

Mt Glorious revisited: secondary succession in subtropical rainforest

M. B. Dale

*Australian School of Environmental Studies, Griffith University, Nathan, Queensland 4111, Australia,
E-mail: m.dale@mailbox.gu.edu.au*

Keywords: Clustering, Explicability, Minimum Message Length, Prediction, Rainforest, Succession.

Abstract: In this paper, I re-examine the subtropical rainforest succession previously studied by Williams, Lance, Webb, Tracey and Dale (1969) (WLWTD) using a clustering procedure based on the *Minimal Message Length* principle of induction. This principle permits the optimal number of clusters to be estimated automatically. Optimality is defined here as a trade-off between quality of fit and complexity of model, both measured in message length units.

Because of the common unit of measurement, we can assess the numerical effectiveness of the procedures adopted in the previous study and compare the results obtained by using density as against presence/absence data or the value of numeric data independent of presence/absence effects. The results also bear on the "principle of explicability" which posits that users seek interpretable results, even if they are less efficient in purely numerical terms.

The optimal density result identified 8 clusters, although these were further clustered into 3 higher level groupings. The pattern of 2 temporal stages followed by spatial segregation is clear, with extra detail concerning aberrant stands and temporal dependency in the third spatial stage also apparent. This analysis was the most effective at recovering structure in the data, of those examined.

Imposing the WLWTD analysis on density data was markedly suboptimal and even the number of clusters recognised (7) was strictly incorrect. However, by subjective interpretation WLWTD selected a number of clusters which was very close to the optimal density solution. For this reason insight gained into the processes operating was not overly compromised. The optimal density result cleans up a few corners and adds more detail but the main outlines are sufficiently clear in the subjectively assessed presence data.

The results from optimal presence/absence analysis were understandable and effective, though considerably less detailed than those obtained using the density data or those from WLWTD's original analyses. Indeed the 3 clusters established using the presence data reflect the higher level of structure which is recognisable in the density result. Using numeric data with 0 values set to missing values, showed little of interest.

Invocation of Kodratoff's principle of explicability, which argues for interpretability to dominate efficiency, was unnecessary since the efficient analyses were directly interpretable. The introduction of domain knowledge during the subjective interpretation in the original analysis was apparently sufficient to counter any losses due to the inefficiency of the clustering method. Given more effective clustering methods and using the density data, it becomes unnecessary.

Abbreviations: MML-Minimal Message Length, N0M- Numeric data with 0 values set to missing values, WLWTD-Williams, Lance, Webb, Tracey and Dale (1969)

Introduction

Studying change

Change in vegetation is ubiquitous. It ranges from growth, ontogeny, birth and death to the long term evolutionary changes reflecting climatic patterns and selection. Rainforest is particularly difficult to study because of the large number of species involved. Such high dimensionality either requires an exorbitant amount of data collection or a change in the manner of description before many

techniques of analysis can be applied. Clustering is one means of dimensionality reduction which is commonly employed. In this paper, I shall use clustering to investigate short term changes in rainforest after clearance.

WLWTD suggested four different approaches to studying change in vegetation. They are:

Palynology: Studies of fossil and sub-fossil material provide an historic record, although there has been too little study of the details of change. Dale and Walker (1970)

examined problems in zoning diagrams, which provides an alphabet for the sequences (see also Dale and Barson 1989). However, pollen records are restricted to particular environments, and there are problems of interpretation since the pollen flora may not clearly represent the actual flora present. Further, the taxonomic precision of pollen identification is different from that used in vegetation studies. For long time periods, pollen provides an effective tool, but its temporal resolution may not be sufficient to link it with short term studies and we cannot always be certain of the effects of possible evolutionary changes.

Spatial surrogates: Spatial gradients are often assumed to reflect temporal stages; i.e., they are regarded as a spatial representation of a temporal change. The validity of such an assumption is obviously heavily dependent on the domain knowledge of the investigator. But this assumption of the equivalence of space and time is a dangerous reification; the appearance of an environmental series may be deceptive, as in the case with presumed hydroseres. In any case, patterns need not be environmentally determined (Boerlijst and Hogeweg 1991, Dale and Hogeweg 1998, Dale 1999), which means that changes in direction of the series can occur without environmental correlates.

Inference from snapshots: Inferring processes and making predictions from a sample taken at a single time is characteristic of many vegetation studies. This again relies heavily on the ability and experience of the investigator and their reliability will also depend markedly on the vegetation processes. Dale and Hogeweg (1998) identified 3 major types of process, only one of which would be amenable to analysis through snapshots; the other two are innately resilient to invasion and involve a cyclic disturb-and-recover sequence, so samples taken at one time will likely have different origins and different futures. Critchley (2000) has suggested that vegetation types are inappropriate for management purposes, which would suggest that they are not 'predictive enough' to be valuable. However, short term forecasting might be possible through introducing ontogenetic information (Gatsuk et al. 1980).

Tracking disturbances: Examination of vegetation change over a short period of observation is the method adopted by WLWTD and in many other studies. Often the time of study is an enforced limitation because of granting restrictions and the exigencies of these! Ideally this fits into a BACI framework (Before-After, Control-Intervention) but it is also common to make opportunistic usage of impacts on vegetation, such as fire or clearance, without knowledge of the pre-existing state. However, the time scale is often inappropriate for the organisms, such as trees which are long-lived or species with poor disper-

sal but with ability to invade. The time limit is mitigated where at least one of the investigators maintains a long-term interest in the vegetation. In that case, it is possible to observe whether predictions of the future course of the change processes are indeed correct. Attempts to evaluate the effectiveness of such short series are therefore of interest.

The effectiveness of limited observational periods can also be extended if several overlapping sequences are available. Multiple historical clearances of known date provide such a series of fragmentary descriptions which might be melded into a coherent series. Such would also be possible for pollen diagrams if the temporal resolution was sufficiently fine. But there are difficulties associated with using such data, since it raises several questions. How do we determine if the vegetation was homogeneous initially and was that homogeneity preserved during and after clearance (Wildi and Schütz 2000). Formally, too, the problem is difficult. Finding supersequences for collections of strings was a big problem in human genome sequencing; the complexity is known to be NP-complete. There is also the problem, not present in DNA, that we may be observing several different series. Visual assessment through techniques such as Principal curves, Multi-dimensional Scaling or Principal Coordinates Analysis may however be sufficiently persuasive as to allow us to proceed without too much fear.

The Mt Glorious study

The original analysis of the Mt Glorious data falls into the fourth class. WLWTD employed several numerical methods, all of which provided some information. Finally they recommended clustering all samples together as the most informative, followed by construction and analysis of transition matrices derived from these cluster sequences.

