MULTIVARIATE CLUSTERING FOR OBJECTIVE CLASSIFICATION OF VEGETATION DATA[1]

Ken Aho[2]

Abstract. Objective classifications of pre-mine vegetation provide unbiased standards for mine revegetation, and therefore defensible bond release criteria. In this paper I present a summary of multivariate classification techniques and briefly explain several fundamental algorithms. In addition I discuss three items imperative for effective classification. These are: 1) an appropriate and effective dissimilarity/distance measure, 2) an effective classification methodology, and 3) an objective way to decide on an appropriate/optimum number of clusters. I address these considerations by testing distance measures, evaluating classification methods, and finding appropriate pruning levels for classifications of alpine and pre-mine steppe vegetation data from the Northern Rocky Mountain region.

1) With respect to distance measures; Bray-Curtis and Jaccard dissimilarities were the most effective as determined by methods which compared environmental gradients with phytosociological distance.
2) With respect to classification typologies; eight commonly used methods were compared with nine classification evaluator algorithms. Three hierarchical agglomerative methods: flexible beta ($\beta = -0.25$), average linkage, and Ward's linkage generally outperformed other methods across a large number of clustering solutions.
3) With respect to appropriate number of clusters; classification evaluators identical to those used for comparing classification typologies were adapted for straightforward graphical presentation of optimal classification solutions.

---

[1] Paper was presented at the 2006 Billings Land Reclamation Symposium, June 4-8, 2006, Billings MT and jointly published by BLRS and ASMR, R.I. Barnhisel (ed.) 3134 Montavesta Rd., Lexington, KY 40502.

[2]Ken A. Aho, PhD candidate, Department of Ecology, Montana State University, Bozeman MT 59715, email: kaho@montana.edu.

Proceedings America Society of Mining and Reclamation, 2006 pp 1-24
DOI: 10.21000/JASMR06010001
http://dx.doi.org/10.21000/JASMR06010001

Introduction

Identification of vegetation patterns (e.g. communities) is vital to effective land reclamation. This is true since recognition and description of vegetation patterns provides both objectives for reclamation, and potential information about vegetation/environmental relationships important to reclamation. Objective vegetation classifications are particularly important for mine reclamation. Such community descriptions can provide standards which regulators can use as criteria for revegetation success and bond release.

There have been both non-quantitative and quantitative attempts to find patterns in vegetation data. One classic non-quantitative approach is relevé table-sorting (Braun-Blanquet, 1964). Using this method one subjectively locates clusters by moving columns and rows of vegetation data matrices. A second method involves designating communities in the field based on subjective interpretation of vegetation physiognomy and species composition.

Obviously categorization using non-quantitative methods relies largely on the skills and expertise of the categorizer and quality of results may vary greatly. Furthermore, it should be apparent that subjective classifications are not reproducible (perhaps even by the original categorizer). It should be emphasized that non-quantitative approaches may be the best option in some circumstances; for instance when categorizing combined datasets which were created using entirely different sampling methods (see Weaver and Aho, 2006)

Quantitative categorization (i.e. cluster analysis) encompasses a large number of mathematical methods which allow objective classification of vegetation data. It can be argued that I use the term "objective" somewhat loosely since cluster analysis models each have their own set of assumptions and constructs (I will show that some model constructs may be more valid than others). Nonetheless, it cannot be argued that quantitative methods 1) allow far more objectivity than non-quantitative techniques, and 2) provide reproducible and defensible results.

Cluster analysis considerations

Three considerations are necessary for effective cluster analysis. First, an appropriate distance/dissimilarity measures must be chosen to describe variability in the data. Second an effective classification strategy (hierarchical or non-hierarchical) and linkage methodology (see explanation below) must be chosen. Third, an appropriate number of clusters must be decided on.

1. Deciding on an appropriate distance measure. Before addressing this topic an explanation of distance measures may be helpful. The simplest quantitative distance measure is generally acknowledged to be Euclidean distance. This algorithm is shown as Equation 1 for sites *i* and *h*, where a_{ij} = abundance of species *j* (*j* = 1, 2,..*p*) in site *i*. The calculation of Euclidean distance matrix for three sites using two species is described in Fig. 1.

$$ED_{i,h} = \sqrt{\sum_{j=1}^{p} (a_{ij} - a_{hj})^2} \qquad \textbf{(1)}$$

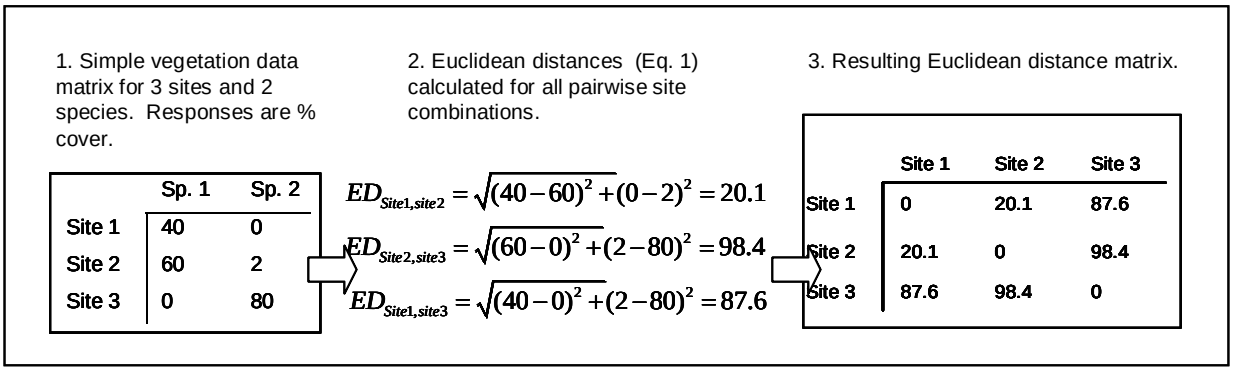

Figure 1. Procedure for calculating a Euclidean distance matrix comparing three hypothetical sites.

To use cluster analysis one must decide on a distance measure to represent variability in the vegetation data (possible exceptions include k-means analysis and TWINSPAN which use Euclidean and chi-squared distance respectively as an implicit part of their algorithm). A number of ecologists have evaluated the performance of distance measures by comparing environmental distance (i.e. differences along gradients) to distance in ordination space (Beals, 1984; Faith et al., 1987; De'ath, 1999). If species distributions are well described by environmental variables, then a strong association should exist between distances in species space and differences in environmental space. Thus, poor correlations represent poor performance by the distance measure. Obviously this analysis becomes more meaningful as the number of important environmental variables increases in ones measure of environmental distance.

2. Comparing classification efficacy methods using internal tests. Clustering methods can be broken into two general strategies: hierarchical and non-hierarchical. Hierarchical methods find groups nested hierarchically within other groups. Non-hierarchical methods group data in discrete clusters without regard to hierarchical structure. Aside from major differences between hierarchical and non-hierarchical constructs, classification methods differ mainly with respect to their method of linkage (Lance and Williams, 1967). Linkage is the way classifications define distance between clusters (i.e. average, median, single, complete, centroid etc.). For further explanation see Kaufman and Rousseeuw (1990), and McCune and Grace (2002). The linkage of three hierarchical methods (single, complete and average linkage) are compared in Fig. 2.

Classification methods can be compared with external or internal tests (Gauch and Whittaker, 1981; Dale, 1991). External tests compare the results of a classification with previously established standards. Examples include recovery of clusters embedded within simulated datasets (Milligan and Cooper, 1985; Belbin and McDonald, 1993; Hirano et al., 2002), or recovery of subjective *a priori* clusters from ecological data (Cao et al., 1997). In contrast, internal tests use the characteristics of the classifications themselves to gauge effectiveness. Examples include cophenetic correlation (Sneath and Sokal, 1973), and maximization of between cluster variance (Orloci, 1967). Since *a priori* "correct" cluster solutions are not available in non-synthetic data, I will use internal evaluators to compare classification efficacy in this report.

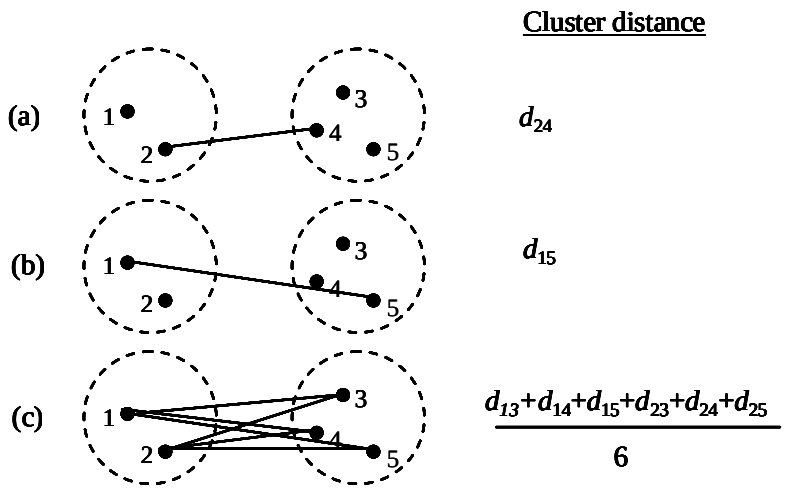

Figure 2. Distance between two clusters for three hierarchical methods (a) single linkage, (b) complete linkage, and (c) average linkage. Single linkage uses the distance of the two closest objects between the clusters to represent intercluster distance. Complete linkage uses the two most distant objects. Average calculates the average distance. Adapted from Johnson and Wichern (1998).

3. Deciding on an appropriate number of clusters "pruning analysis". Classification results using hierarchical methods can be presented in a tree-format (Fig. 3). This diagram presents an overview of the hierarchy and relatedness of sites, but gives only limited information about how many "real" clusters may exist. Lines drawn on Fig.3 indicate where the tree would need to be "pruned" for solutions with 3, 5, and 11 clusters. It should be obvious that any other clustering solution from 1 to N clusters (where N = number of classified objects) is also possible.

