.* 28:65-82.
- Poulin, R. and J. F. Guégan. 2000. Nestedness, anti-nestedness, and the relationship between prevalence and intensity in ectoparasite assemblages of marine fish: A spatial model of species coexistence. *International J. Parasitol.* 30: 1147-1152.
- Rietkerk, M., M. C. Boerlijst, F. van Langevelde, D. HilleRisLambers, J. van der Koppel, L. Kumar, H. H. T. Prins and A. M. de Roos. 2002. Self-organization of vegetation in arid ecosystems. *American Naturalist* 160:524-530.
- Rodin, R., H. W. Mili, R. Mineau, R. Missaoui, A. Arfi and T-T. Chau. 1998. Design of class hierarchies based on concept (Galois) lattices. *Theory & Application Object Systems* 4:117-134.
- Sandland, R. L. and P. C. Young. 1979. Probabilistic tests and stopping rules associated with hierarchical classification techniques. *Australian J. Ecol.* 4: 399-406.
- ter Braak, C. J. M. and I. C. Prentice. 1988. A theory of gradient analysis. *Adv. Ecol. Res.* 18: 271-318.
- Van der Maarel, E. and M. T. Syke. 1993. Small-scale plant species turnover in a limestone grassland: the carousel model and some comments on the niche concept. *J. Veg. Sci.* 4: 179-188.
- Vandvik, V. 2004. Gap dynamics in perennial subalpine grasslands: trends and processes change during secondary succession. *Journal Ecology* 92:86-96.
- Wallace, C. S. 1998. Intrinsic classification of spatially-correlated data. *Computer Journal* 41: 602-611.
- Wallace, C. S. and D. L. Dowe. 2000. MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. *Statistics Computing* 10, 73-83.
- Webb, L. J., J. G. Tracey, W. T. Williams and G. N. Lance. 1967. Studies in the numerical analysis of complex rain-forest communities I a comparison of methods applicable to site/species data. *Journal Ecology* 55: 171-191.
- Westhoff, V. and E. van der Maarel. 1978. The Braun-Blanquet approach. In: R. H. Whittaker (ed.) *Classification of Plant Communities*. Dr. W. Junk, Den Haag. pp. 287-399.
- Williams, W. T. and M. B. Dale. 1962. Partition correlation matrices for heterogeneous quantitative data. *Nature* 196: 602.
- Williams, W. T., G. N. Lance, L. J. Webb, J. G. Tracey and J. H. Connell. 1969. Studies in the numerical analysis of complex

rain-forest communities IV. A method for the elucidation of small-scale forest pattern. *J. Ecol.* 57:635-654.

Wishart, D. 1969. Mode analysis: a generalisation of nearest neighbour which reduces chaining effects. In: A. J. Cole (ed.), *Numerical Taxonomy*. Academic Press, London. pp. 282-234.

Wright, D. H., B. D. Patterson, G. M. Mikkelson, A. Cutler and W. Atmar. 1998. A comparative analysis of nested subset patterns of species composition. *Oecologia* 113: 1-20.

Wright, D. H. and J. Reeves. 1992. On the meaning and measurement of nestedness of species assemblages. *Oecologia* 92:416-428.

Xiao, L., D. Wissmann, M. Brown and S. Jablonski. 2002. Hierarchical Concept Description and Learning for Information Extraction. Proceedings 6th Natural Language Pacific Rim Conference, NLP2001, Tokyo. pp. 299-306

Appendix

Because of sickness it was not possible for Chris Wallace to update this description of the hierarchical clustering process. This description follows the original paper of Wallace and Boulton and the program actually used employs a more sophisticated search strategy. The MML formulae remain the same.

Minimum message length for an hierarchical clustering

Continuous attribute d is measured with accuracy $\epsilon[d]$, and with mean $m[d,t]$ and standard deviation $s[d,t]$ for class t .

Multistate attribute d has $M[d]$ states and distribution $\{p[m,d,t]\}$ ($m=1..M[d]$) where $p[m,d,t]$ is the probability that a thing in terminal class t will have attribute d in state m .

If we have a population of U members which is split into sub-clusters A and B, $|A| \leq |B|$, the description of this split will require a message of length

$$\frac{1}{2} \left(\ln \left(\frac{U(U-1)^2}{12AB} \right) + 1 \right) + \ln(2).$$

This is based on the occurrence of the split and the relative abundance of the subclasses.

Next we need to specify the subclass distribution functions. Let class A with N members divide into B and C with L and M members, means a , b and $s.d.$ r and t , respectively. For some continuous attribute, specify the class mean of A and the class standard deviation of A to accuracies of $\pm 0.5s \sqrt{12/N}$ and $\pm 0.5s \sqrt{6/(N-1)}$, respectively. For the subclasses, we must use $a \pm 0.5 \sqrt{12/L}$ and $b \pm 0.5 \sqrt{12/M}$ and for standard deviation $r \pm 0.5r \sqrt{6/(L-1)}$ and $t \pm 0.5r \sqrt{6/(M-1)}$. For simplicity drop the -1 . The total message length needed for a continuous attribute is

$$\ln \left(\frac{\pi s^2}{12\sqrt{2}r^2t^2} \sqrt{N(Ns^4 + Lr^4 + Mt^4)} \right).$$

For a discrete attribute with Q states, the full expression is given by

$$0.5 \ln \left\{ \frac{LM}{N} \prod_{i=1}^Q \frac{n_i}{l_i m_i} \right\} + \ln(Y) - \frac{Q-1}{2} \ln(12)$$

where Y is an occupancy constant and given in Boulton and Wallace (1973).