The clustering method used, Williams et al.'s (1966) Agglomerative Information Analysis, was based on presence data only and the choice of number of clusters was a subjective decision (cf. Austin 1970). WLWTD do not in fact provide any precise information on the criteria actually used to establish the number, which was based largely on interpretability and interestingness. The former includes conformity with known beliefs or facts (Pazzani and Kibler 1992) and possibly simplicity, while the latter relates to deviation from expectation and, in goal-directed situations, to actionability (Barsalou 1995, Hilderman and Hamilton 1999). It is impossible to remove all subjectivity from an analysis as MacKay (1969) and Watanabe (1969) have shown, though we can clearly demarcate it and sometimes replace some of it by objective criteria.

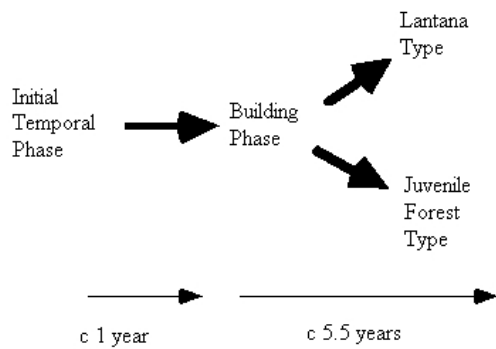


Figure 1. Simplified successional sequence.

The WLWTD results can be briefly summarised as follows. They elected to accept 7 clusters. 2 of these clusters represented initial temporal phases, one cluster occupying the first 2 time periods and the other the next three. There is then an abrupt change from ‘pioneer’ to ‘building’ phase. Subsequently the remaining clusters distinguish spatial patterns without temporal significance. This spatial pattern indicates that stands 8, 9, 10 (and perhaps 5) have some special distinctiveness, attributed by WLWTD to the presence of *Lantana camera*. I shall refer to these as aberrant stands. The whole can be simply represented as shown in Fig. 1.

Questions for re-analysis

In re-analysing these data, I have not followed the original in using a transition matrix-based analysis to investigate the temporal sequences themselves. In most cases, the patterns so distinguished can be visually appreciated from appropriate diagrams where the cluster assignment labels are displayed using the spatial and temporal coordinates. Further, there are alternative means of examining sequences (see Dale and Barson 1989) and analysis of transition matrices may no longer be the best choice.

The original data recorded the density of species in 10 stands at 12 times, but density was not used in the clustering, and the number of groups was subjectively assessed. I have sought to use an improved clustering method, based on the MML principle, to model the data. This permits the evaluation of the quality of a clustering. This leads to the following questions:

- What is the optimal number of clusters for a density-based solution? Is it interpretable? How effective is it in recovering structure?

- How good was the original presence/absence analysis? That is, if I impose the 7-cluster solution on the numeric data does this give an acceptable result?
- What is the optimal presence-based solution and how does it differ from the 7-group result of WLWTD or the density analysis?
- How does this optimal presence clustering differ from the density-based result?
- It is possible using Williams and Dale’s (1962) partitioning to separate the effects due to presence from those due to abundance conditional on presence. If the presence data are separated out, do the residual NOM provide an adequate basis for analysis?

In addition, I have a further question to raise. This is concerned with trading efficiency of analysis with interpretability in accordance with Kodratoff’s (1986) principle of explicability, which posits that users should be able to interpret results, even if they are less efficient. There is, after all, little point in obtaining a ferociously efficient result which is inexplicable! Diday (1988) has used this principle as an argument for logical as opposed to statistical models.

Collecting presence data is possibly easier than collecting density data; ‘possibly’ because in general most individual plants would have to be examined to confirm that no further species are present. If the results are of roughly the same interest then even if they are less efficient descriptions of the vegetation they may be adequate.

The minimum message length principle

Instead of Information Analysis, the clustering method employed here is based on the MML principle, as implemented in the Snob program (Wallace and Boulton 1968, Boulton and Wallace 1970, Wallace and Dowe 2000). Put simply, this principle argues that to find a good model of some data we need to balance the quality of fit of a model against the complexity of that model, in effect providing an operational version of the principle attributed, probably falsely, to William of Occam’s ‘*Entia non sunt multiplicanda praeter necessitatem*’

As a simple example consider a simple regression equation $y = a + b(x) + e$. The simplest form would involve only the constant term a , but in general this would not give us a good fit to the data. Alternatively we can fit a polynomial of order $(n-1)$ to n points $y = a + \sum b_i x^i + e$ which will give perfect fit, though one we are unlikely to feel would apply to any new information. Obviously we need to compromise between the simplicity of the former

and the quality of fit of the latter. MML provides a principled way of doing this.

Somewhat more formally, we start from the notion of an 'explanation'. An explanation encodes a body of data by first stating a 'theory' about the data and then encoding the data in a code which would be optimal were the theory is true. Obviously, the theory, or model, is chosen from a class of possible models, since we cannot envisage investigating every possible model. MML selects a model of the data from a class of models each of which has a particular value of a parameter. MML proceeds by estimating the parameter value so as to minimise the message length needed to optimally encode the data. This can be shown to be equivalent to determining the model with the maximum *a posteriori* probability, with a probability p being encoded by a message of length $-\log_2(p)$.

Examples of classes of model include clustering, where the parameter is the number of clusters, ordination, where the parameter is the number of axes and Markov models where the parameter is the number of states. We might also have considered the class of ordination models, with a parameter for the number of axes, or even a combination of clusters and axes. MML methods for these are indeed known (Wallace 1995, Edwards and Dowe 1998), though the programs are not presently available for general use. Note that this combined solution provides a means of comparing clustering with ordination.

The class of models considered here is that of mixture models with a single parameter, the number of clusters. Of course, while choosing a model we have other values to estimate, such as cluster parameters or factor loadings. MML provides consistent estimates of the parameters of the clusters and determines the clusters to which the things being clustered are to be (partially) assigned. MML is concerned with mixture modelling and not with segmentation where crisp clusters are demanded and no overlap of cluster membership is permitted (MML methods for segmentation are known). For vegetation, samples might contain elements of several clusters due to physical misplacement of sample boundaries so that several types are included, to the existence of ecotones and ecoclines at spatial boundaries or of temporal changes where perhaps one set of species is expanding while another is retracting.

With continuous data, MML employs a coding based on the accuracy of measurement; after all, it seems reasonable that imprecise data should tell us less. This requires some extra terms to be added to the message but does not modify the underlying principles. Note that in choosing between models the actual encoded message need never be constructed, since to compare models we

need only to know the length of their associated messages. Since shorter messages mean better models, only the variable components of the message lengths need be estimated. Indeed, the difference between 2 such message lengths is related to the odds in favour of the shorter, given the data.