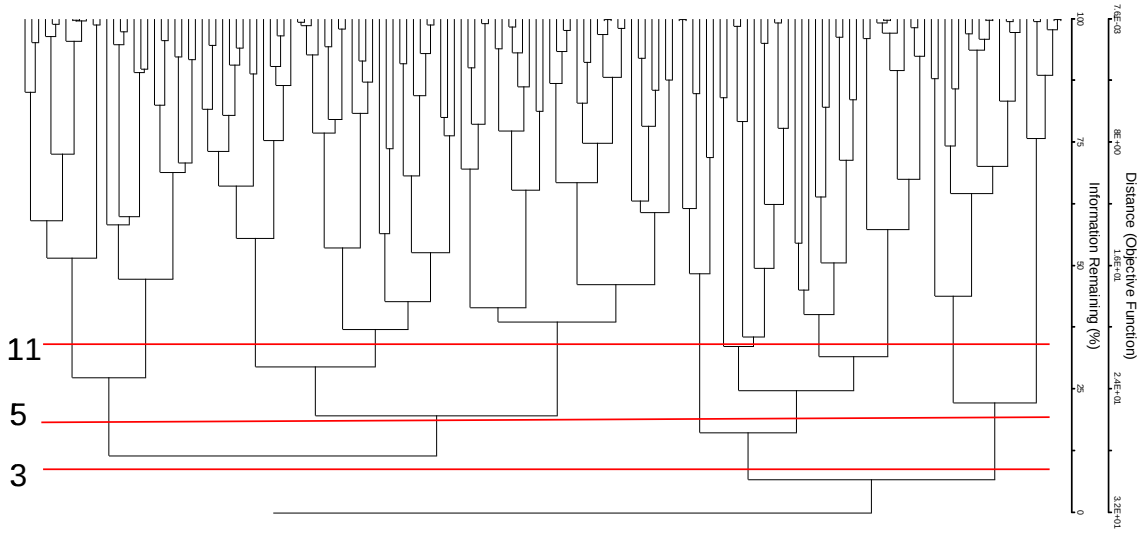

Figure 3. Classification with 3, 5 and 11 cluster solutions indicated on the hierarchical tree.

A large number of methods have been proposed as stopping/pruning criteria for cluster analyses (Dale, 1991; Milligan and Cooper, 1985). Since internal tests (discussed in the section above) measure the intrinsic effectiveness of classifications, they may also be used as pruning

criteria, by testing the effectiveness of individual pruning level/clustering solutions (Milligan and Cooper, 1985).

Goals

I addressed the three considerations listed above by 1) testing the performance of distance measures, 2) evaluating the effectiveness of different classification methods, and 3) finding appropriate pruning levels for classification solutions. I did this by classifying and evaluating two distinct vegetation datasets from the Northern Rocky Mountain region of the United States.

Methods

Datasets

Two vegetation datasets were used for this study. The first consisted of alpine data from three high altitude ranges in the Northern Rocky Mountains (the Washburn, Absaroka and Beartooth Ranges), and was provided by the author's dissertation work. This dataset contained 178 plots and 180 species. A second dataset was acquired for a prairie steppe ecosystem from Southeast Montana which contained 113 plots and 173 species. Quantitative responses for both datasets were ocular estimates of species cover within transects.

Data analysis

1. Distance comparisons. Six commonly used distance measures: Euclidean, Bray-Curtis, Canberra, Jaccard, Kulczynski, and Manhattan were compared. Distance measures were evaluated by comparing sociological distance calculated by each measure to environmental distance. Environmental distance was calculated using two steps.

First environmental variables were standardized using Equation 2 so that they would not unequally weight environmental distance.

$$STx_i = \frac{x_i - \mu}{\sigma} \quad \textbf{(2)}$$

Where $ST(x_i)$ = the standardized value for response i, $(i = 1,2,....,n)$, from variable x $(x = 1,2,...,z)$

Second, the sum of absolute values of the differences between environmental responses at all sites was calculated (*ENVDIST*; Equation 3).

$$ENVDIST = \sum_{x=1}^{z} \sum_{\substack{i=1 \\ j \neq i}}^{n} \left| STx_i - STx_j \right| \quad \textbf{(3)}$$

The association between the vector *ENVDIST* and dissimilarity matrices resulting form the six distance measures was measured with Spearman's rank tests.

Due to the strength of environmental data from the Washburn Range, only this data (40 plots, 80 species) was used for distance measure comparisons. Sixteen environmental variables were used to describe environmental distance. These were: soil %N, soil %C, soil P (mg/kg), elevation (m), slope(degrees), aspect (degrees from north), annual solar radiation (MJ cm^2 yr^{-1}), %rock cover, %soil cover, pH, conductivity (mmhos cm^{-1}), %sand, %silt, %clay, number of days

soil were warmer than 10˚C at 10cm depth, and number of days soils were wetter than -0.12 MPa at 10cm depth.

2. Classification comparisons. Eight classification methods commonly used by vegetation ecologists were evaluated. They consisted of five hierarchical agglomerative methods: flexible β linkage (Lance and Williams, 1967), Ward's method (Ward, 1963), complete linkage (McQuitty, 1966), average linkage (Sokal and Michener, 1958), and single linkage (Sokal and Sneath, 1963), a hierarchical divisive method: TWINSPAN (Hill, 1979), and two non-hierarchical methods: partitioning around medoids (PAM; Kauffman and Rousseeuw, 1990), and k-means analysis (Hartigan and Wong, 1979). The value $\beta = -0.25$ was used for flexible β linkage as recommended by McCune and Grace (2002). TWINSPAN was performed using default parameters from PC-ORD (McCune and Mefford, 1999). Random starting points were used as initial cluster centers for k-means analysis.

Classification methods were compared using nine internal classification evaluators, most of which have been well tested and reviewed in the literature. The evaluators are summarized in Table 1. For evaluator equations see Aho (2006), or Aho et al. (2006).

Both datasets were classified with each of the 8 classification methods, and classifications were pruned to find their 20 simplest solutions (2 to 21 clusters). Each of these solutions was evaluated by each of the nine classification evaluators to compare solution efficacy. For both datasets, median values across the 20 solutions, confidence intervals for medians, and maximum values (best solution) across the 20 solutions were determined for each evaluator assessment of each method. Medians were used as estimates of evaluator central tendency since distributions of evaluator responses were generally non-normal. Confidence intervals for medians were calculated from the interquartile range, and are based on the asymptotic normality of the median for roughly equal sample sizes for two medians being compared (McGill et al., 1978, pg. 16).

Differences among methods with respect to median evaluator response were determined by Mann Whitney non-parametric pairwise tests. Significance levels were adjusted for simultaneous inference using Dunn's (1964) procedure. While Kruskal-Wallis tests were run on raw (non-standardized responses), standardized values are presented in the results to allow better comparability among evaluators. Responses were standardized using Equation 2.

3. Pruning analysis. The classification evaluators listed in Table 1 were also used to find the best clustering solutions from the best overall methodologies (the best overall methodologies being determined from the classification comparison step described above). Comparisons were made in two ways: 1) responses of standardized evaluator responses were compared across cluster levels (i.e. 2 clusters, 3 clusters,...., 21 clusters); and 2) residuals from standardized evaluators responses from fits to linear models were compared across cluster levels (Equation 4).

$$e_i = Y_i - \hat{Y}_i \qquad \textbf{(4)}$$

Where Yi are observed values and $\hat{Y}_i$ are fitted values from a linear regression model.

Responses were standardized using Equation 2. The second comparison method was used to test and compensate for linear artifact in evaluator responses across cluster levels (i.e evaluators whose responses always increase/decrease with number of clusters). An example of this is demonstrated in Fig. 4.

Table 1. Summary of classification evaluators used in this paper. For evaluator equations see Aho (2006), or Aho et al. (2006).

Evaluator	Optimality criteria	Comments [reference]*
Average Silhouette Width (ASW) (ASW, Rousseeuw 1987)	Low within-cluster dissimilarity and high dissimilarity of samples to nearest neighbor cluster.	Use of nearest neighbor cluster provides "local" criterion. Useful for seeking compact, widely separated clusters [1]
Partition Analysis ratio (PARTANA) (Roberts 2005)	Low within-cluster dissimilarity and high dissimilarity of samples within clusters to samples outside of clusters.	"W/B", a highly similar algorithm [2] was found effective for cluster recovery [3], but ineffective as a stopping criterion [4].
C-Index (Hubert and Levin 1975)	Low within-cluster dissimilarity, with respect to d_{min} and d_{max} coefficients (see Aho et al. 2006).	Found effective for cluster recovery, and as a stopping criterion [3,4]. Minimum response = optimal solution.
Gamma (Goodman and Kruskal 1954)	High number of concordant compared to non-concordant quadruples (see Aho et al. 2006).	An adaptation of this algorithm [5] was found to be effective for cluster recovery, and as a stopping criterion [3,4].
Point Biserial Correlation (PBC) (Brogden 1949)	Low within cluster dissimilarity.	Found effective for cluster recovery [3,6], May underestimate number of clusters as a stopping criterion [4].
Indicator Species Analysis (ISA) (Average *p*-value) (Dufrêne and Legendre 1997)	High fidelity and abundance of species within particular clusters. ISA-values are tested for significance with Monte-Carlo procedures to calculate *p*-values.	Found useful as a stopping criterion [7,8]. Proposed as a procedure for comparing classification methodologies [8].
Indicator Species Analysis (ISA) (Number of significant indicators, $\alpha = 0.05$) (Dufrêne and Legendre 1997, McCune and Grace 2002)	See explanation above.	Found useful as a stopping criterion [7].
Morisita's index of niche overlap (adapted from Horn 1966)	High proportional occurrence of species within single clusters.	An unbiased estimator of niche overlap [9]. Highly similar to ISAMIC [10].
Indicator Species Analysis Minimizing Intermediate Constancies (ISAMIC) (Roberts 2005)	Consistent presence or absence of species within single clusters.	Highly similar to Morisita's index of niche overlap [10].

* 1 = Kaufman and Rousseeuw (1990), 2 = McClain and Rao (1975), 3 = Milligan (1981), 4 = Milligan and Cooper (1985), 5 = Baker and Hubert (1975), 6 = Milligan (1980), 7 = McCune and Grace (2002), 8 = Dufrêne and Legendre (1997), 9 = Smith and Zaret (1982), 10 = Aho et al. (2006).

