

Minimum message length clustering, environmental heterogeneity and the variable Poisson model

M. B. Dale

*Australian School of Environmental Studies, Griffith University, Nathan, Qld 4111, Australia.
Tel.: +61 7 3371 4414, Fax: +61 7 3870 5681, E-mail: m.dale@mailbox.gu.edu.au*

Keywords: Fuzzy clustering, Gaussian distribution, Mixture modelling, Pattern.

Abstract: One possible explanation of variation in vegetation is based on the variable Poisson model. In this model, species occurrence is presumed to follow a Poisson distribution, but the value of the Poisson parameter for any species varies from point to point, as a result of environmental variation. As an extreme, this includes dividing the given habitat into areas favourable to a community and areas which are unfavourable, or at least not occupied. The spatial area can then be viewed as a series of patches within which each species follows a Poisson distribution, although different patches may have different values for the Poisson parameter for any particular species.

In this paper, I use a method of fuzzy clustering (mixture modelling) based on the minimum message length principle to examine the variation in Poisson parameter of individual species. The method uses the difference between the message length for the null, 1-cluster case and the message length for the optimal cluster solution, appropriately normalised, as a measure of the amount of pattern any analysis captures. I also compare the Poisson results with results obtained by assuming the within patch distribution is Gaussian. The Poisson alternative consistently results in a greater capture of pattern than the Gaussian, but at the expense of a much larger number of clusters. Overall, the Gaussian alternative is strongly supported. Other mechanisms that might introduce extra clusters, for example within-cluster correlation or spatial dependency between observations, would presumably apply equally to both models. The variable Poisson model, in the limit, converges on the individualistic model of vegetation, the Gaussian on something like the community unit model. With these data, the individualistic model is strongly rejected. Difficulties with comparing model classes mean this conclusion must remain tentative.

Abbreviations: MML - Minimum Message Length, ptp - point-to-point.

Introduction

In his discussion of pattern in vegetation, Greig-Smith (1983, p. 80) notes two approaches to non-randomness, one due to point-to-point (ptp) variability in a Poisson parameter, the other due to some form of dependence between observations, termed contagion by Pólya (1930). He cites Feller (1943) as showing that these cannot be distinguished on the basis of observed frequency distributions, and comments that ptp variability has been little studied. Indeed, in contrast to the large number of references to studies of contagion, he cites only 4 concerned with point-to-point variability – Ashby (1935), Stevens (1937), Singh and Das (1938) and Erickson and Stehn (1945). The Poisson parameter is simply the mean (and variance) of a Poisson distribution and presumably variation in this parameter reflects environmental heterogeneity, with the probability of occurrence of a species varying with the harshness, or otherwise, of the habitat. It would

seem sensible, therefore, to attempt to cluster observational data so that, within clusters, the Poisson parameter is constant. In addition, it would also seem sensible to compare the use of a Poisson distribution with some other possible distribution, such as the Gaussian, to determine the extent to which the Poisson model is supported. This is the objective of this paper.

Method

Previous methods

For the four approaches noted above, the first three proceed via the number of empty squares $E = n(1-n^{-1})^s$ [1+s(s-1)c] expected when a quadrat is divided into n smaller squares and s individuals occur, assumed randomly distributed. The constant c is estimated to give best fit to the data and acts as an overall measure of Poisson parameter variability. Greig-Smith (1983) comments that

the applications tested contagion more than ptp variation, that it would be difficult to apply and that the meaning of c is difficult to interpret.

Erickson and Stehn (1945) are concerned only to divide the area into favourable and unfavourable parts for each species. They plotted $\log(x!y)$ against x , where x is some number of individuals and y is the number of quadrats containing x individuals. This, for random data, is a linear relationship, and they fitted a linear relationship to the larger values of x . This permits an estimation of the number of, and mean density within, samples assigned to the favourable area, from which the unfavourable elements can be determined by difference. This certainly concentrates attention on the variability of the Poisson parameter, but assumes a discontinuous pattern of variability and can be invalidated if contagion is also present. Since neither of these previous methods can be considered satisfactory, I have developed an alternative.

Outline of method

The overall notion is as follows. I first derive a number of clusters using all species data to define subsets of observed samples that are similar, the number of clusters being objectively determined. The clusters are defined assuming that all species have a specified distribution within a cluster. In the present case the distribution will be either Poisson or Gaussian. So, within these clusters, an estimate of the parameter(s) needed to define the distribution of each species is possible. With suitable spatially arranged observations, we can then examine the spatial distribution and variation of the Poisson or other parameters for each species. Although the clusters are presumed homogeneous, the parameter values for any particular species need not differ between clusters. This assumes that no single species dominates the cluster formation¹ since, if one did, then the clusters and the parameter values would be equivalent.