The MML principle is avowedly Bayesian. This means that we can express our prior beliefs, and the data will reveal its unexpectedness and hence interest. In the present examples, I have not introduced any personal beliefs as priors and instead have relied on colourless ones; for example, any number of clusters is regarded as equally likely *a priori*. In practice, only extreme values for prior probabilities markedly affect the results obtained.

The Snob program

The Snob program uses the MML principle to determine the number and contents of each cluster of things. In the example presented here the vegetation of one sample at one time is a "thing" while "stand" refers to a particular spatial location sampled several times. Essentially, Snob calculates a message length for each cluster, for the attributes within clusters and for each thing being classified with an additional optimisation reflecting the precision of measurement. Things are assigned to all clusters to some degree, so that the resulting clusters are overlapping and fuzzy in nature. Somewhat surprisingly, such fuzziness can be used to reduce the message length as well.

Briefly, Snob obtains the message length as follows. For a probability p the message length required is $-\log(p)$ bits, so we have, for model H and data D

Message Length = $-\log(H) - \log(D|H) + \text{other terms relating to the precision of measurement.}$

Minimising the message length is equivalent to maximising $Pr(H).Pr(D|H)$ which, by Bayes rule, equals $Pr(D).Pr(H|D)$ and since $Pr(D)$ is independent of H , MML maximises the Bayesian posterior probability $Pr(H|D)$. More precisely, assuming a locally flat prior and a quadratic likelihood function, we have

$E(\text{Message Length}(y, \theta)) = -\log(h(\theta)) - \log(f(y, \theta)) + \text{precision terms,}$

where $h(\theta)$ is the assumed known prior density on θ (the model class parameter which here is the number of clusters) and $f(y|\theta)$ is the likelihood of y given θ . This expression remains an approximation and there may be modifications and extensions necessary in other applications of the principle.

I have assumed a Gaussian within-cluster distribution for the numeric data, so the data are regarded as being sampled from one or more, possibly overlapping, Gaussian distributions. Other possibilities relate to multistate data, angular data and Poisson distributions. Snob can change the number of clusters through fusion and splitting operations, as well as reallocating things between clusters. The result from Snob is based on the Bayesian posterior distribution of the model, so that we are effectively performing a significance test on the number of clusters. The null hypothesis is represented by the data forming a single cluster.

The present program omits many interesting possibilities, such as within cluster variation, the incorporation of temporal and spatial dependencies between things and the possibility of hierarchical arrangements of clusters. It is hoped that these will become available in the near future. However, even in its present form, Snob offers many features which are simply unavailable with any other clustering method, notably the estimation of number of clusters and the consistent estimation of model parameters, so that it is certainly competitive, and possibly superior to, any of its competitors. Its performance on datasets with over 40000 members is certainly adequate, although the search methods used have not received a great deal of attention. It is known, for example, that simulated annealing offers an alternative search possibility which has yet to be examined.

Output

The output includes information on the following.

1. The message length cost of encoding the data as a single cluster which forms a null hypothesis. If there is no cluster model which has a shorter message length, then the null hypothesis is accepted.
2. The estimated number of clusters and the relative probability of each, together with the total cost of encoding the data using that selected number of clusters. I have used the difference between the 1-class message length and the n -class message length as a measure of the amount of structure recovered. In effect it is a measure of redundancy removed by the clustering.
3. For each attribute, within each cluster, the message length required to encode distributional parameters, to some optimal precision, and whether these values differ significantly from those of the population as a whole. For numeric values the parameters are mean and standard deviation (both coded to an optimal precision), while for multistate attributes the probabilities of each class within each cluster are estimated. Appropriate parameters

are encoded for numeric Poisson distributions and angular distributions.

If the attribute does not differ from the population in at least one cluster, a message to that effect is printed, and the full output is suppressed.

4. For each thing the cluster(s) to which it is assigned together with the relative probability of assignment. However, if this probability is $<.01$ then output is suppressed. The number of things which are ambiguously assigned is itself interesting since it reflects the degree of overlap of the groups. I have elected to regard any thing which has a probability ≤ 0.1 for more than one group as being ambiguously assigned.

Problems

Small Clusters: In taking the result as optimal, some caveats are required. First, very small clusters are suppressed, although they can be identified by examining the distribution of message lengths associated with the things in a particular cluster. As outliers, the cost of encoding them using the cluster parameters is generally large.

Within cluster correlation: Snob assumes no within-cluster correlation between attributes and if such correlation exists, extra clusters may be produced. Edwards and Dowe (1998) have proposed incorporating a single axis of (linear) variation within clusters. Not all clusters may demand such an axis, and coding a non-existent axis would increase the message length. Several clusters may share a common axis while others might require several axes, which increases the search space considerably. Finally, the message length expression becomes considerably more complex to evaluate and the introduction of an axis poses problems for the multiple assignment of things to clusters leading to inconsistency in parameter estimates. Techniques for surmounting most of these problems are known, but not implemented in the present program.

Independence: Snob assumes that the things are independent samples. This is clearly not so in the present case. Many plants will persist through time in presence at least. They might also grow larger or more numerous, and with density as our performance measure, we are concerned with number. They may also grow less numerous, become less vital or cover less area, though rarely do they become smaller. Gatsuk et al (1980) have discussed ontogenies which includes break-up of patches. In any case, we must expect some sort of temporal dependency between our things.

In the same way, because the plots are physically adjacent, we might expect some spatial correlation. This will differ from the temporal dependence because in gen-

eral it will not be directional and furthermore it is possible that relatively remote things can interact, for example through transfer of propagules, whereas in temporal dependence, remote (ancestral) interactions would be filtered through intermediates. The MML approach has been extended to cover simple examples of both these cases. Edgoose and Allison (1999) employed a first order Markov process to introduce dependency in time, while Wallace (1998) has looked at some simple forms of spatial dependency. These authors all indicate that further extensions are possible to cope with more complex dependencies, although the amount of data necessary may become large.

Suboptimal solutions: There are computational difficulties because we are seeking a global optimum, since it is possible to obtain local minima rather than global ones. How likely this is depends in part on the separation of the clusters. There are various alternative algorithms which might be employed, but their effectiveness has not been studied. The Snob program allows the user considerable control over its operations and initiating several searches will minimise this effect, but it cannot be eliminated entirely.