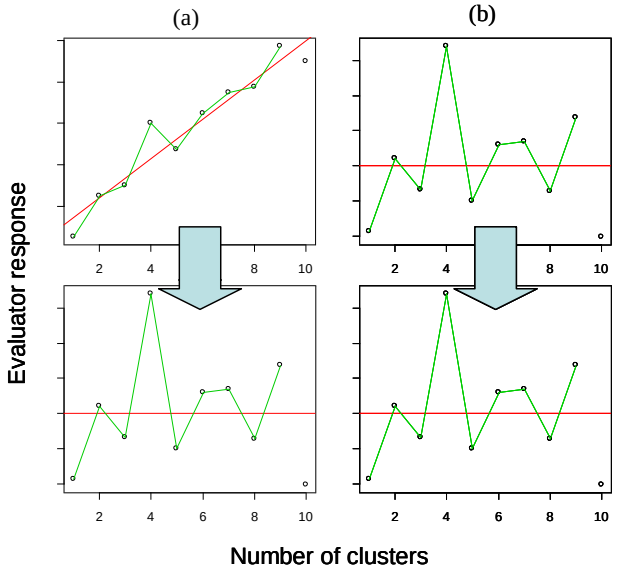

Figure 4. Converting evaluator responses (upper figures) to residuals from linear models lower figures). This is demonstrated for evaluators with possible linear artifact (a), and without linear artifact (b). Note that while predicted optima for the evaluator with linear artifact (a) are radically adjusted, (b) is unchanged.

Computational methods

All classifications were run in PC-ORD (McCune and Mefford, 1999) except for PA M and k-means analysis which were run in R (R development core team, 2005). Coding of evaluators and all other statistical programming was done in R.

Results

1. Distance comparisons

While all distance measures showed significant association with environmental differences using Spearman's rank correlation tests ($p < 0.001$; Fig. 5), Bray-Curtis and Jaccard's dissimilarity explained more of the variability in environmental difference than the other four distance measures ($r_s = 0.4$, Fig. 5). The worst measures for explaining variability were Euclidean and Manhattan distance ($r_s = 0.28$ and 0.214 respectively, Fig. 5). All six distance measures showed a propensity to lose sensitivity (i.e. asymptote) as environmental distance increased (Fig. 5).

2. Classification comparisons

The majority of the classification evaluators found average linkage, flexible $\beta = -0.25$ and Wards linkage to be the best classification methods. Out of 18 Kruskal-Wallis tests comparing evaluator responses (i.e. 9 evaluators · 2 datasets = 18 tests; Table 2), average linkage and flexible $\beta = -0.25$ were the best method in 13 tests, while Ward's linkage was the best method in 12 tests (Table 2). Note that several methods may "tie" for best evaluator response since their medians may not differ significantly. Poorest methods were k-means analysis and TWINSPAN which had the best response in only 1/18, and 3/18 cases respectively (Table 2).

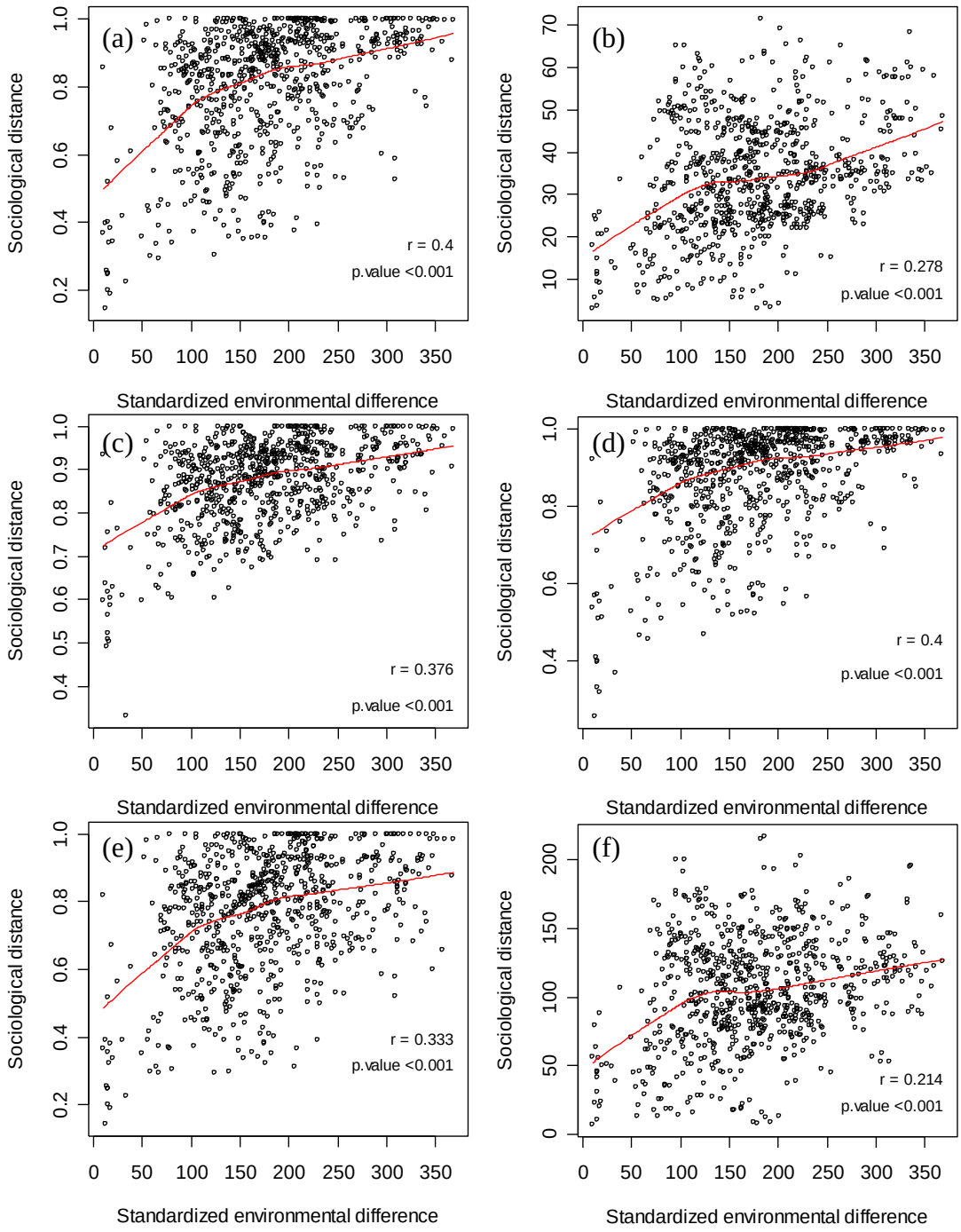

Figure 5. The association between sociological distance and the sum of standardized environmental differences (Equations 2, 3), using 6 different methods to calculate sociological distance: (a) Bray-Curtis, (b) Euclidean, (c) Canberra, (d) Jaccard, (e) Kulczynski, and (f) Manhattan.

Table 2. Medians for standardized evaluator responses and 95% confidence intervals for medians (see McGill et al., 1978 for procedure) calculated across 20 different clustering solutions (2-21 clusters) for each methodology. Methods with the same letter are not significantly different at $\alpha = 0.05$ using Kruskal-Wallis multiple comparisons [simultaneous inference adjusted using Dunn's (1964) procedure]. Best responses in each column are bolded.

	ASW		1-C index		Gamma		PARTANA		PBC	
	alpine	prairie	Alpine	Prairie	Alpine	Prairie	alpine	prairie	alpine	prairie
Average	**0.5 ± 0.7 AB**	**1.0 ± 0.5 A**	**0.6 ± 0.3 AB**	**0.9 ± 0.3 A**	**0.7 ± 0.2 A**	**0.6 ± 0.3 A**	**-0.2 ± 0.1 AB**	**0.3 ± 0.1 A**	**0.7 ± 0.7 A**	**1.5 ± 0.3 A**
Complete	-0.6 ± 0.8 BC	0.1 ± 0.2 CD	0.0 ± 0.3 C	0.1 ± 0.4 BC	**0.5 ± 0.4 AB**	**0.4 ± 0.4 A**	-0.2 ± 0.1 BC	-0.1 ± 0.3 BC	**0.4 ± 0.8 A**	0.0 ± 0.1 B
Flexible	**1.1 ± 0.3 A**	**0.8 ± 0.1 A**	0.2 ± 0.3 BC	**1.1 ± 0.3 A**	**0.7 ± 0.6 AB**	**1.0 ± 0.4 A**	**-0.2 ± 0.1 ABC**	**0.5 ± 0.1 A**	**1.1 ± 0.4 A**	**0.8 ± 0.4 AB**
k-means	-0.6 ± 0.2 C	-0.4 ± 0.2 CD	-1.7 ± 0.3 D	-1.3 ± 0.2 C	-1.9 ± 0.1 C	-0.6 ± 0.4 C	-0.4 ± 0 D	-0.8 ± 0.2 C	-1 ± 0.2 B	-1 ± 0.2 C
PAM	**0.2 ± 0.2 AB**	0.2 ± 0.1 BC	-0.2 ± 0.1 C	0.2 ± 0.3 BC	**0.5 ± 0.2 AB**	**0.6 ± 0.3 AB**	**-0.2 ± 0 ABC**	0.1 ± 0.2 BC	**0.6 ± 0 A**	-0.4 ± 0.2 C
Single	-0.8 ± 0.4 C	-1.6 ± 0.3 D	**1.6 ± 0.2 A**	-0.5 ± 0.1 C	**0.6 ± 0.3 A**	-1.3 ± 0.1 C	**0.2 ± 0.7 A**	-0.1 ± 0.3 BC	-1 ± 0.1 B	-1 ± 0.3 C
TWINSPAN	-0.9 ± 0.1 C	-1.6 ± 0.7 CD	-0.7 ± 0.0 D	-0.5 ± 0.2 C	0.1 ± 0.1 BC	0.0 ± 0.1 BC	-0.3 ± 0 C	-0.3 ± 0.1 C	**0.3 ± 0 A**	0.0 ± 0.6 BC
Ward	**1.0 ± 0.2 AB**	**0.7 ± 0.1 AB**	0.1 ± 0.2 D	**0.7 ± 0.4 AB**	**0.7 ± 0.5 AB**	**0.8 ± 0.4 A**	**-0.2 ± 0 ABC**	**0.3 ± 0.3 AB**	**0.8 ± 0.3 A**	0.1 ± 0.3 BC