The clustering program used allows each species to be associated with a separate model, so mixtures of Poisson and Gaussian models are possible, if the user can justify such choice. It might be sensible to allow rarer species to be Poisson distributed and commoner ones Gaussian, but there is no obvious means of identifying a suitable threshold.

The MML clustering method

The method I use for clustering falls into the category of mixture modelling or model based clustering. For a general discussion of such methods, see Banfield and Raftery (1993), Bensmail et al. (1997) and Fraley and Raftery (1998). Here I have chosen to use Wallace and Dowe's (2000) Snob program, which is based on the minimum message length principle. This principle argues that a user should determine the quality of a model by calculating the length of an optimally encoded representation of the data, given the model. Overall the Snob program incorporates a Bayesian view coupled with information-theoretic ideas derived from Kolmogorov (1965) and combines a measure of model complexity with one of quality of fit to the observed data.

Using this principle has several advantages:

- It takes into account the precision with which the observed data were recorded to determine the precision with which the distribution parameters have to be estimated; coarse measurement means coarse estimates, precise measurement means precise estimates. However, this does mean that when comparing results we have to normalise the comparison in some way to adjust for varying precision.
- The clusters identified are fuzzy clusters to which things belong with some given probability. Fuzzy clustering is necessary if the parameters of the clusters are to be consistently estimated. In addition, the use of fuzzy clusters will usually reduce the length of the message, which is our criterion of model excellence.
- By establishing a trade-off between quality of fit to the observed data and complexity of the model the program can determine an optimal number of clusters (cf. Dale 1987). Complexity of model here means simply the number of clusters. Thus, the method provides a means for establishing the number of clusters, and incorporates a test against the null hypothesis of a single cluster.
- Snob allows the user to specify the required distribution within clusters and presently permits the use of Bernoulli (multistate), Poisson, Gaussian and von Mises (angular) distributions².

1 It would also be possible to adopt a 'leave one out' approach. This would use all species except a nominated one to form clusters, then estimate the parameters for the omitted species, which is rather tedious if a large number of species are examined. In this case, the model for the selected species need not be the same as the model used for the remaining species.

2 A recent development (Dowe pers. commun.) also allows a t-distribution to be chosen. This provides a 'thick-tailed' distribution. I hope to examine this possibility later.

Assessment of results

In order to assess the analysis overall I use the difference in message length between the 1-cluster and the optimal n -cluster solution as a measure of the amount of pattern captured by the clustering. To avoid problems with differences in precision between analyses, I normalise this difference using the 1-cluster message length to provide a suitable index of pattern captured.

It should be noted that the difference in value of the 2 message lengths is intimately associated with the *a posteriori* probability of the model. Specifically, the exponential of the difference represents the odds in favour of the model with the shorter length. Finally, it is possible that the 1-cluster null hypothesis can be accepted, if no other solution with smaller message length is found.

Data and analyses

The data were originally collected by Goodall (1953) and consist of 256 samples arranged in a stratified random sample over an area of Mallee vegetation growing on dunes; *Eucalyptus oleosa* dominated the ridges, *Eucalyptus dumosa* the hollows. An area of 640 x 640 m² was divided into 64 grid squares of 10 x 10 m². Four 5 x 5 m² randomly selected samples were then taken from these. 61 attributes were used, mostly species although some growth forms were distinguished and dead material of *Triodia irritans* was also recorded. The number of species was reduced to 32 in some analyses by rejecting very rare species.

Percentage cover was used as a performance measure, measured using point quadrats to 0.05%. Since this represents extremely precise recording, an analysis was also carried out with the precision reduced to 1%. The Snob program uses the precision value together with fit to the data to determine an optimal coding for the parameter estimates.

To find the optimal clustering solution, several searches were made, starting from different initial estimates of the number of clusters and from differing random initial allocation of samples to clusters. In most cases, these converged to the same, or very similar, MML solutions.

In total, five analyses were made. These involved using Gaussian and Poisson within-cluster models with both the full and the reduced data. The extra analysis was of the full data using the Poisson model but with the precision reduced as noted earlier. No substantive interpretation is attempted here, since the main objectives are concerned with differences in model within clusters.

Results

Snob output

The results from a Snob clustering include the following information:

- The message length for the 1-cluster (null) case. This in effect is the message length if all patterns in the data are ignored
- The message length for the optimal number of clusters and a statement of that number n .
- Information on the association between attributes and clusters including the appropriate distributional parameters for each attribute and whether these are significantly different from the population parameter values.
- The assignment of the samples proportionally to the various clusters. This will reflect the degree of overlap between clusters, which is recorded as the probability that the sample belongs to a specific cluster.
- For each sample, the message length required to code it, conditional on the estimated parameters and cluster assignments. This information is useful for detecting outliers in clusters but is not important here.