Data and analyses

Mt Glorious data

The data, more fully described in WLWTD, consist of 10 stands recorded at 12 different times with the 10 stands arranged in 2 adjacent rows of 5. I shall refer to the units being clustered as things, each thing being the description of one plot at one time. 118 taxa were recorded although some, more especially seedlings, were not identified to species level. So the initial data matrix is 120 x 118. The clustering assigns each thing a relative probability of belonging to each class, but generally the assignments were crisp with the largest probability being 0.99 or greater. I shall comment on cases where any considerable ambiguity of assignment exists. The thing assignments can be simply displayed, where necessary, in a stand x time array of 10 x 12 elements although it must be remembered that the time intervals between samples is not constant. Table 1 provides the actual dates of sampling and for some purposes I have made use of actual number of days from

Table 1. Dates of sampling.

4/2/1958	28/2/1958
6/5/1958	9/7/1958
2/10/1958	31/3/1959
9/12/1959	7/3/1960
14/12/1960	30/11/1961

commencement while for others only the order in sequence is used.

Quantity, presence and optimality

There are four analyses in addition to the original WLWTD results.

- The full density data.
- The density data forced into the groups of WLWTD.
- The optimal presence data solution.
- The numeric data.

Each analysis may differ in number of clusters as well as assignment of things to clusters and the attributes deemed significantly different from the population values might also change. I have left the cluster labels untouched rather than trying to establish similar clusters from several different analyses. Each analysis is represented by a stand x time map of cluster labels, and usually by a list of species showing significant differences from population values at the 1% level.

For comparing results, I have calculated contingency tables between group labels and where possible used correspondence analysis to provide illustrations of these relationships.

Results

General comparison

Table 2 shows the general results obtained for the 4 analyses. It is immediately obvious that the density data contain considerably more information, with the 1-class message length, about 3 times that for presence and NOM data. Equally obvious is the difference in number of clusters estimated for the various analyses. While the density and original WLWTD analyses give a similar number of clusters, both presence and NOM give rather few. WLWTD's original estimate of 7 clusters for the presence data was apparently a gross overestimate which has some implications for the use of suboptimal solutions which are more complex, and for Kodratoff's principle. The WLWTD solution imposed on the density data is also suboptimal as might be expected. A difference of 10 between the n -class message lengths would give odds of more than 2200:1 in favour of the true density solution and the difference here is around 14000!

What is also important is that the differences between 1-class and n -class show that the density analysis is recovering a greater amount of structure, the WLWTD result is almost as good at recovering structure but the presence

Table 2. General characteristics. Message lengths are in nits. The 1-class length represents the null hypothesis.

	1-class message length	Number of clusters	n-class message length	difference 1-class – n-class	difference as proportion of 1-class
Density	11928.8	8	8548.0	3380.8	0.28
WLWTD	11928.8	7	9819.0	2109.8	0.18
Presence	3769.7	3	3346.0	423.7	0.11
NOM	4025.4	2	3611.8	413.6	0.10

and still more the NOM are considerable less effective. However, these results suggest that there is a considerable amount of residual variation remaining.

Density

The density data results provide 8 clusters. These are shown in spatio-temporal distribution in Table 3. Note that only one thing has a notably ambiguous assignment; in fact 3 other things show some small affinity with other clusters. The pattern identified by WLWTD is obvious but considerably more detail has now been incorporated. The initial temporal stage is predominantly composed mostly from things from cluster 4, and rarely clusters 3 and 5. Cluster 3 represents what WLWTD recognised as shoot regrowth from tree stumps and recurs at a late stage in the series in stands 1 and 8. Cluster 5 probably represents species poor vegetation, associated with *Lantana camera*. But while WLWTD identified stands 9 and 10 as aberrant, in this analysis there is a sign of change in the final stage towards juvenile forest as exemplified by cluster 13.

Clusters 10 and 11 are restricted to the second temporal stage, and are themselves spatially segregated - there is only 1 transition between them. Both are abruptly replaced between times 5 and 6. Cluster 12 is almost restricted to stand 5, which WLWTD suggested might be related to the aberrant stands 8, 9 and 10, but here forms a coherent sequence and possible represents another type. Clusters 13 and 14 also form temporally coherent sequences with rather few transitions between them. It seems that temporal dependence is quite strong, and an analysis which permitted the incorporation of such dependence would be extremely interesting.

Species showing significant differences from population values at the 1% level are shown in Table 4. The first point of interest is the high number of negative relationships - species absent or with reduced abundance. Of the positive species groups 10 (especially) and 11 are best indicated. Groups 3, 4, 12, 13 and 14 each have about the same number of positively associated species while group

Table 3. Spatio-temporal distribution of clusters - density data. The 10 stands are shown in their correct spatial relationships. Entries in () indicate other groups to which the thing is assigned with probability p = 0.1.

Time	Days from Start	Stands				
		1	2	3	4	5
1	0	6	7	8	9	10
		3	4	4	5	4
		5	4	4	4	4
2	24	3	4	11	4	4
		4	4	4	4	4
		10	11	11	4	4
3	91	11	10	10	11	11
		10	11	10	11	11
		11	10	10	11	11
4	155	10	11	11	11	11
		11	10	10	11	11
		10	11	11	11	11
5	240	11	10	10	11	11
		10	11	11	11	11
		11	10	10	11	11
6	420	14	14	14	14	5
		11	14	14	5	5
		13	13	13	14	12
7	674	14	14	3	5	12
		13	13	13	14	12
		14	14	3	5	5
8	763	3	13	13	13	12
		14	14	3	5	5
		13	13	13	13	12
9	1045	14	14	3	5	5
		13	13	13	13	12
		14	14	3	5	5
10	1396	14	14	3	5	5
		3	13	13	13	12
		14	14	3	5	13
11	1925	3	13	13	13	12
		14	14	3	5	13
		14	14	3	5	13
12	2339	3	13	13	14	12
		14	14(13)	3	13	13
		14	14(13)	3	13	13

5 has but one! In most cases each group has some unique positively associated species.

Perusal of this table also suggests that in many cases the clusters are themselves clustered so that species are associated with collections of clusters - (3,4,5); (10,11); (12,13,14). This suggests that a level structure or hierarchy may be useful. Again, an MML procedure is known (Boulton and Wallace 1973) but not implemented in the present program.

Overall the results confirm the general outlines of the original WLWTD analyses, and provide, I believe, more detail in matters such as the appearance of rainforest elements in the aberrant stands 8, 9 and 10, while also clearly separating stand 5 from the other aberrant stands.

Table 4. Attributes significant at the 1% level - density data.