	ISAMIC		1-ISA p-val.		ISA sig indicators		1 – Mor.index			High scores		
	Alpine	prairie	Alpine	Prairie	alpine	prairie	alpine	prairie		alpine	prairie	total
Average	**1.1 ± 0.1 A**	0.2 ± 0.4 CD	-0.7 ± 0.3 CD	0.0 ± 0.1 B	-1.3 ± 0.0 D	-0.9 ± 0.2 C	**1.2 ± 0.1 A**	**0.7 ± 0.2 A**		**7**	6	**13**
complete	**0.4 ± 0 A**	-0.4 ± 0.2 D	-0.1 ± 0.3 CD	0.2 ± 0.1 B	-1.1 ± 0.1 D	0.1 ± 0.2 B	**0.8 ± 0.1 A**	-0.2 ± 0.3 C		4	1	5
flexible	-0.5 ± 0.2 B	-0.5 ± 0.2 D	**0.8 ± 0.1 A**	**0.4 ± 0.1 AB**	**1.2 ± 0.2 A**	**0.5 ± 0.2 A**	-0.7 ± 0.2 B	-0.4 ± 0.2 C		6	**7**	**13**
k-means	-0.8 ± 0.3 B	-0.1 ± 0.2 D	0.3 ± 0.1 BC	0.1 ± 0.2 B	0.0 ± 0.1 C	**0.3 ± 0.1 A**	-0.9 ± 0.2 B	-0.6 ± 0.3 C		0	1	1
PAM	-0.7 ± 0.2 B	-0.4 ± 0.2 D	**0.8 ± 0 A**	**0.5 ± 0.2 A**	**1.1 ± 0.1 AB**	**0.7 ± 0.3 A**	-0.8 ± 0.1 B	-0.3 ± 0.3 C		6	3	9
single	**1.9 ± 0.1 A**	**2.6 ± 0.4 A**	-1.9 ± 0.1 D	-2.1 ± 0.2 C	-1.1 ± 0.1 B	-2.1 ± 0.3 D	**1.4 ± 0.1 A**	**1.7 ± 0.1 A**		5	2	7
TWINSPAN	-0.6 ± 0.3 B	0.0 ± 0.2 BC	**0.6 ± 0 AB**	**0.5 ± 0.1 A**	0.0 ± 0.3 BC	-0.2 ± 0.3 BC	-0.2 ± 0.3 B	0.5 ± 0.3 B		2	1	3
Ward	-0.5 ± 0.2 B	-0.5 ± 0.1 D	**0.9 ± 0.1 A**	**0.6 ± 0.1 AB**	**1.2 ± 0.1 A**	**0.6 ± 0.2 A**	-0.7 ± 0.2 B	-0.5 ± 0.2 C		6	6	12

3. Pruning analysis

Since flexible β = -0.25 generally outperformed other methods in tests described above, classifications using this method for the prairie and alpine datasets were examined with pruning analysis (Figs. 6,7, Tables 3, 4). Some evaluators (particularly ISAMIC and Morisita's index of prey preference) showed propensity for linear artifact (i.e. evaluator responses appeared to always increase with increasing numbers of clusters). As a result compensation for linear increase caused predictions of cluster optima to differ dramatically from those of simple standardized responses, particularly in the alpine dataset (compare Figs. 7b and 7d). Indeed, patterns of linear increase with numbers of clusters may be valid for the alpine dataset. This is demonstrated by the fact that almost all evaluators predicted the optimum number of clusters to be greater than the maximum number of clusters tested (i.e. > 21 clusters, Fig. 7a, 7b, Table 4).

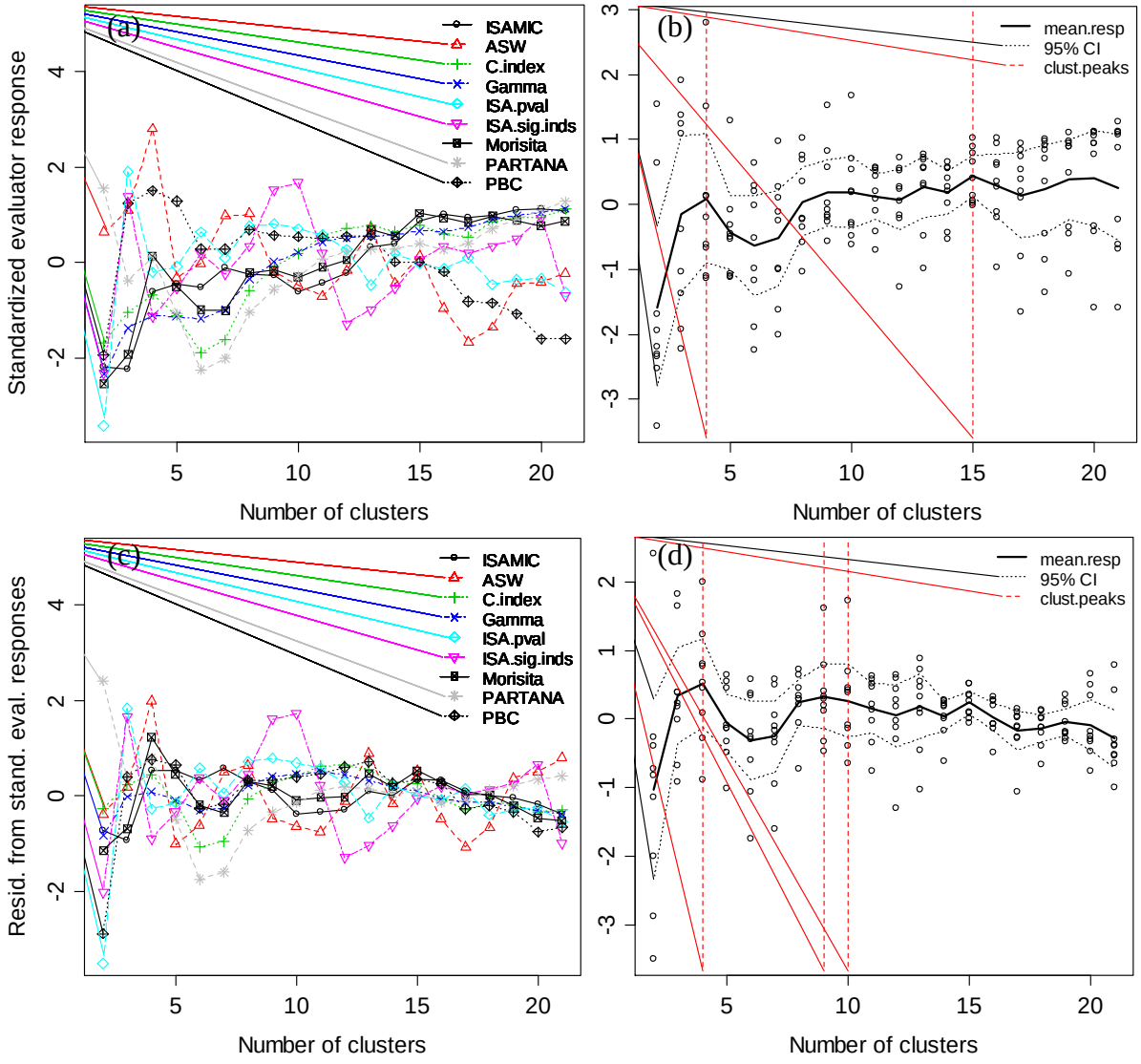

Figure 6. Pruning analysis for flexible β = -0.25 classification of the pre-mine dataset. (a) Standardized evaluator responses. (b) Average standardized evaluator responses. (c) Residuals from evaluator responses plotted against linear models. (d) Average residual responses.

Table 3. Summary of Fig. 6. A = ISAMIC, B = ASW, C = C.index, D = Gamma, E = ISA.pval, F = ISA.sig.inds, G = Morisita, H = PARTANA, I = PBC.

	A	B	C	D	E	F	G	H	I	**Avg.**
Standardized evaluators (Figs. 6a and 6b)										
Max cluster	20	4	21	21	3	10	15	2	4	**15**
Min cluster	3	17	6	3	2	2	2	6	2	**2**
Residuals from stand. evals. (Figs. 6c and 6d)										
Max cluster	7	4	12	11	3	10	4	2	4	**4**
Min cluster	3	17	6	2	2	2	2	6	2	**2**

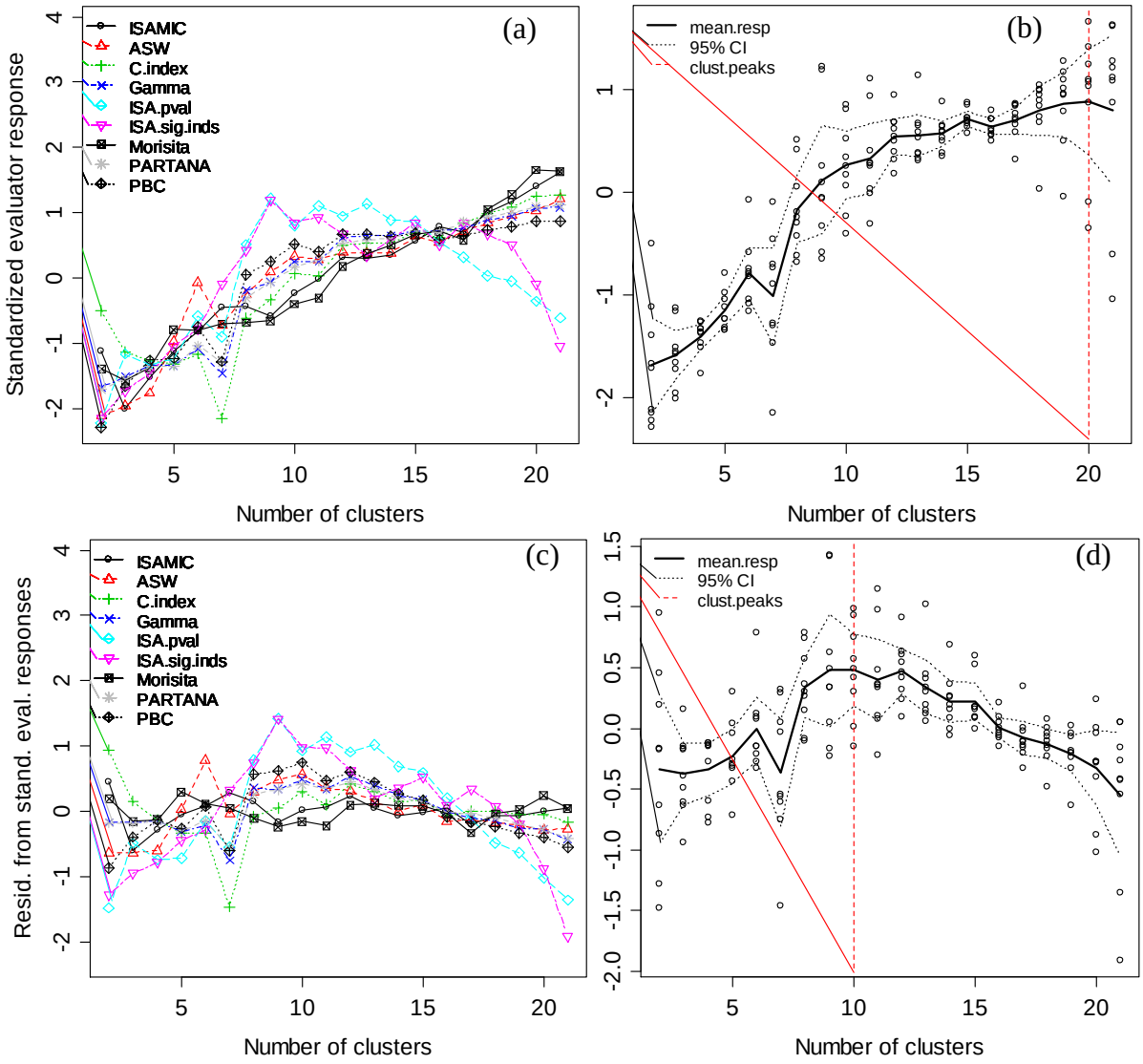

Figure 7. Pruning analysis for flexible β = -0.25 classification of the alpine dataset. (a) Standardized evaluator responses. (b) Average standardized evaluator responses (c). Residuals from evaluator responses plotted against linear models. (d) Average residual responses.