Comparison of Gaussian and Poisson descriptions

The general features of the analyses are presented in Table 1. Several things are immediately obvious. The Poisson models all have considerably larger message lengths than the Gaussian solutions and consistently identify an optimal solution with considerably more clusters, in fact about 3 times as many. The odds in favour of the Gaussian solution are very large indeed, being e^{227248} :1! The choice of within-cluster model overrides any differences in number of species and in precision of measurement, and the differences in message length indicate that the observed differences are certainly significant. Reducing the precision produces a longer message and thus a less desirable result. Reducing the number of species decreases the message length, as might be expected.

While the Poisson data have a higher total message length, the analyses under this model clearly capture much more structure. With a large number of small clusters (the largest has only 11 members), the fit to the data can be extremely good. In contrast, the Gaussian clusters are larger, ranging in size from 4 to 64, and, individually, fit less well. I should add that this result is at the high end of percentages when compared with other studies of vege-

Table 1. Message length information. Message lengths are in nits. The normalised effectiveness index = difference between 1-class and n -class lengths / 1-class length.

Analysis	1-class message length	Number of clusters	n -class message length	Difference in message lengths	Normalised Effectiveness index	Number of samples ambiguously assigned
Gaussian	45,550.4	16	20,652.2	24,898.2	0.55	15
Red. Gaussian	33,631.2	18	16,457.8	17,173.4	0.51	7
Poisson High Precision	301,545.1	50	49,299.1	252,246.0	0.84	1
Poisson Low Precision	327,957.2	51	56,420.5	271,536.7	0.83	0
Red. Poisson	291,310.6	51	44,439.7	246,870.9	0.85	0

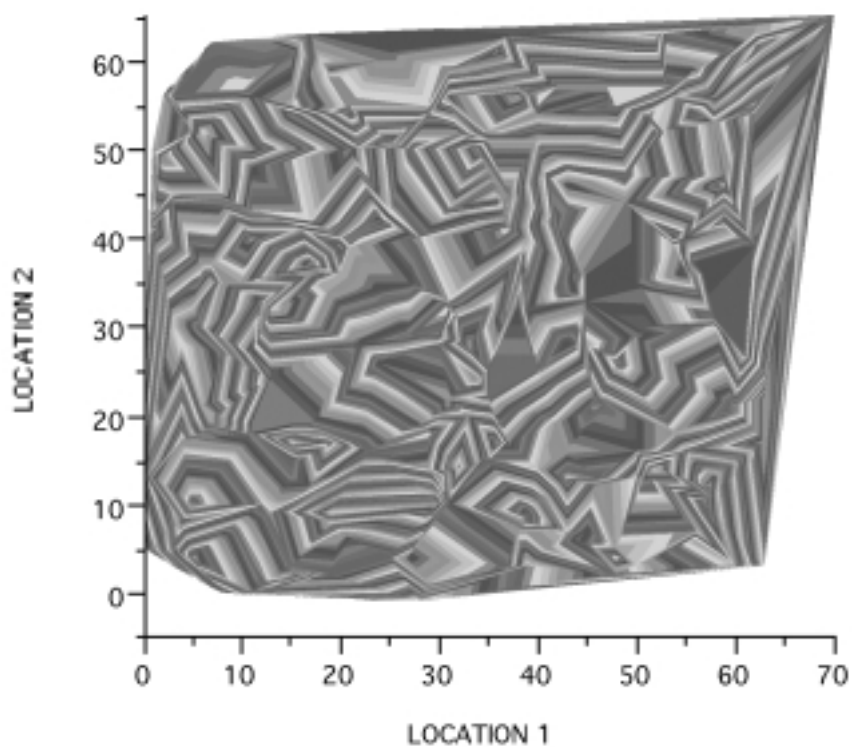


Figure 1. Mallee Poisson clusters: spatial distribution. This map illustrates the fragmentary nature of the Poisson clusters and is obtained by “contouring” the cluster labels.

tation data I have attempted. In fact, 20-30% pattern capture is a more common result; contingent variation is very large in vegetation data (Brokaw and Busing 2000).