Attribute	Significantly low group	Significantly high group
Acacia melanoxylon	3, 5, 13, 14	11
Alangium villosum	11, 13	12
Alocasia macrorrhiza	4, 5, 11, 14	3
Alphitonia excelsa	4, 5, 12, 13, 14	10
Archontophoenix cunninghamii	4, 5, 11	13
Carduus nutans		3, 10
Cayratia clematidea	13, 14	10
Commelina cyanea	4, 5, 14	13
Cucubitateae Gen et Sp	13, 14	4
Cynoglossum sp	4, 5	
Dioscorea transversa	3, 4, 5, 12, 13, 14	10, 11
Duboisia myoporoides	5, 13	
Dysoxylum fraseranum	4, 10, 11, 14	5
Endiandra pubens		12
Erigeron canadensis	4, 5, 12, 13, 14	10, 11
Ervatamia angustisepala	4, 5, 11, 13, 14	3
Eugenia brachyandra	3, 4, 5, 10, 11, 13	12
Eugenia corynantha		10
Eugenia luehmannii	3, 4, 5, 10, 11, 13, 14	12
Eustrephus latifolius	3, 5, 12, 13, 14	10
Guioa semiglauca		3
Hibbertia scandens		12
Homalanthus populifolius	3	13
Hydrocotyle hirta	3, 4, 5	10
Lantana camera	5, 12, 13	10
Legnephora moorei	11, 13	10, 14
Lobelia trigonocaulis	4, 5, 11, 13	14
Lonchocarpus blackii	4	13
Melodinus acutiflorus	4, 5, 11	13, 14
Menispermaceae Gen et Sp?		10
Muhlenbeckia cunninghamii	13, 14	4
Pandorea pandorina	3, 5, 13, 14	10, 11
Panicum pygmaeum	3, 4, 5, 11	
Parsonia ventricosa	3, 4, 5	10
Passiflora alba	3, 5, 12, 13	4, 11
Physalis peruviana	3, 5, 12, 13, 14	4, 10
Phytolacca octandra	3, 12, 13, 14	11
Piper novae-hollandiae	4, 5, 10, 11, 14	13
Podocarpus elatus		14
Pseuderanthemum Sp?	4, 5, 12, 13, 14	11
Rubus rosifolius	4, 5, 12, 14	10
Solanum auriculatum	3, 12, 13	10, 11
Solanum aviculare	3, 5, 12, 13, 14	10, 11
Solanum nigrum	3, 5, 12, 13, 14	10, 11
Solanum sporadotrichum	12, 13, 14	10
Stellaria media	3, 4, 5, 12, 13, 14	10
Trema aspersa	4, 5, 13, 14	10, 11
Urtica incisa	3, 4, 5, 13	10, 11
Zanthoxylum brachycanthum		12
Elaeocarp/Gmelina		3
Panax seedlings	3, 4, 10, 11, 13, 14	12
Seedling with hairy leaf		12

WLWTD Imposed

Results for numeric classes derived from WLWTD (Table 5) are similar to the MML solution with the number of ambiguous assignments slightly greater than the density analysis but fewer than the presence analysis. The proportion of structure recovered is similar to that of the density analysis. Having one less cluster, the idiosyncrasy of stand 5 is lost, while density groups 10 and 11 have been merged into a single group. The 2 temporal stages and the other aberrant stands are still recognisable. There is confusion in the segregation of density groups 12, 13 and 14 as can be seen in Table 6a but the correspondence analysis (Table 6b and Fig. 2) indicates considerable correlation of the results. Species information is not available for this analysis As for the possible level structure, instead of 3 higher level clusters, there are 4.

Table 5. Spatio-temporal distribution of clusters. - density data. WLWTD groups superposed.

Time	Stands				
	1	2	3	4	5
6	y				
1	3	4	4	4	4
2	3	4	6	4	4
3	6	6	6	4	6
4	6	6	6	4	6
5	6	6	6	4	6
6	7	7	7	7	5
7	8	7	7	9	8
8	8	7	8	9	8
9	8	7(9)	3	5	5
10	8	8	8	9	8
11	8	8	8	9	8
12	8	8	8	9	8
	9	9	3	8	8

When comparing this result with other analyses we can recognise various ways in which divergence is possible. In one, called here omission, the alternative analysis recovers some, but not all, of the original clusters with little mixing. In the other the original clusters become merged to provide a blurred result. In a third the things can be markedly differently assigned and provide an alternative structure. These may be combined, of course if some clusters are recovered more or less intact while others are blurred. In this case we have partial recovery of some clusters and a melting pot for the remainder, but there is no evidence for large scale reassignment which might presage an alternative structure.

Presence

The MML presence/absence analysis identifies 3 classes (Table 7) and these are strongly associated with combinations of the clusters of the MML numeric analysis though with increased ambiguity of assignment. The presence data capture the higher level structure although it proportionately captures less structure than the density analysis. Presence group 5 combines density groups 10

Table 6a: Contingency analysis: density and WLWTD solutions. Prediction of density clusters by WLWTD clusters has $R^2 = 0.7798$. **b:** Correspondence analysis of contingency table.

a

		Density Clusters									
		Clusters	3	4	5	10	11	12	13	14	Total
WLWTD	3	8	0	0	0	0	0	0	0	0	8
	4	0	16	0	0	0	0	0	0	0	16
	5	0	0	13	0	0	1	0	0	0	14
	6	0	1	0	10	20	0	0	1	0	32
Clusters	7	0	0	0	0	0	0	4	10	0	14
	8	3	0	0	0	0	6	14	0	0	23
	9	0	0	0	0	0	0	3	10	0	13
Total			11	17	13	10	20	7	21	21	120

b

	Singular Value	Inertia	Portion				
	0.98746	0.97507	0.2434				
	0.96159	0.92465	0.2308				
	0.94660	0.89608	0.2236				
	0.86667	0.75111	0.1875				
	0.67777	0.45937	0.1146				
	0.02440	0.00060	0.0001				
clus0	e1	e2	e3	cluswebb	e1	e2	e3
3	0.8109	-0.6905	0.6409	3	0.8212	-0.7181	0.6771
4	-1.5919	0.8558	1.5452	4	-1.6121	0.8900	1.6323
5	1.2747	2.3423	-0.6344	5	1.2587	2.2523	-0.6006
10	-0.9407	-0.2600	-1.3227	6	-0.9289	-0.2501	-1.2520
11	-0.9407	-0.2600	-1.3227	7	0.5305	-0.7399	0.2755
12	0.8298	-0.1286	0.2888	8	0.7462	-0.5197	0.4190
13	0.6811	-0.6179	0.3905	9	0.5183	-0.7474	0.2650
14	0.4610	-0.7489	0.2089				

and 11, presence group 11 combines density clusters 12, 13 and 14 while presence group 12 unites density clusters 3, 4 and 5. Thus the optimal analysis of the presence data is identifying somewhat gross clusters at a different spatial scale, clusters which are actually recognisable in the density data as well but are there treated in more detail. These features are clearly shown in Table 8ab and Fig. 3.