Table 4. Summary of Fig. 7. A = ISAMIC, B = ASW, C = C.index, D = Gamma, E = ISA.pval, F = ISA.sig.inds, G = Morisita, H = PARTANA, I = PBC.

	A	B	C	D	E	F	G	H	I	**Avg.**
Standardized evaluators (Figs. 7a and 7b)										
Max cluster	21	21	21	21	9	9	20	21	21	**20**
Min cluster	3	2	7	2	2	2	3	2	2	**2**
Residuals from stand. evals. (Figs. 7c and 7d)										
Max cluster	2	6	2	12	9	9	5	12	10	**10**
Min cluster	3	2	7	7	2	21	17	7	2	**21**

Discussion

1. Distance comparisons

Bray-Curtis and Jaccard's distance performed well for the experimental dataset, while Manhattan, and Euclidean distance performed poorly. The strength of Bray-Curtis and Jaccard for measuring distance in ecological datasets has been demonstrated by other authors (Beals, 1984; Boyce and Ellison, 2001). The poor performance of Euclidean and Manhattan distance may have to do with their incompatibility with vegetation data. Vegetation species data across sites tends to be positively skewed, i.e. there are many zeros (absences) where a species does not occur. Unlike the other four measures which are indexes which range from 0 to 1, Manhattan and Euclidean distance do not have a fixed upper bound for sample units that have nothing (no species) in common (Roberts. 2005).

The loss of sensitivity of distance measures as environmental distance increases (demonstrated by the curvilinear relationship of sociological to environmental distance in Fig. 5), is a characteristic apparently inherent to all distance measures (McCune and Grace, 2002). Euclidean distance not only loses sensitivity, but is prone to introducing considerable error at moderate to high environmental differences (McCune and Grace, 2002).

2. Classification comparisons

The nine classifications evaluators generally found flexible $\beta = -0.25$, average linkage, and Ward's method to be effective (Table 2). A number of authors have recommended flexible $\beta = -0.25$ since it produces effective and space conserving clusters similar to Ward's method, but is compatible with non-Euclidean distances (Lance and Williams, 1967; McCune and Grace, 2002). Average linkage has been widely recommended as an intuitive and effective method for detecting clusters with spherical or ellipsoidal shapes (Cunningham and Ogilvie, 1972; Milligan and Isaac, 1980; Milligan, 1980; Kaufman and Rousseeuw, 1990). Ward's method is generally effective with spherical multivariate distributions (Kaufman and Rousseeuw, 1990; Hirano et al., 2002), however, since it is based on a sum of squares criterion, Ward's method may perform poorly if cluster populations are unequal in size or have unequal cluster diameters (Kuiper and Fisher, 1975; Kaufman and Rousseeuw, 1990; Milligan, 1980). In addition, although Ward's method may perform well with non-Euclidean distance measures (Cao et al., 1997), this may not be entirely appropriate since Ward's method calculates internal dendrogram distances with the Euclidean method (Peilou, 1984).

Classification evaluators found TWINSPAN and k-means analysis solutions to be ineffective. TWINSPAN may suffer from a number of problems present in its parent method,

correspondence analysis (CA), which performs poorly in finding patterns for multi-gradient datasets (McCune and Grace, 2002), and implicitly uses chi-squared distance, a measure which accords high weights to species with low total abundance (Faith et al., 1987). It should be noted that Dale (1995) argued TWINSPAN should not be compared to other classification methods solely on the basis of clustering since its primary goal is characterization through indicator species, not creating a cluster structure.

The poor performance of k-means analysis may be due to several factors. The problems associated with Euclidean distance (the measure used by k-means analysis) and vegetation data were noted above, i.e. errors are introduced at higher distances, and there is no fixed upper bound for sample units that have nothing in common. Dufrêne and Legendre (1997) used the Principal Coordinates Analysis (PCoA) species and site scores from an ordination of a Stienhaus similarity matrix instead of Euclidean distances to avoid this problem. In addition, while random k-means starting points (our methodology) generally results in poor cluster recovery with synthetic datasets, much better recovery characteristics may be exhibited when using valid seed points, e.g. average linkage centroids (Milligan, 1980). Finally, for our datasets, the use of means may be disadvantageous to k-means analysis, since this statistic is often a poor indicator of centrality in ecological data (Kaufman and Rousseeuw, 1990). The much better performance of PAM, a non-hierarchical method which addresses several of these deficiencies, is evident in Table 2.

While my results indicate that flexible $\beta = -0.25$, average linkage, and Ward's method are effective classification methods for vegetation data, they also indicate that methods can vary in efficacy with different ecological datasets. Thus, our results support the idea that the "fit" between classification methods and data should be verified (using methods similar to those in this paper) preceding a definitive classification of data.

3. Pruning analysis and Community summaries

Although evaluators were often not in agreement over exact optima locations, most had peaks in generally the same area, indicating legitimate pruning regions. Exceptions include ISAMIC and Morisita's index whose linear artifact causes them to be poor choices for pruning analysis. Indeed, eliminating these evaluators resulted in plots of simple standardized responses and plots of residuals from standardized responses to be essentially identical for the pre-mine dataset (Fig. 8, Table 5).

Pruning analysis indicated that the optimum number of clusters may be greater than the maximum number of clusters tested for the alpine dataset (i.e. > 21 clusters, Fig. 7a, 7b, Table 4). In contrast, several legitimate pruning solutions were indicated for the prairie dataset (Figs. 6, 8, Tables 3, 5). The multiple optima suggest a hierarchical structure in the pre-mine dataset with divisions at around 4 and 10 clusters (Figs. 6c, 6d, 8, Tables 3, 5).

The four clusters indicate general physiognomic types within the pre-mine landscape. The clusters represent 1) grassland; 2) disturbed pastureland; 3) bottomland; and 4) skunkbush shrub/grassland. These types are clearly shown with NMDS ordination scatterplots (Kruskal and Wish, 1978; Fig. 9). Ordination is a mathematical approach to data that allows samples to be organized on a plot so that those samples which are most similar in terms of species composition and relative abundance will appear closer together, while those which have different species composition will be positioned further apart. For more detail on ordination see Gauch (1982), Johnson and Wichern (1998), and/or McCune and Grace (2002). The species composition of the four physiognomic types can be demonstrated with relevé tables which summarize species

constancy and cover within the communities (Table 6). The ten clusters represent subdivisions of the four cluster solution. These clusters can also be well represented by ordination (Fig. 10), and relevé tabling (Table 7).

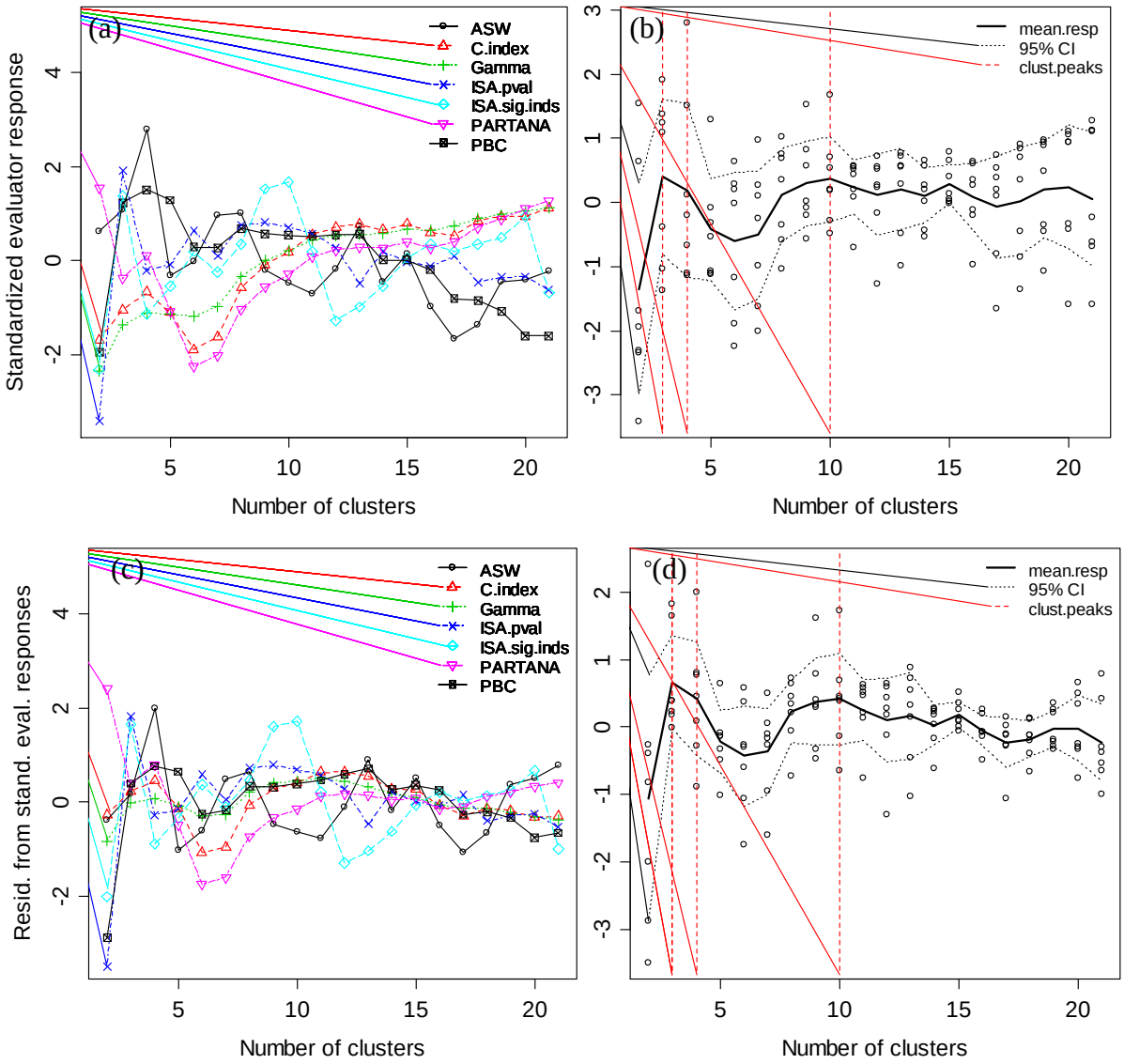

Figure 8. Pruning analysis for flexible β = -0.25 classification of the prairie (pre-mine) dataset without ISAMIC and Morisista's index of niche overlap. (a) Standardized evaluator responses. (b) Average standardized evaluator responses. (c) Residuals from evaluator responses plotted against linear models. (d) Average residual responses.