The question is whether the intricate pattern of the Poisson clusters when spatially arranged (Fig. 1) is a result of environmental heterogeneity or a function of the plants themselves (see Boerlijst and Hogeweg [1991] for possibilities of autopoietic pattern generation). Some of the variation is certainly environmental, for Goodall notes the existence of sandy ridges up to 10 m high running east-west. This dune-swale system can be easily identified in the Gaussian solution (Fig. 2). In contrast to the large number of clusters of the Poisson analysis, the Gaussian analysis is more frugal, although the number of groups may still seem somewhat large.

They are less obvious in the Poisson result partly because of the way the maps were prepared but also because of the greater fragmentation due to the large number of clusters formed. But these ridges and hollows are associated with different dominants, if dominance is a suitable concept when total cover is only around 50%. *Eucalyptus dumosa* and *Triodia irritans* are important on the ridges, *Eucalyptus oleosa* and some scattered chenopods in the hollows.

Individual species parameter distributions

Figures 3, 4 and 5 show the spatial variation in the Poisson rate parameter for 3 species, *Eucalyptus dumosa*, *Eucalyptus oleosa* and dead *Triodia irritans*. In all cases, it is clear that there are large patches with similar Poisson rates and that these patches are spatially larger than the cluster patches. This suggests that the cluster patches are not a reflection of these particular species but do represent the combined effect of all species.

In all three results, the linear dune-swale pattern is distinguishable though sometimes with difficulty. Overall, the spatial distribution of Poisson rates suggests that the clusters formed from all species subdivide the area very finely. Individual species have patches with a smoother response bounded by areas of more rapid change, but these patches seem larger than those indicated when all species are taken into account.

Approximation by reduced data

Comparison of reduced and full data show correlations between results whether Gaussian or Poisson models are used, although the Gaussian correlation is somewhat lower. Using the reduced data clusters to predict the

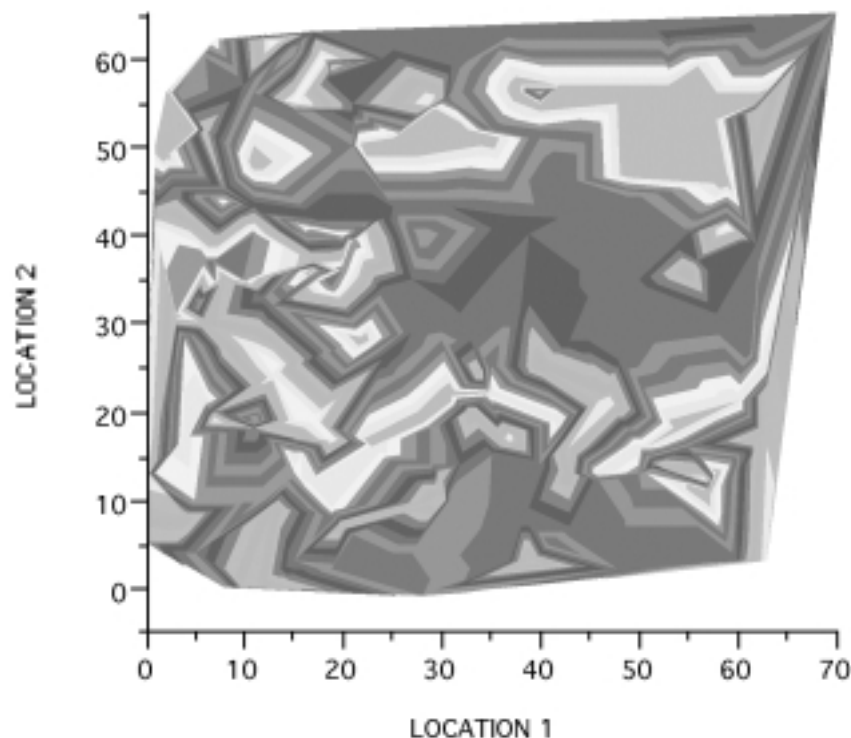


Figure 2. Mallee Gaussian clusters: spatial distribution.