For a binary split $Y = 1 + \min(L, M, n_1, n_2)$

If N_t is the population of terminal class t , then the message for the things is given by $\ln(S/N)$

Values for attributes are:

continuous $N \left(\ln \left(\frac{\sqrt{2\pi} s}{\epsilon} \right) + 0.5 \right) + 0.5$

discrete $\sum_{q=1}^Q n[q] \ln \frac{N+Q}{n[q]+1} + 1 + 0.5(Q-1)$

We then sum all the pieces to get an overall message length for the entire hierarchy: i.e., sum for each class the attribute values and add $N \ln(S/N)$ to code the class labels, sum over all terminal classes and add the additional cost of the hierarchic encoding of class properties and dendrogram structure.

For clustering, we examine the various message lengths for class A and its subclasses B & C together with the conditional information given the properties of A and if $I_{B^+} + I_{C^+} - I_{(B,C|A)} > I_A$, we merge B and C. By maintaining subclasses we can also look for possible splits. A standard SAHN hierarchy can also be used agglomeratively. However, an agglomerative strategy leads to difficult decision rules and the hierarchy can be fragmented into several disconnected subtrees. A divisive strategy can also be obtained although things may not be assigned ideally because once a division is made it is not reconsidered. However divisive strategies allow easy decision rules.

Davidson (1996) presents an MML evaluation that incorporates context, suggesting the use of

$$-\log[\text{Prob}(H_i|D, c)] = -\log[\text{Prob}(H_i|c)] - \log[\text{Prob}(D|H_i, c)] + \text{const.}$$

for context c . Note that $-\log(\text{probability})$ is in fact the minimal coding length required. Thus, in principle we can adjust the clustering according to the particular vegetation context, while capturing our own views on the probability of a hierarchy in that context in the prior probabilities assigned. The context, for example, might represent certain

laws that we feel are acceptable. By comparing with the unconstrained result, we can assess the value of the laws.

Minimisation strategies

If we fix the class membership and hierarchic structure we can choose the estimates of class parameters to minimise the information by using essentially the maximum likelihood estimates. For fixed class properties and fixed hierarchy we can minimise the information by assigning each thing to its most probable class. Both these operations are independent of the hierarchical structure and so can be applied cyclically to find a local optimum. Davidson (1996) has considered using simulated annealing for the non-hierarchical case.

Given any two terminal classes we can ask if merging them will improve matters. Given non-terminal A with size N and its terminal subclasses B and C of sizes L and M, we can calculate I_B and I_C the message lengths to give the attributes of B and C, and also I_A , for specifying the attributes of A. We can also work out $I_{(B,C|A)}$ the information needed to specify the existence relative sizes and properties of B and C given the properties of A. It is then advantageous to combine B and C if $I_B + I_C + I_{(B,C|A)} > I_A$. All these elements are independent of the hierarchic structure and any classes other than A. If the hierarchy is not yet established we can try all pairs - note that this does not examine if maintaining B and C might not be preferable if a hierarchy was *not* required. It is possible that the combination is preferable to maintaining B and C as subclasses of A.

To increase the number of classes we maintain subclasses of every class as 'half-classes' that are treated as if they were established and can be tested to see if promotion to 'real' classes is worthwhile. Testing using $I_B + I_C + I_{(B,C|A)} > I_A$ is then applied to the half-classes.

An agglomerative strategy based on the terminal groups is then used to develop a hierarchy. However, such a hierarchy may be disjoint, thus combining non-hierarchical and hierarchical elements.

The agglomerative hierarchy has a significant drawback. Assignment of new things has to be made to the terminal classes and there is no easy 'decision-making' rule proceeding down the hierarchy as the distribution within the non-terminal classes is undefined (maybe some non-parametric rule might be advanced?). By using a divisive strategy of splitting classes into subclasses wherever this is profitable does provide a decision rule but at the cost of losing the assignment of things to their most probable terminal class - once a split is made it is never reconsidered and it can be wrongly assigned at an early stage (maybe a monothetic split could also be introduced by constraining the subclasses to be monothetically definable?) Also it assumes that the distribution within each high level class is the same as that for a terminal class, an assumption that is certainly dubious. In general, the message length for the divisive result will be greater than that for the agglomerative result (and this latter will probably exceed that for a non-hierarchical solution!).

In SNOB it is usual to use a fuzzy assignment to classes with things partially belonging to classes in order to provide consistent estimates. It might be possible to allow partial assignment in the hierarchical form by using weighted means and S. D's which should make the estimates consistent.

Single linkage hierarchy

A common, and once recommended (Jardine and Sibson 1971) method of obtaining a hierarchy, based on dissimilarity measures, is single linkage clustering. The MML single linkage approach of Boulton and Wallace (1975) was not used here, since the tree obtained using this method does not address the hierarchy as a means of minimising message length. In fact it is usually used to develop a cluster tree based on a hierarchy of dense (modal) clusters. (see Wishart 1969; Hartigan 1975, 1981). Wallace and Boulton show that underlying this approach is an assumption that the occurrence of a thing in the space means that it is likely other things will be found 'near' it. The choice of dissimilarity measure is itself a difficult one that is best avoided if possible.