Table 5. Summary of Fig. 8. B = ASW, C = C.index, D = Gamma, E = ISA.pval, F = ISA.sig.inds, H = PARTANA, I = PBC.

	B	C	D	E	F	H	I	**Average**
Standardized evaluators (Figs. 8a and 8b)								
Max cluster	4	21	21	3	10	2	4	3
Min cluster	17	6	2	2	2	6	2	2
Residuals from stand. evals. (Figs. 8c and 8d)								
Max cluster	4	12	11	3	10	2	4	3
Min cluster	17	6	2	2	2	6	2	2

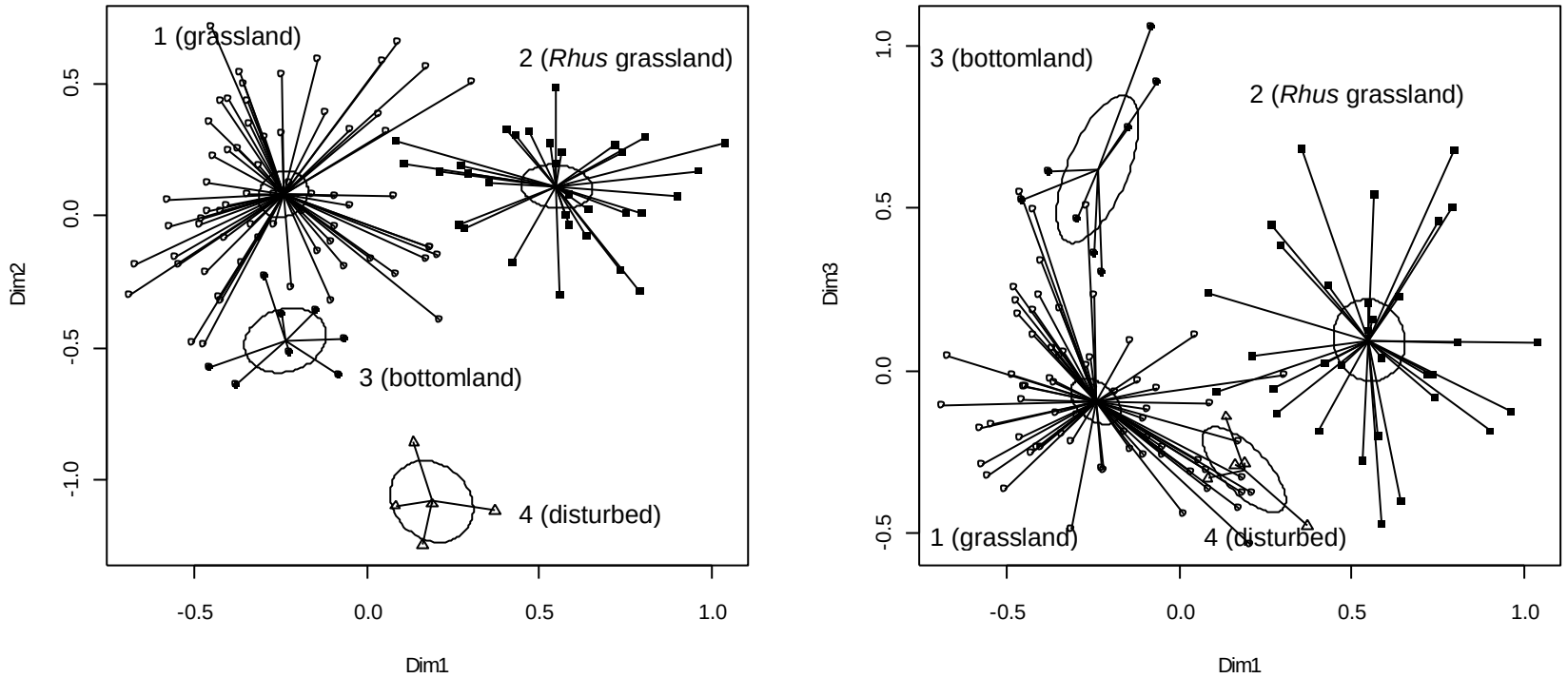

Figure 9. Prairie (pre-mine) dataset 3-dimensional ordination solution. Four cluster solution from classification analysis superimposed. Ellipses are 95% confidence intervals around cluster centroids.

Table 6. Summarized relevé table for the pre-mine dataset 4 cluster solution. This table lists all species that occur with >30% constancy in at least one of the four types. Species occurrences with > 50% constancy are shaded. Indicator species are bolded. A two character cipher[1] is included in each cell which indicates constancy (percentage of sites that contain the species), and cover respectively.

	Community 1 (grassland)	Community 2 (*Rhus* grassland)	Community 3 (bottomland)	Community 4 (disturbed)
Vulpia octoflora	4A	+A	..	1A
Artemisia dracunculus	4A	++	1A	1A
Bouteloua gracilis	8C	3A	3A	..
Carex filifolia	5D	4B	1B	..
Phlox hoodii	5A	6A	..	..
Lygodesmia juncea	3A	+A	2A	1A
Carex pensylvanica	2C	4B	..	..
Aster falcatus	2A	3A	..	..
Yucca glauca	1A	5B	..	..
Andropogon scoparius	1A	3B	..	..
Astragalus adsurgens	1A	3A	..	..
Bouteloua curtipendula	+A	4C	..	..
Comandra umbellata	+A	3A	..	..
Calamovilfa longifolia	3B	3A	2A	..
Artemisia cana	4B	3A	6C	..
Cerastium arvense	2A	3A	2A	..
Solidago missouriensis	2A	3A	1A	..
Echinacea angustifolia	2A	3A	1A	..
Artemisia ludoviciana	2A	2A	7A	..
Hedeoma hispidum	2A	++	1A	..
Rhus trilobata	1B	9D	3C	..
Linum lewisii	1A	4A	1A	..
Agropyron spicatum	5C	9D	1A	1A
Agropyron smithii	9D	7B	9D	5A
Bromus japonicus	9D	7B	8C	5A
Koeleria cristata	9D	7B	3B	5A
Stipa comata	9D	6B	3B	1A
Gutierrezia sarothrae	5C	6B	1A	5B
Bromus tectorum	3B	2A	1A	1A
Stipa viridula	5C	4A	6C	3A
Tragopogon dubius	5A	4A	4A	1A
Gaura coccinea	5A	3A	4A	1A
Artemisia frigida	6B	3A	3A	5B
Sphaeralcea coccinea	7A	3A	1A	5A
Psoralea argophylla	5A	3A	1A	3A
Poa pratensis	5C	2A	9E	3A
Taraxacum officinale	5A	1A	6A	7A
Poa secunda	6B	1A	2A	3A
Plantago argyraea	7A	1A	1A	5A
Achillea millefolium	3A	5A	9C	1A
Melilotus officinalis	3B	2A	4C	5D
Rosa arkansana	2A	3A	2A	1A
Aristida purpurea	1A	+A	1A	3A
Lithospermum incisum	3A	++	..	3A
Aster campestris	1A	3A	..	1A
Lactuca oblongifolia	1A	++	3A	..
Symphoricarpos occidentalis	+A	4B	8E	..
Rosa woodsii	..	+A	6D	..
Androsace occidentalis	1A	++	..	3A
Agropyron cristatum	1A	..	1A	9E
Alyssum alyssoides	1A	..	..	7C
Vicia americana	+A	1A	..	3B
Convolvulus arvensis	..	..	..	3C

[1]For each cell in the body of the table, constancy is indicated by the first symbol, while cover is indicated by the second symbol. For constancy: 0% = ".", 0-10% = +, 10-20% = 1, 20-30% = 2, 30-40% = 3, 40-50% = 4, 50-60% = 5, 60-70% = 6, 70-80% = 7, 80-90% 8, 90-100% = 9. For cover: 0% = ".", 0-0.01% = +, 0.01-1% = A, 1-2% = B, 2-5% = C, 5-25% = D, >25% = E.