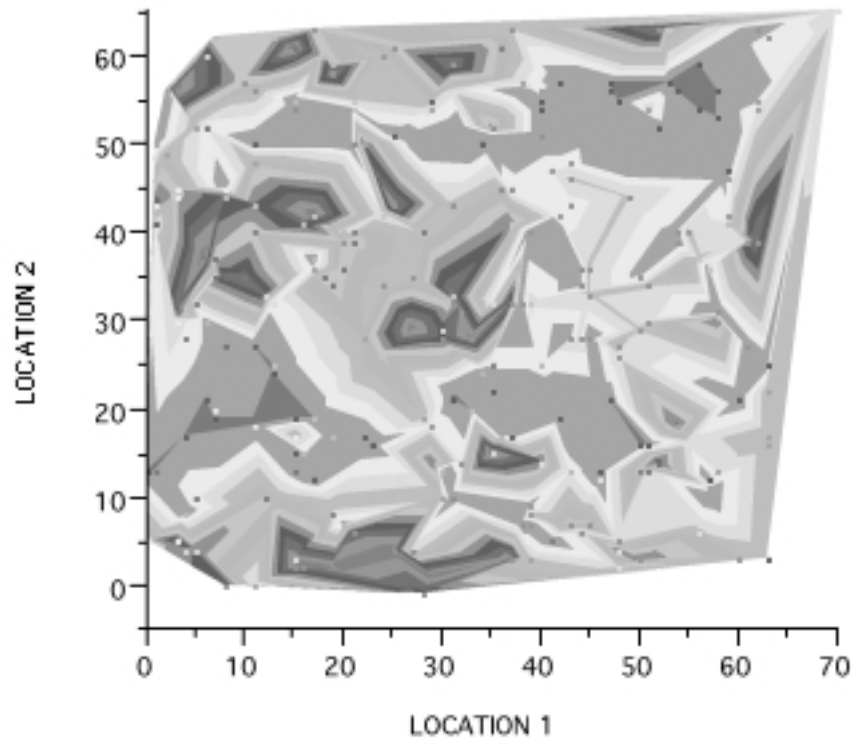


Figure 3. *Eucalyptus dumosa*: contour map of the Poisson rate parameter.

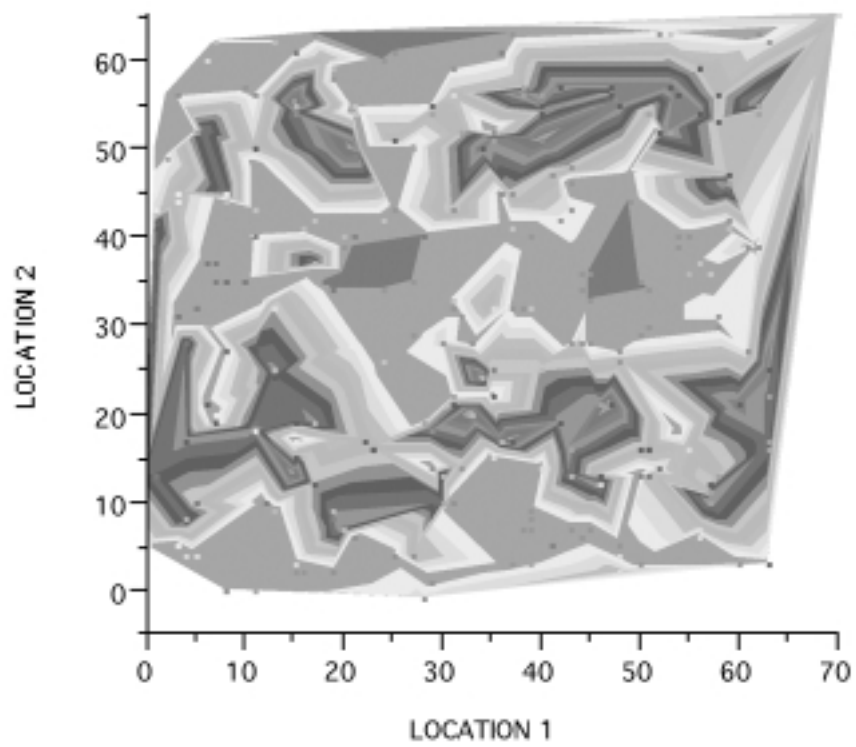


Figure 4. *Eucalyptus oleosa*: contour map of Poisson parameter.