The presence analysis has several more ambiguous assignments, and is most remarkable because strong links between the later aberrant stands 8, 9 and 10 and the initial

temporal stage. These would be relatively species poor, which might be sufficient to explain the relationship. There are certainly indications of such a relationship in the other analyses although WLWTD indicated that their recognition of relationships between early and late stages related to stem regrowth from felled forest trees. The species associated with the clusters (Table 9) are, with a single exception, a subset of those used in the Density analysis.

NOM analysis

The NOM analysis is largely uninteresting and captures the least proportion of structure of any of the analyses. Only 2 clusters are identified (Table 10) one of which (16) is commoner at the second temporal stage and in the aberrant stands. There is no marked correlation with the density clusters (Table 11) although the correspondence analysis suggests a relationship between cluster 8 and density clusters 5 and 12 and cluster 11 with density cluster 11.

In the light of the other analyses some patterns do seem to be detectable, but the importance of absence information is obvious. Possibly a crude encoding of abundance (e.g., none, a few, a lot) would be more useful and efficient than counting plants! The species involved (Table 12) are a still smaller subset of those used in the density analysis than was found in the presence analysis.

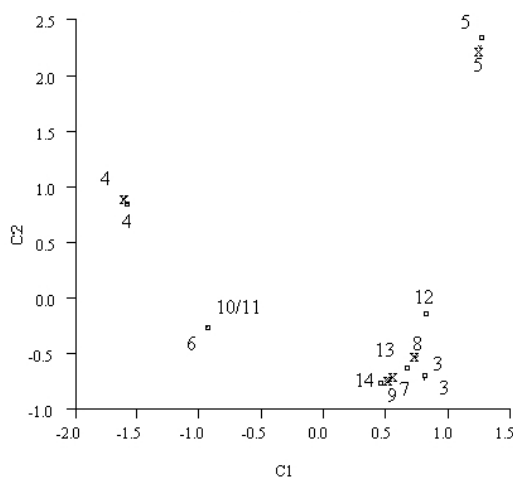


Figure 2. Correspondence Analysis. Density (o) and WLWTD (x).

Table 7. Spatio-temporal distribution of clusters - optimal presence. Entries in () indicate other groups to which the thing is assigned with probability $p \leq 0.1$.

Time	Stands				
	1	2	3	4	5
1	6	7	8	9	10
1	12	12	12	11(12)	12
	12	12	12	12	12
2	12	12	12	12	12
	12	12	12	12	12
3	5	12	5	12	12(5)
	5	5	5	5(12)	5
4	5	5	5	5	5
	5	5	5	5	5
5	5	5	5	5(12)	5
	5	5	5	5	5
6	12	11	11	11	12
	12	12	12	12	12
7	11	11	11	11	11
	11	11	12	12	11
8	11	11	11	11	11
	12	11	12	12	12(11)
9	11	11	11	11	11
	11	11	12	11(12)	12
10	11	11	11	11	11
	11	11	12	12(11)	12
11	11	11	11	11	11
	11	11	12	12(11)	11
12	11	11	11	11	11
	11	11	12	11(12)	11

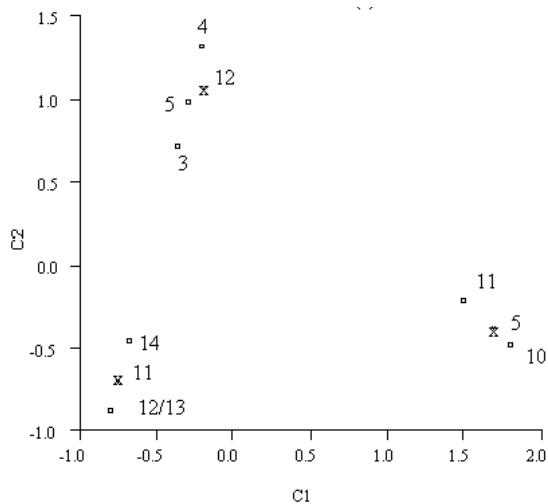


Figure 3. Correspondence Analysis: Density (o) and Optimal Presence (x).

Table 8a: Contingency analysis density and presence $R^2 = 0.7604$. **b:** correspondence analysis.

a

Presence	Density Clusters									Total
	3	4	5	10	11	12	13	14		
Clusters	5	0	0	0	10	17	0	0	0	27
	11	3	0	2	0	0	7	21	17	50
	12	8	17	11	0	3	0	0	4	43
Total		11	17	13	10	20	7	21	21	120

b

Singular Value	Inertia	Portion			
0.94590	0.89472	0.5815			
0.80242	0.64389	0.4185			
clus0	c1	c2	cluspres	c1	c:
3	-0.3655	0.7218	5	1.6965	-0.383
4	-0.2057	1.3223	11	-0.7488	-0.705
5	-0.2959	0.9835	12	-0.1946	1.061
10	1.7935	-0.4772			
11	1.4936	-0.2073			
12	-0.7916	-0.8794			
13	-0.7916	-0.8794			
14	-0.6800	-0.4601			

Table 9. Significant Species Optimal Presence data 1% level. Boldface identifies the unique species not found in the density analysis list.

Attribute	Significantly low group	Significantly high group
Acacia melanoxylon	11	5
Alphitonia excelsa	11	5
Cayratia clematidea		5
Dioscorea transversa	12	5
Erigeron canadensis	11	5
Lantana camera	11	5
Pandorea pandorina	12	5
Panicum pygmaeum	12	11
Physalis peruviana	11	5
Phytolacca octandra	11	5
Rubus rosifolius	12	5
Solanum aviculare	11	5
Solanum nigrum	11	5
Solanum sporadotrichum	11	5
Stellaria media		5
Trema aspersa	11	5
Urtica incisa	12	5

Discussion

The large amount of residual variation remaining can be attributed to two reasons. Either there is considerable random or contingent variation (cf. Brokaw and Busing 2000) or another model of vegetation is needed which will better fit these data. One possibility is the variable Poisson model (Stevens 1937) which regards the spatial environment as composed of patches each with its own quality as a habitat for any species, as in Dale and Anderson's (1973) two-parameter analysis. I am presently investigating this possibility. Another is the more constrained gradient model which in effect orders patches by their quality

Table 10. Spatio-temporal distribution of clusters - NOM data. Entries in () indicate other groups to which the thing is assigned with probability $p \leq 0.1$.