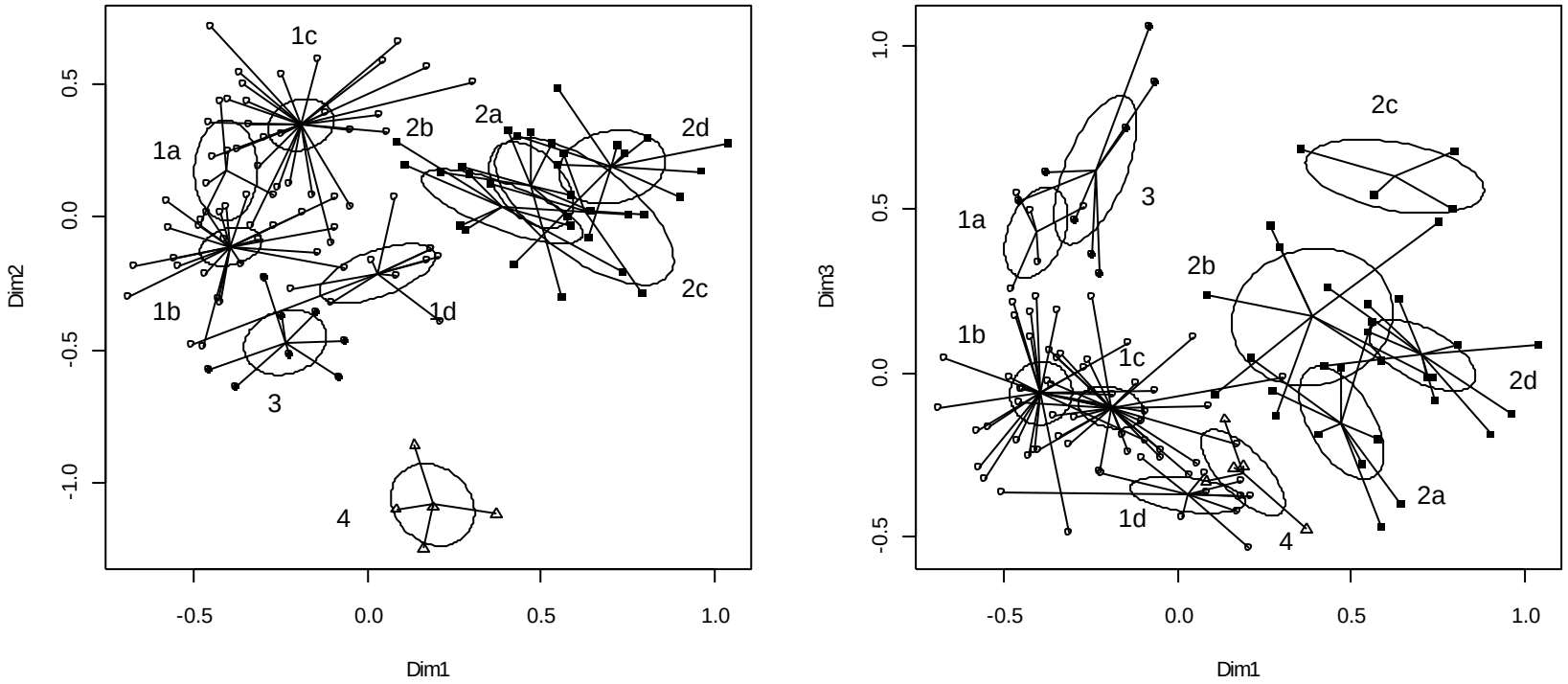

Figure 10. Rosebud NMDS 3 dimensional solution. Final stress = 13.76. 10 cluster solution from classification analysis superimposed. 1a = STCO-AGSM-BOGR-CAPE, 1b = STCO-AGSM-BOGR-ARFR, 1c STCO-AGSM-BOGR-CAFI, community 1d = ARTR-STVI), 2a = RHTR-AGSP-CAFI, 2b = RHTR-AGSP, 2c = RHTR-AGSP-CAPE-FEID, 2d = RHTR-AGSP-BOCU-ANSC), 3 = AGSM-POPR-ROWO, 4 = AGCR. Ellipses are 95% confidence intervals around cluster centroids.

Table 7. Summarized relevé for the pre-mine dataset 10 cluster solution. This table lists all species that occur with >30% constancy in at least one of the ten types. Species occurrences > 50% constancy are shaded. Important indicator species are bolded. A two character cipher[1] is included in each cell which indicates constancy and cover respectively.

	Grassland				*Rhus* grassland				bottomland	disturbed
	1a STCO-AGSM-BOGR-CAPE	1b STCO-AGSM-BOGR-ARFR	1c STCO-AGSM-BOGR-CAFI	1d ARTR-STVI	2a RHTR-AGSP-CAFI	2b RHTR-AGSP	2c RHTR-AGSP-CAPE-FEID	2d RHTR-AGSP-BOCU-ANSC	3 AGSM-POPR-ROWO	4 AGCR
Collomia linearis	3A	+A	..	..	..	..	..	..	..	..
Astragalus adsurgens	3A	+A	+A	4A	4B	2A	4A	3A	..	..
Rosa arkansana	5C	2A	2A	1A	4A	3A	4A	2A	2A	1A
Artemisia frigida	1A	8C	6B	6A	2A	3A	4A	2A	3A	5B
Psoralea argophylla	5A	5A	6A	5A	2A	3A	4A	2A	1A	3A
Gaura coccinea	7A	6A	5A	3A	5A	2A	..	4A	4A	1A
Bouteloua gracilis	7A	8C	9D	6B	4A	6A	..	2A	3A	..
Sphaeralcea coccinea	7A	8A	5A	7A	2A	6A	2A	4A	1A	5A
Stipa comata	9D	9D	9D	6B	7C	4C	7A	7B	3B	1A
Agropyron smithii	9D	9D	8D	9D	5B	9C	4A	8B	9D	5A
Stipa viridula	3B	6C	2B	9D	7B	6A	4A	1A	6C	3A
Koeleria cristata	9A	8D	9C	8D	9C	9C	2A	5A	3B	5A
Gutierrezia sarothrae	1A	5B	5B	8C	8C	7A	2A	6B	1A	5B
Artemisia cana	9D	6C	2A	2A	3A	4B	7B	1A	6C	..
Bromus tectorum	5B	4C	3B	1A	2A	2A	..	2A	1A	1A
Poa pratensis	9D	7D	3B	3A	1A	3A	2A	1A	9E	3A
Bromus japonicus	9D	9D	9D	9D	6A	9C	4A	7A	8C	5A
Achillea millefolium	5A	4A	2A	6B	6B	7A	9B	1A	9C	1A
Tragopogon dubius	9A	4A	5A	7A	8A	3A	2A	1A	4A	1A
Taraxacum officinale	3A	7A	3A	3A	3A	1A	..	++	6A	7A
Melilotus officinalis	3A	4B	2B	4A	3A	3A	..	1A	4C	5D
Alyssum alyssoides	..	1A	2A	3A	..	..	..	..	..	7C
Agropyron cristatum	..	1A	1A	+C	..	..	..	..	1A	9E
Lygodesmia juncea	5A	2A	4A	++	1A	..	..	++	2A	1A
Artemisia dracunculus	7B	4A	5B	+A	..	1A	..	..	1A	1A
Vulpia octoflora	5A	4A	5A	2A	..	2A	..	++	..	1A
Astragalus gracilis	1A	+A	4A	2A	1A	..	..	+A	..	1A
Plantago argyraea	3A	9A	8A	3A	..	4A	..	+A	1A	5A
Aristida purpurea	1A	1B	1A	..	..	..	..	+A	1A	3A
Erysimum asperum	3A	2A	3A	1A	..	1A	..	1A	..	1A
Lithospermum incisum	3A	1A	5A	1A	..	1A	..	..	..	3A
Androsace occidentalis	1A	3A	++	1A	..	1A	..	..	..	3A
Hedeoma hispidum	1A	4A	2A	2A	..	1A	..	..	1A	..
Lactuca oblongifolia	3A	1A	+A	..	1A	..	..	..	3A	..
Solidago mollis	3A	+A	1A	..	..	..	..	++	1A	..
Artemisia tridentata	..	+A	+A	8D	2A	1A	..	+A	..	..
Calamovilfa longifolia	5C	1A	5C	..	2A	2A	4A	4A	2A	..
Calamagrostis montanensis	3B	1A	4A	++	5A	1A	2A	..	..	..
Carex pensylvanica	9E	1A	4C	..	3B	3C	9D	3A	..	..
Dalea purpurea	1A	++	+A	+A	2A	3A	..	++	..	..
Agropyron spicatum	..	2B	6C	8D	9D	9D	9C	9D	1A	1A
Cerastium arvense	3B	1A	2A	2A	5A	2A	9A	1A	2A	..
Artemisia ludoviciana	5C	3A	1A	..	3A	3A	7A	..	7A	..
Rhus trilobata	3C	1A	1B	..	8C	9D	9B	9C	3C	..
Solidago missouriensis	1A	1A	3A	..	5A	1A	9A	1A	1A	..
Bouteloua curtipendula	..	+A	+A	..	3B	2B	4A	8D	..	..
Yucca glauca	..	1A	2B	+A	6C	3A	2A	6A	..	..
Aster falcatus	..	2A	2A	1A	6A	3A	2A	++	..	..
Phlox hoodii	..	6A	4A	8B	8A	8A	4A	2A	..	..
Antennaria parvifolia	..	1A	1A	++	3A	2A	2A	++	..	..
Andropogon scorapius	..	..	3B	..	3A	2A	..	6C	..	..
Liatris punctata	..	+A	3A	+A	3A	..	..	3A	..	..
Carex filifolia	..	2B	9D	4B	9C	4B	..	1A	1B	..
Linum lewisii	..	1A	+A	4A	5A	6A	4A	2A	1A	..
Comandra umbellata	..	+A	++	1A	3A	4A	2A	1A	..	..
Echinacea angustifolia	..	1A	1A	4A	4A	3A	2A	4A	1A	..
Poa secunda	..	6C	7A	6A	2A	2A	..	+A	2A	3A
Aster campestris	..	1A	1A	2A	4A	3A	2A	2A	..	1A
Vicia americana	..	++	+A	3A	1A	3A	2A	+A	..	3B
Symphoricarpos occidentalis	..	+A	+A	..	4A	3B	9C	2A	8E	..
Prunus virginiana	..	+A	..	..	3A	2A	..	+A	2C	..
Crepis acuminata	1A	..	+A	1A	4A	1A	..	..	..	..
Muhlenbergia cuspidata	..	+A	1A	..	2A	4B	..	+A	..	..
Oxytropis sericea	..	..	+A	..	1A	..	4A	..	..	..
Campanula rotundifolia	..	..	+A	..	2A	..	4A	++	..	..
Penstemon nitidus	..	..	++	2A	3A	..	2A	+A	..	..

Table 7. cont.

	Grassland				*Rhus* grassland				bottomland	disturbed
	1a STCO-AGSM-BOGR-CAPE	1b STCO-AGSM-BOGR-ARFR	1c STCO-AGSM-BOGR-CAFI	1d ARTR-STVI	2a RHTR-AGSP-CAFI	2b RHTR-AGSP	2c RHTR-AGSP-CAPE-FEID	2d RHTR-AGSP-BOCU-ANSC	3 AGSM-POPR-ROWO	4 AGCR
Festuca idahoensis	..	..	..	++	..	..	7D	..	..	..
Pinus ponderosa	..	..	..	..	1A	..	4A	2A	..	..
Amelanchier alnifolia	..	..	..	..	2A	1A	4A	..	..	..
Solidago nemoralis	..	..	..	..	2A	1A	4A	..	..	..
Phlox alyssifolia	..	..	..	..	3A	1A	..	..	..	..
Juniperus scopulorum	..	..	..	..	1A	3C	..	+A	..	..
Chrysothamnus nauseosus	..	..	..	..	..	3A	..	+A	..	..
Rosa woodsii	..	..	..	..	..	1B	..	..	6D	..
Convolvulus arvensis	..	..	..	..	..	..	..	..	..	3C

[1]For each cell in the body of the table, constancy is indicated by the first symbol, while cover is indicated by the second symbol. For constancy: 0% = ".", 0-10% = +, 10-20% = 1, 20-30% = 2, 30-40% = 3, 40-50% = 4, 50-60% = 5, 60-70% = 6, 70-80% = 7, 80-90% 8, 90-100% = 9. For cover: 0% = ".", 0-0.01% = +, 0.01-1% = A, 1-2% = B, 2-5% = C, 5-25% = D, >25% = E.