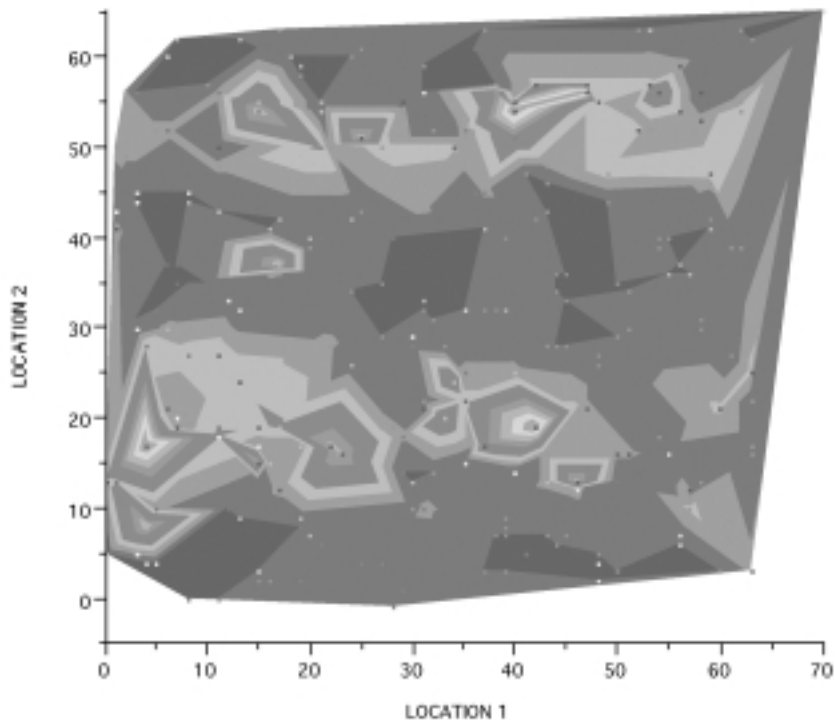


Figure 5. *Triodia irritans* (dead): contour map of Poisson rates.

full data clusters gives an $R^2 = 0.79$ for the Poisson model and $R^2 = 0.61$ for the Gaussian.

However, the differences in message length show that there are significant differences between these results and that the contributions of the rarer species are not insignificant. Almost all species show differences from the population Poisson rate (mean and variance), though in many cases this reflects absence in most of the area with the species concentrated in just a few samples.

Attribute significance

There are marked differences between the Poisson and Gaussian solutions in how the attributes reflect the cluster differentiation. In the Poisson analysis, the commoner species show significant differences for almost all clusters. The rare species also make their contribution to the distinction of clusters. Many of these differences are based on absence; that is the species is of lower mean abundance in the cluster than in the population in general. In contrast, in the Gaussian solution the commoner species each distinguish only a few clusters, and the rare species contribute much less to differentiation. Still, only 4 species in the Gaussian analysis show uniformity across the entire area and all are very rare.

Fuzziness of assignment to clusters

The amount of ambiguity in assignment to clusters is very small, with the Gaussian solutions worse (Table 1). The Poisson assignments are almost entirely completely crisp. However, in almost all cases only one other cluster had any reasonable probability (there is a single case where 2 other clusters were involved). The probabilities of assignment to clusters other than the most likely were all relatively small; the largest being a probability of 0.18. If there had been serious ambiguity then drawing maps of cluster distributions would become problematic, but in the present case no serious difficulties arise.

Examination of the message lengths for each sample (not shown here) does not suggest that there are any outliers, except in a single case for the complete Gaussian analysis. This means that there are very few, if any, small clusters which have been merged into other clusters by the program search procedure.

Discussion

Methodological assessment

The strategy of using all species to develop clusters seems to have worked well. The clusters found are in gen-

eral very small for the Poisson model, and the individual species form patches of relatively constant Poisson parameter which are larger than the clusters identified. In addition, the discrimination of the clusters by the attributes shows that almost all make some contribution. The Gaussian solution seems to rest more on the commoner species and perhaps reflects dominance characteristics. This difference is reminiscent of the contrast of floristic (presence) and dominance (abundance) approaches to vegetation classification; on these results dominance is a clear winner. Historically in vegetation classification the Braun-Blanquet approach (cf. Westhoff and van der Maarel 1973), which emphasises floristic presence, has been more widely adopted than the Scandinavian system of Du Rietz (cf. Trass and Malmer 1973) which more strongly emphasised dominance. It seems that for these data this was an error!

Another feature of note is that, in selecting the optimal number of clusters, the 1-cluster null model has no special status, a stance which is at variance with the views of, say, Simberloff (1980), Wilson (1991) and Keddy (1993). These authors have argued for an approach based on falsification of simple clear null hypotheses. But even if such hypotheses are available in the complexity of ecosystems, they do not identify which alternative hypothesis should be adopted when the null hypothesis is falsified. We are told *what is not*, but *not what is*!

Instead, the problem being addressed here is to estimate the optimal number of clusters within a class of models parameterised by that number and there is an implicit falsification of many models within a class. We could also consider another class of models, such as an ordination model, whose parameter is the optimal number of orthogonal axes³.

In fact, we might go further, for we could attempt to identify the appropriate class of models as well. Both the minimum message length principle and the minimum description length principle (Rissanen 1999) permit such comparisons, the latter being expressly developed to do just that. So, contrary to Shipley and Keddy (1987), it is possible to compare cluster representation with axis representation and select a 'best' result.

Whether such a comparison is of much use is a different matter. That it might be done is perhaps worth knowing. That it should be done is a very different question - and one based in deontic logics that notoriously have no

truth value. It is likely, for example, that scale effects, due to changes in the area of the primary sample unit or successional development, could lead to the appropriate family of models varying. Neither clusters nor axes might be universally preferable and some combination of the two models could well be preferable to either alone (cf. Edwards and Dowe 1998).