Stage	Days from Start	Stands				
		1	2	3	4	5
1	0	6	7	8	9	10
		8(16)	16	16	8(16)	16
2	24	8	8	8	8	8
		8(16)	16	16	8	16
3	91	8	16	16	16	16
		16(8)	16	16	16	8
4	155	16	16	16	16	16
		16	16	16	16	8
5	240	16	16	16	16	16
		16	16	16	16	8
6	420	16	8	8	8	16
		16	16	8	8	8
7	674	8	8	8	8	16(8)
		8	8(16)	16	8	8
8	763	8	8(16)	16(8)	8	8(16)
		8(16)	8	16	8	8
9	1045	8	8(16)	16	8	8
		8	8	16	8	16
10	1396	8	8	16	8	8
		8	8	16	8	8
11	1925	8	16	16	8	8
		8(16)	8	16	8	16
12	2339	8	16	16	8	16
		8	8	16	16	16

Table 11a: Contingency analysis for NOM. $R^2=0.10$. **b:** Correspondence analysis, singular value = 0.58.

a

Density	Numeric with 0=missing		Total
Clusters	8	16	
3	5	6	11
4	8	9	17
5	11	2	13
10	0	10	10
11	3	17	20
12	5	2	7
13	11	10	21
14	10	2	21
Total	62	58	120

b

Density Clusters	Numeric with 0=missing Clusters	
3	0.1243	8 -0.5666
4	0.0922	16 0.6057
5	-0.6593	
10	1.0339	
11	0.7337	
12	-0.3955	
13	-0.0143	
14	-0.7766	

which is assumed to parallel some environmental gradient(s).

The residual variation would also seem to support the notions of complexity espoused by Anand (2000). While

Table 12. Significant species for NOM at the 1% level.

Attribute	Significantly low group	Significantly high group
Acacia melanoxylon	●	
Dioscorea transversa	●	
Erigeron canadensis	●	
Homalanthus populifolius	●	
Pandorea pandorina	●	
Phytolacca octandra	●	
Rubus rosifolius	●	
Urtica incisa	●	

it is true that at the species level there is considerable complexity, the overall pattern is one which, I would suggest, is much simpler. In general terms the rain-forest re-establishes itself, although some plots have been, temporarily, diverted to *Lantana* dominance. Thus, at a structural or physiognomic level the situation is relatively simple; indeed they conform to the linear phase of Anand (loc. cit.). Further, the presence analysis suggests that even the aberrant stands may be the result of the initial state after clearance permitting a *Lantana* invasion. It seems that, if we chose to describe the vegetation in other ways than species, a simple pattern might emerge

Use of Kodratoff's principle would seem, from these analyses, to be unnecessary. The most complicated analysis captured the greatest amount of structure, and provided clear and interpretable results. There was little need to consider suboptimal solutions at all, although, as the original WLWTD results show, these could be themselves quite illuminating. There may be an argument for examining other local optima which might identify different structures, but there is no need to exclude the optimal solution. There is empirical evidence which purports to show that models slightly more complex than the simplest are actually more effective (Quinlan and Rivest 1989), but here Occam's Razor is vindicated

The density analysis identifies proportionately more structure than any of the other analyses. WLWTD approximates to this with reasonable efficiency. The gross structure of changes can be retrieved using presence data only and the presence clusters also form a higher level of abstraction with each of the 3 clusters found further split by the density analysis. Combining the presence and density result identifies a 2-level structure but this is also visible in the density analysis alone. Such a 2-level structure related to change of performance measure from presence to density was unexpected, and is probably related to the importance of absences.

The 2-level structure is not a hierarchy which strictly invokes dominance relationships between different levels, not equivalence between classes at any one level. In-

Table 13. Ordering of performance measures.

Measure	Qualities measured
Presence	Existence
Density (as expected frequency)	Existence and Number = expected frequency
Valence (as rooted frequency)	Existence, Number and Distribution
% Cover (shoot frequency)	Existence, Number, Distribution and Size
% Cover repetition (shoot) (frequency by layers)	Existence, Number, Distribution, Size and Height Distribution
Biomass (separating components by kind as well as position)	Existence, Number, Distribution, Size, Height Distribution and Physical Density
Function (support, absorptive photosynthetic and reproductive tissues)	Existence, Number, Distribution, Size, Height Distribution, Physical Density, Physiological Function

stead, it is a level structure (Bunge 1969, Dale 1976) which *does* identify equivalence classes at the various levels, but for different levels the equivalence involves qualitative changes in relevant characters. Of course a hierarchical clustering method might be used to identify such a level structure; Boulton and Wallace (1973) have described an MML hierarchical clustering procedure, and the Snob program is being modified to produce an intercluster dissimilarity matrix based on the Kullback-Leibler (1951) distance together with an agglomerative clustering of this matrix. These developments were not available at the time of analysis.

Regarding the interestingness of the results, we might consider comparing the results obtained with those expected by the interpreters. Unfortunately, Webb and Tracey's (subjective) prior probabilities are unavailable and would in any case be difficult to quantify, so unexpectedness is hard to assess. It would have been preferable, in hindsight, to have recorded the expectations before the analyses were completed. Actionability depends on the goals of the analysis which presumably were to identify patterns of change and from them infer possible processes. If we try to distinguish between the relay and the initial floristic composition models of succession, for example, the density data do show weak suggestions of waves of 'invasion' although there are certainly species which were initially present and maintain themselves throughout.

Is MML optimality preferable to the suboptimal hierarchical method used by WLWTD? I suspect that the success of the original WLWTD analysis was the result of the selection of the appropriate, though actually not well supported, number of clusters, and that minor misallocation was insufficient to impede the determination of broad

structures. This almost certainly reflects the domain knowledge brought to the determination of cluster number by Webb and Tracey. Interestingly, initialising the Snob search using the WLWTD clusters identified the optimal density solution whereas initialisation using 8 random clusters often did not. In message length terms the density analysis, even with more clusters, is markedly superior and it is also much more successful in structure identification.

MML relies on balancing the message length cost of complexity of a model with quality of fit. The results here obtained suggest that this principle is effective in providing estimates of the number of clusters and estimating cluster parameters. But MML is not restricted to determining the number of clusters. We might, for example, wish to examine various measures of performance to seek an optimal one, though here we ought also to examine the practicalities and costs of data collection which would complicate the assessment of complexity. If we look at the performance measures commonly employed in vegetation study, we can interpret them as forming a cumulative sequence of qualities whose structural content is to be assessed as shown in Table 13. All of these can be regarded as probabilities. There would be some interesting problems in providing appropriate prior probabilities.