Conclusions

In this report I summarize the three steps that must be followed for effective cluster analysis.

1. An effective distance measure was found. Bray-Curtis and Jaccard's distance were effective dissimilarity/distance measures for the vegetation dataset tested, while Manahattan, and Euclidean distance performed poorly. In accordance with these results, Bray-Curtis dissimilarity was used whenever possible to create classifications used for analyses in steps 2 and 3 below.

2. Effective classification methods were identified. Three hierarchical agglomerative methods: flexible $\beta = -0.25$, average linkage, and Ward's method were the most effective classification methods for datasets tested. In response to this flexible $\beta = -0.25$ classifications were used for the final cluster analysis step: pruning analysis.

3. Optimal cluster numbers (pruning optima) were determined. While the optimum number of clusters appeared to be greater than the maximum number of clusters tested for the alpine dataset, legitimate pruning solutions were found for the prairie dataset. The multiple optima suggest a hierarchical structure with ten clusters nested within four general physiognomic types.

It is inevitable that land managers who require vegetation community analyses that are objective and reproducible will eventually turn to multivariate cluster analyses for recognition of vegetation patterns. If used thoughtfully in conjunction with comprehensive and accurate field data, such analyses may provide quantitative middle ground between regulators and the mine industry for agreement on mine reclamation guidelines. Thus cluster analysis may provide legitimate criteria for bond release in mining land reclamation.

Acknowledgements

I thank the Reclamation Research Unit at MSU (e.g. Dennis Neuman and Pam Blicker), members of a committee studying categorization and characterization of Southeast Montana coal mines: S. Regle (DEQ), and R. Montgomery, D. Myran, P. Martin, and B. Green, and T. Weaver and D. Roberts at MSU for ecological insights and help with statistical programming.

Literature cited

Aho, K. 2006. Alpine and cliff ecology in the Northern Rocky Mountains. PhD dissertation. Montana State University, Bozeman MT, USA.

Aho, K., Roberts, D. and T. Weaver. 2006. Using geometric and non-geometric internal evaluators to compare eight commonly used vegetation classification methods. Submitted to Journal of Vegetation Science, spring 2006.

Baker, F. B., and L. J. Hubert. 1975. Measuring the power of hierarchical cluster analysis. J. Am. Stat. Assoc. 70, 31-38. http://dx.doi.org/10.1080/01621459.1975.10480256

Beals, E. W. 1984. Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data. Advances in ecological research. 14: 1-55. http://dx.doi.org/10.1016/S0065-2504(08)60168-3.

Belbin, L., and C. McDonald 1993. Comparing three classification strategies for use in ecology. J. Veg. Sci. 4: 341-348. http://dx.doi.org/10.2307/3235592.

Boyce, R. L, and P. C. Ellison. 2001. Choosing the best similarity index when performing fuzzy set ordination on binary data. J. Veg. Sci. 12: 711-720. http://dx.doi.org/10.2307/3236912.

Braun-Blanquet, J. 1964. Pflanzenoziologia. Grundüge der Vegetationskunde. 3. Aufl. Berlin. Wien. Springer Verlag, New York, USA.

Brogden, H. E. 1949. A new coefficient: application to biserial correlation and to estimation of selective efficiency. Psychometrika. 14: 169-182 http://dx.doi.org/10.1007/BF02289151 PMid:24536228

Cao, Y., Bark, A. W., and W. P. Williams. 1997. A comparison of clustering methods for river benthic community analysis. Hydrobiologia. 347: 25-40. http://dx.doi.org/10.1023/A:1002938721135.

Cunningham, K. M., and J. C. Ogilvie. 1972. Evaluation of hierarchical grouping techniques: a preliminary study. Comp. J. 15: 209-213. http://dx.doi.org/10.1093/comjnl/15.3.209.

Dale, M. B. 1991. Knowing when to stop: cluster concept=concept cluster. In Feoli, E. and L. Orloci (eds.), Computer Assisted Vegetation Analysis, Kluwer Academic Publishers, Dordrecht: 149-171. http://dx.doi.org/10.1007/978-94-011-3418-7_14.

Dale, M. B. 1995. Evaluating Classification Strategies. J. Veg. Sci. 6: 437-440. http://dx.doi.org/10.2307/3236243.

De'eath, G. 1999. Extended dissimilarity: a method of robust estimation ecological distances from high beta-diversity data. Plant ecology. 144: 191-199. http://dx.doi.org/10.1023/A:1009763730207.

Dufrêne M., P., and P. Legendre. 1997. Species assemblages and indicator species: the need for a flexible asymmetric approach. Ecol. Mong. 67(3): 345-367. http://dx.doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2 http://dx.doi.org/10.2307/2963459.

Dunn, O. J. 1964. Multiple comparisons using rank sums. Technometrics 6: 241-252. http://dx.doi.org/10.1080/00401706.1964.10490181.

Gauch, H. 1982. Multivariate Analysis in Community Ecology. Cambridge University Press, New York NY, USA. http://dx.doi.org/10.1017/CBO9780511623332.

Gauch, H. G., and R. H. Whittaker. 1981. Hierarchical classification of community data.J. Ecol. 69: 537-557. http://dx.doi.org/10.2307/2259682.

Goodman, L., and W. Kruskal. 1954. Measures of association for cross-validations. J. Am. Stat. Assoc. 732-764. http://dx.doi.org/10.1080/01621459.1954.10501231 http://dx.doi.org/10.2307/2281536.

Hartigan, J. A., and M. A. Wong. 1979. A k-means clustering algorithm. Appl. Stat.-J. Roy. St. C. 28: 100-108.

Hill, M. O. 1979. TWINSPAN—A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Ecology and Systematics. Cornell University, Ithaca, NY, USA.

Hirano, S., Sun., X. & Tsumoto, S. 2002. Comparison of clustering methods for clinical databases. Inform. Sciences. 159: 155-165. http://dx.doi.org/10.1016/j.ins.2003.03.011.

Horn, H. S. 1966. Measurement of "overlap" in comparative ecological studies. Am. Naturalist. 100: 419-424. http://dx.doi.org/10.1086/282436.

Hubert, L. J., and J. R. Levin. 1976. A general framework for assessing categorical clustering in free recall. Psychol. Bull. 83: 1072-1080. http://dx.doi.org/10.1037/0033-2909.83.6.1072.

Johnson, R. A., and D. W. Wichern. 1998. Applied multivariate statistical analysis, 4th edition. Prentice-Hall, Upper Saddle River, NJ, USA.

Kaufman, L., and P. J. Rousseeuw. 1990. Finding groups in data, an introduction to cluster analysis. John Wiley & Sons, New York,USA.

Kruskal, J. B., M. Wish. 1978. Multidimensional scaling. Sage Publications. Beverly Hills, CA. 93 pp.

Kuiper, F. K., and L. Fisher. 1975. A Monte Carlo comparison of six clustering procedures. Biometrics. 31: 777-783. http://dx.doi.org/10.2307/2529565.

Lance, G. N., and W. T. Williams. 1967. A general theory of classification sorting strategies I. Hierarchical systems. Comp. J. 4: 373-380. http://dx.doi.org/10.1093/comjnl/9.4.373.

McCune, B., and J. B. Grace. 2002. Analysis of ecological communities. MjM Software design. Gelenden Beach OR, US.

McCune B., and M. S. Mefford. 1999. PC-ORD: Multivariate analysis of ecological data, ver. 4. MjM software design, Glendale Beach, Oregon, US.

McClain, J. O., and V. R. Rao. 1975. CLUSTISZ: A program to test for the quality of clustering of a set of objects. *J. Marketing Res.* 12: 456-460.

McGill, R., Tukey, J. W., and W. A. Larsen. 1978. Variations of box plots. Am. Stat. 32: 12-16. http://dx.doi.org/10.1080/00031305.1978.10479236 http://dx.doi.org/10.2307/2683468.

McQuitty, L. L. 1966. Similarity analysis by reciprocal pairs for discrete and continuous data. Educ. Psychol. Meas. 26: 825. http://dx.doi.org/10.1177/001316446602600402 http://dx.doi.org/10.1177/001316446602600303.

Milligan, G. W. 1980. An examination of six types of error perturbation fifteen clustering algorithms. Psychometrika. 45: 325-342. http://dx.doi.org/10.1007/BF02293907.

Milligan, G. W. 1981. A Monte Carlo study of thirty internal criterion measures for cluster analysis. Psychometrika. 46(2): 187-199. http://dx.doi.org/10.1007/BF02293899.

Milligan, G. W., and M. C. Cooper. 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika. 50 (2): 159- 179. http://dx.doi.org/10.1007/BF02294245.

Milligan, G. W., and P. D. Isaac. 1980. The validation of four ultrametric clustering algorithms. Pattern Recogn. 12: 41-50. http://dx.doi.org/10.1016/0031-3203(80)90001-1.

Orloci, L. 1967. An agglomerative method for classification of plant communities. J. Ecol. 55: 193-206. http://dx.doi.org/10.2307/2257725.

Peilou, E. C. 1984. The interpretation of ecological data. Wiley, New York, US.

R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Roberts, D. 2005. Vegetation classification in R, for labdsv ver. 1.1-1, vegetation

ecology package. http://www.cran.r-project.org. unpublished.

Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20: 53-65. http://dx.doi.org/10.1016/0377-0427(87)90125-7.

Smith, E. P., and T. M. Zaret. 1984. Bias in estimating niche overlap. Ecology. 63: 1248-1258 http://dx.doi.org/10.2307/1938851

Sneath, P. H. A., and R. R. Sokal. 1973. Numerical taxonomy, the principles and practice of numerical exploration. W. H. Freeman and Company. San Francisco, CA USA.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin. 38: 1409-1438.

Sokal, R. R., and P. H. A. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman and Company. San Francisco, CA USA.

Ward, J. H., Jr. 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58: 236-244. http://dx.doi.org/10.1080/01621459.1963.10500845.

Weaver, T., and K. Aho. 2005. Identification of community types of SE Montana as targets for mine reclamation. *In* CCV Phase IV final report: characterization of vegetation types and finalization of database. Reclamation Research Unit, Montana State University, Bozeman MT.