In any case, the use to be made of such structure as is found may necessitate the choice of one or the other model class irrespective of their optimality. If all patterns are ultimately for an agent, as MacKay (1969) argued, then interest and 'actionability' may be paramount. Evaluating these properties is a challenging task, though some suggestions already exist in the literature (Barsalou 1995, Hilderman and Hamilton 1999).

Methodological assumptions, constraints and extensions

The Snob program makes some assumptions and we need to consider what impact these may have had. The present program assumes that all attributes are uncorrelated within clusters, which may not be true. Edwards and Dowe (1998) have examined modifications to Snob which permit within-cluster correlations. They note that the effect of the existence of such correlation is to increase the number of clusters. Snob further assumes that the various samples being clusters are independent, i.e., that spatial or temporal dependence is absent. Wallace (1998) and Edgoose and Allison (1999) have considered how this constraint might be overcome in space and time respectively. Again an increase in number of clusters is the result. However, these possibilities would apply equally to both Poisson and Gaussian models and it seems unlikely that they would be responsible for the very large number of clusters identified in the Poisson results obtained here.

The program deletes clusters with less than 4 members, re-assigning the members to other clusters, since parameter estimates are then very unreliable. This will tend to reduce the number of clusters. Overall it seems likely that there has been some overestimation of the number of clusters. I hope to examine this question when suitable programs become available.

The next problem is more difficult although a solution should be attainable. To be completely Bayesian, prior probabilities should be assigned to the class of models based on Poisson and Gaussian distributions. I have assumed here that this probability would be equal for both,

3 Wallace (1995) notes that the 'real' axes are very unlikely to be orthogonal. Presumably some form of rotation is desirable after the orthogonal dimensions have been identified. However, it is possible that the number of oblique axes may exceed the (orthogonal) dimensionality of the space, which makes any estimation difficult.

though other views may differ. I suspect that I should further take account of the fact that for the Poisson solution there is only a single parameter for each cluster, whereas for the Gaussian there are two (mean and variance). Thus, the Gaussian model is more complex than the Poisson and some penalty should be paid for this. I have included no such penalty.

I have assumed that all clusters have the same internal distribution, that is either all Poisson or all Gaussian, although this is not a necessary assumption. Banfield and Raftery (1993; see also Stanford and Raftery 1997) have considered models involving principal curves (Hastie and Stuetzle 1989) and Poisson distributions in order to recover linear manifolds in noisy data. However, I have no justification in the present study for using such complex models.

Practically, the major problem concerns the effectiveness of the search for the optimal model. Snob employs an expectation-maximisation algorithm but it is possible that other algorithms might be more effective, such as simulated annealing. For the present study, the search procedure seems to be acceptably effective.

The variable Poisson model

The basis of the variable Poisson model is that the Poisson parameter for any species varies from place to place, in contrast to the use of other models expressing contagion. Certainly, the results obtained using the Poisson model within clusters suggest that the variable Poisson model is capable of describing the variation in abundance extremely well, and to that extent is justified. However, the most surprising result is the contrast between Poisson model's effective pattern capture and the fact that it is markedly suboptimal in terms of message length. The identification of a large number of clusters means that fit to the data can be very good, but of course it increases the complexity of the description greatly.

In comparison, the Gaussian distribution leaves a greater part of the variation to contingency but is decidedly more parsimonious in defining clusters. The results obtained here suggest that a Poisson model is not a good model of species distribution compared to the Gaussian. There may of course be even better models for within-cluster variation. Kemp and Kemp (1956), following Robinson (1954), have suggested that locally cover percentage may follow a Beta distribution. For neither distribution is the degree of ambiguity of assignment of a large magnitude. However, I would argue that the procedure used here does suffice to examine the variable Poisson model for individual species, if that is the intent of the

investigation. This would be the case if we were interested more in autecological characteristics or niche definition for individual species.

Overall, to accept the variable Poisson model would seem to require a very large prior commitment to this model. In the limit, of course, the variable Poisson model would lead to an acceptance of continuous variation in the Poisson parameter, and an individualistic interpretation. Indeed, if the Poisson model is accepted, then this would be a natural interpretation. The Gaussian model, in contrast, isolates large patches of relatively homogeneous vegetation, which is closer to, but not necessarily identical with, the community-unit model. The results obtained here show that the Poisson model is much less acceptable, although this conclusion must remain tentative because of the problems of model class comparison which have been ignored.