MML has proved effective in several other problem areas. It can provide a consistent estimation procedure for Multiple Factor Analysis (Wallace 1995), which Maximum Likelihood estimation cannot, and the transformations suggested by Legendre and Gallagher (2000) would extend the range of such ordination procedures to various metrics other than Euclidean. MML, or the related Minimum Description Length principle (Rissanen 1995), has also been used in regression studies for feature selection, transformation selection, determining appropriate powers for polynomials and for the detection of outliers as well as in the construction of decision trees and graphs. This means that it can be used to select optimal response functions from families such as that proposed by Huisman et al. (1993). It has also been used in model-averaging approaches to prediction, to string similarity, and in the inference of causal (path) models.

MML can be extended to cover collections or families of model classes. In principle it could choose between axis-based models and cluster-based models or plexus models. Turning this principle into practice could provide extremely useful tools for vegetation science.

References

- Anand, M. 2000 Fundamentals of vegetation change: complexity rules. *Acta Biotheoretica* 48:1-14.

- Austin, M. P. 1970. An applied ecological example of mixed data classification. In: R. S. Anderssen and M. R. Osborne (eds.), *Data Representation*, Univ. Queensland Press, Brisbane. pp. 113-117.
- Barsalou, L. W. 1995. Deriving categories to achieve goals. In: A. Ram and D. B. Leake (eds.), *Goal Directed Learning*. MIT Press Cambridge MA. pp. 121-176.
- Boerlijst, M. C. and P. Hogeweg. 1991. Spiral wave structure in prebiotic evolution: hypercycles stable against parasites. *Physica D* 48: 17-28.
- Boulton, D. M. and C. S. Wallace 1970 A program for numerical classification *Comput. J.* 13: 63-69.
- Boulton, D. M. and C. S. Wallace. 1973. An information measure for hierarchic classification. *Comput. J.* 16: 254-261.
- Brokaw, N. and R. T. Busing. 2000. Niche versus chance in tree diversity in forest gaps. *TREE* 15: 183-188.
- Bunge, M. 1969. Metaphysics, epistemology and methodology of levels. In: L. L. Whyte, A. G. Wilson and D. Wilson (eds.), *Hierarchical Structures*. American Elsevier, New York. pp. 17-28.
- Critchley, C. N. R. 2000. Ecological assessment of plant communities by reference to species traits and habitat preferences. *Biodiversity and Conservation* 9:87-100.
- Dale, M. B. 1976. Hierarchy and level: prolegomena to a cladistic classification Tech. Memo. 1, CSIRO Division of Tropical Crops and Pastures, St. Lucia, Brisbane.
- Dale, M. B. 1999. The dynamics of diversity: mixed strategy systems *Coenoses* 13:105-113.
- Dale, M. B. and D. J. Anderson. 1973. Inosculate analysis of vegetation data. *Austral. J. Bot.* 21: 253-276.
- Dale, M. B. and M. M. Barson. 1989. On the use of grammars in vegetation science *Vegetatio* 81: 79-94.
- Dale, M. B. and P. Hogeweg. 1998. The dynamics of diversity: a cellular automaton approach. *Coenoses* 13:3-15.
- Dale, M. B. and D. Walker. 1970. Information analysis of pollen diagrams. *Pollen et Spores* 2: 21-37.
- Diday, E. 1988. The symbolic approach in clustering and related methods of data analysis: the basic choices. In: H. H. Bock (ed.), *Classification and Related Methods of Data Analysis*, North Holland, Amsterdam. pp. 673-683.
- Edgoose, T. and L. Allison. 1999. MML Markov classification of sequential data. *Statistics and Computing* 9:269-278.
- Edwards, R. T. and D. Dowe. 1998. Single factor analysis in MML mixture modelling. *Lecture Notes in Art. Intell* 1394 Springer, pp. 96-109.
- Gatsuk, L. E., O. V. Smirnova, L. I. Vorontzova, L. B. Zaugolnova and L. Zhukova. 1980. Age states of plants of various growth forms: a review. *J. Ecol.* 68: 675-696.
- Hilderman, R. J. and H. J. Hamilton. 1999. Heuristics for ranking the interestingness of discovered knowledge. Proc. 3rd Pacific-Asia Conf. Knowledge Discovery PKDD'99, Beijing, Springer Verlag Berlin. pp. 204-209.
- Huisman, J., H. Olf, and L. F. M. Fresco. 1993. A hierarchical set of models for species response analysis. *J. Vegetation Science* 4, 37-46.
- Kodratoff, Y. 1986. *Leçons d'apprentissage symbolique*, Cepadued., Toulouse.
- Kullback, S. and R. A. Leibler. 1951. On information and sufficiency. *Ann. Math. Statist.* 22: 79-86.
- Legendre, P. and E. Gallagher. 2000. Ecologically meaningful transformations for ordination biplots of species data. *Ecology* (submitted).
- Mackay, D. M. 1969. Recognition and action. In: S. Watanabe (ed.), *Methodologies of Pattern Recognition*. Academic Press, London. pp. 409-416.
- Pazzani, M. J. and D. Kibler. 1992. The utility of knowledge in inductive learning *Machine Learning* 9: 57-94.
- Quinlan, R. and R. L. Rivest. 1989. Inferring decision trees using the Minimum Description Length Principle. *Information and Computation* 80:227-248.
- Rissanen, J. (1995) Stochastic complexity in learning. In: P. Vitányi (ed.), *Computational Learning Theory*, Lecture Notes in Computer Science, 904. Springer Verlag, Berlin. pp. 196-201.
- Stevens, W. L. 1937. Significance of grouping. *Ann. Eug. Lond.* 8:57-69.
- Wallace, C. S. 1995 Multiple factor analysis by MML estimation Tech Rep. 95/218, Dept. Computer Science, Monash University, Australia.
- Wallace C. S. 1998. Intrinsic classification of spatially-correlated data. *Comput. J.* 41: 602-611.
- Wallace, C. S. and D. M. Boulton. 1968. An information measure for classification. *Comput. J.* 11:185-195.
- Wallace, C. S. and D. L. Dowe. 2000. MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. *Statistics and Computing* 10: 73-83.
- Watanabe, S. 1969. *Knowing and Guessing*. J. Wiley, New York.
- Wildi, O. and M. Schütz. 2000. Reconstruction of a long-term recovery process from pasture to forest *Community Ecology* 1:25-32.
- Williams, W. T. and M. B. Dale. 1962. Partitioned correlation matrices for heterogenous quantitative data. *Nature* 196: 502.
- Williams, W. T., J. M. Lambert, and G. N. Lance. 1966. Multivariate methods in plant ecology. V. Similarity analysis and information analysis. *J. Ecol.* 54: 427-446.
- Williams, W. T., G. N. Lance, L. J. Webb, J. G. Tracey and M. B. Dale. 1969. Studies in the numerical analysis of complex rain-forest communities. III. The analysis of successional change. *J. Ecol.* 57: 513-535.