References

- Ashby, E. 1935. The quantitative analysis of vegetation. *Ann. Bot.* 49: 779-802.
- Banfield, J. D. and A. E. Raftery 1993. Model-based Gaussian and non-Gaussian clustering. *Biometrics* 49:803-821.
- 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.
- Bensmail, H., G. Celeux, A. E. Raftery and C. P. Robert. 1997. Inference in model-based cluster analysis. *Statistics and Computing* 7:1-10.
- Boerlijst, M. and P. Hogeweg. 1991. Spiral wave structure in prebiotic evolution: hypercycles stable against parasites. *Physica D* 48: 17-28.
- Brokaw, N. and R. T. Busing. 2000. Niche versus chance in tree diversity in forest gaps. *TREE* 15: 183-188.
- Dale, M. B. 1987. Knowing when to stop: cluster concept-concept cluster. *Coenoses* 3: 11-32.
- 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 Artificial Intelligence 1394* Springer Verlag. pp. 96-109.
- Erickson, R. O. and J. R. Stehn. 1945. A technique for analysis of population density data. *Amer. midl. Nat.* 33:781-787.
- Feller, W. 1943. On a general class of 'contagious' distributions. *Ann. Math. Statist.* 14:389-400.
- Fraley C. and A. E. Raftery 1998. How many clusters? Which clustering method? - Answers via Model-Based Cluster Analysis. Technical Report no. 329, Department of Statistics, University of Washington.
- Goodall, D. W. 1953. Objective methods for the classification of vegetation 1. The use of positive interspecific correlation. *Austral. J. Bot.* 1: 39-63.
- Greig-Smith, P. 1983. *Quantitative Plant Ecology*, 3rd Edition, Blackwell, Oxford.
- Hastie, T. and W. Stuetzle. 1989. Principal curves. *Amer. Statist. Assoc. J.* 84: 502-516.

- Hilderman, R. J. & Hamilton, H. J. 1999. Heuristics for ranking the interestingness of discovered knowledge. Proc. 3rd Pacific-Asia Conf. Knowledge Discovery PKDD'99, Beijing, Springer, Berlin. pp. 204-209.
- Keddy, P. A. 1993. Do ecological communities exist? A reply to Bastow Wilson. *J. Veg. Sci.* 4: 135-136.
- Kemp, C. D. and A. W. Kemp. 1956 The analysis of point quadrat data. *Austral. J. Bot.* 4:167-174.
- Kolmogorov, A. N. 1965. Three approaches to the quantitative description of information. *Prob. Inform. Transmission* 1: 4-7. (translation).
- Mackay 1969. Recognition and action. In: S. Watanabe (ed.), *Methodologies of Pattern Recognition*, Academic Press, London. pp. 409-416.
- Pólya, G. 1930. Sur quelques points de la théorie des probabilités. *Ann. Inst. Poincaré* 1: 117-161.
- Rissanen, J. 1999. Hypothesis selection and testing by the MDL principle. *Comput. J.* 42:260-269.
- Robinson, P. 1954. The distribution of plant populations. *Ann. Bot.* 19:59-66.
- Shipley, B. and P. A. Keddy. 1987. The individualistic and community-unit concepts as falsifiable hypotheses. *Vegetatio* 69: 47-55.
- Simberloff, D. 1980. A succession of paradigms in ecology: Essentialism to materialism and probabilism *Synthese* 43:3-29.
- Singh, B. N. and K. Das. 1938. Distribution of weed species on arable land *J. Ecol.* 26: 455-466.
- Stanford, D. and A. E. Raftery. 1997. Principal curve clustering with noise. Tech. Rep. 317, Dept. Statistics, University of Washington.
- Stevens, W. L. 1937. Significance of grouping. *Ann. Eug. London.* 8: 57-69.
- Trass, H. and N. Malmer. 1973. North European approaches to classification. In: R. H. Whittaker (ed.), *Classification and Ordination of Plant Communities*, Dr. W Junk, The Hague. pp.529-575.
- Wallace, C. S. 1995. Multiple factor analysis by MML estimation. Tech. Rep. 95/218, Dept Computer Science, Monash University, Clayton, Victoria 3168, Australia.
- Wallace C. S. 1998. Intrinsic classification of spatially-correlated data. *Comput. J.* 41: 602-611.
- 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.
- Westhoff, V. and E. vand der Maarel 1973. The Braun-Blanquet approach. In: R. H. Whittaker (ed.), *Classification and Ordination of Plant Communities*, Dr. W. Junk, The Hague. pp. 617-707.
- Wilson, J. B. 1991. Does vegetation science exist? *J. Veg. Sci.* 2:289-290.