Large scale related effects on the determination of plant communities and relationships with environmental variables

P. B. Drewa¹,² and G. E. Bradfield¹,³

¹ Department of Botany, University of British Columbia, 6270 University Blvd., Vancouver, BC, Canada V6T 1Z4. E-mail: garyb@interchange.ubc.ca.
² Present address: Jornada Experimental Range, USDA-Agricultural Research Service, Box 30003, MSC 3JER, New Mexico State University, Las Cruces, NM, USA 88003. E-mail: pdrewa@nmsu.edu.
³ Author responsible for correspondence.

Keywords: Environmental variance explained. Grain size. Minimum increase of sum of squares cluster analysis. Principal component analysis. Sampling scale. Vegetation classification.

Abstract: The influence of scale on the discernment of plant community patterns was examined using vegetation-environment data collected from a subalpine wet meadow in south-coastal British Columbia. Species cover data were recorded in 225, 0.25 m² quadrats systematically located at 5 m intervals in a 40 m x 120 m sampling grid. Environmental data consisted of quadrat elevations as well as measured and kriged estimates of five soil variables (carbon content, pH, electrical conductivity, percent sand, and percent clay). Sampling scale was adjusted by aggregating neighbouring quadrats into composite sampling units; analytical scale was altered by varying the intercept level in dendrograms corresponding to the different sampling scales. The resulting classifications were evaluated for their ability to explain variation in the vegetation data and in the environmental data. The vegetation variation explained by the classifications was highest at the smallest sampling scale indicating that vegetation heterogeneity is fine grained. In contrast, the environmental variation explained was higher for the classifications based on the larger composite sampling units implying a coarser scaling of abiotic conditions within the study area. These results were consistent with the recognition of three main zones along a drainage gradient within the sampling grid — upper mixed-forb, middle heath, and lower sedge. There was also evidence that the orientation of rectangular sampling units parallel to the drainage gradient leads to higher levels of explained variation. This study reaffirms the need for careful consideration of alternatives both in field sampling and analytical phases of vegetation research to ensure that description and interpretation of patterns adequately address study objectives and that vegetation-environment relationships are more completely investigated from a hierarchical perspective.

Abbreviations: MISSQ-Minimum increase of sum of squares cluster analysis; PCA-Principal component analysis.


Introduction

Hierarchical classifications of plant community data rarely explore the influence of changing the sampling scale and dendrogram level on the proportion of vegetation and environmental variation explained. Sampling unit size has long been known to affect the form and interpretation of vegetation pattern (Evans 1952, Goodall 1961, Kershaw 1961, Noy-Meir et al. 1970, Greig-Smith 1983). In particular, block size variance analysis (Greig-Smith 1952, 1961, Austin 1968) and multiple pattern analysis (Noy-Meir and Anderson 1971) not only demonstrated how varying quadrat size alters the elucidation of vegetation patterns but also affects the interpretation of vegetation-environment relationships. More recently, the manipulation of sampling unit size has become a standard method to examine scale related influences in studies of pattern-process relationships (Reed et al. 1993, Schaefer and Messier 1995, Qi and Wu 1996, Rescia et al. 1997, Fortin 1999).

Classification may provide a basis for invoking process-oriented explanations for the groups generated. In this context, the statement by Diggle (1983) should be heeded: “The nature of the pattern generated by a biological process can be affected by the physical scale on which the process is observed.” Similarly, selection of generated
groups from a cluster analysis influences pattern interpretation and, in turn, affects the clarity with which pattern-process relationships are interpreted.

Selection of an appropriate dendrogram level from hierarchical cluster analysis is a difficult problem for two reasons. First, hierarchical clustering methods provide little guidance in choosing the number of groups in a data set. Second, the effectiveness of procedures used to determine the number of clusters may be a function of inherent qualities of the data (Everitt 1980, Milligan and Cooper 1985). Numerous methods have been developed to select the number of groups in a data set (Marriot 1971, Calinski and Harabasz 1974, Ratliff and Pieper 1981). In particular, Thorndike (1953) and Gower (1975) introduced similar methods to determine the number of clusters where mean within-cluster distance is plotted against the number of data partition groups. Multiple discriminant analysis has been used to test the goodness of classification (Matthews 1979); however, this method is limited by the assumption of equal variance-covariance matrices, and by arbitrary decisions relating to probability levels and the dimensionality of the discriminant space (Podani 1984). Stopping rules have also been developed to determine the number of clusters in a classification (Hogeweg 1976, Hill 1980, Popma et al. 1983). More recently, Pillar (1999) proposed a method that statistically validates the selection of a dendrogram level through simulated resampling of vegetation data via bootstrapping.

Levels in a hierarchy of varying observational scales are a function of observer-chosen criteria (Allen and Starr 1982, O'Neil et al. 1986). We recognize two hierarchies in this study: a hierarchy of sample unit size (termed 'grain' size), and a classification hierarchy related to the sequence of linkages joining the sampling units in a dendrogram. A sampling unit hierarchy is comprised of levels determined by different sizes and shapes of sampling units: fine-scaled observations are a function of small quadrat sizes; coarser-scaled observations may be made using larger quadrats in the field or, as described here, by combining neighbouring quadrats to construct coarser-scaled composite sampling units (Gauch 1982). A classification hierarchy refers to the results generated by a hierarchical cluster analysis corresponding to a particular sampling scale, diagrammatically represented as a dendrogram. In addition, plant community patterns can be viewed from different reference points while maintaining the same observational scale (Bradfield and Orloci 1975, Allen and Starr 1982). For example, this may be accomplished by using a rectangular quadrat and altering its orientation.

In this paper, we present a methodological approach to exploring how vegetation patterns may shift with changes in sampling scale and dendrogram levels. We examine also the consequences of those changes for explaining total variation in species composition and environmental data. A sampling unit hierarchy was constructed by combining adjacent units together from a grid sampling design used to collect information on species distribution and abundance patterns in relation to environmental conditions. A cluster analysis of the vegetation data at each sampling scale yielded different classification hierarchies of the same vegetation for comparison. We address two questions related to vegetation classification: 1) What effect does altering the sampling scale have on the discernment of plant community patterns? and 2) What is the relationship between sampling scale, dendrogram level, and the proportion of variation explained in species composition and environmental factors? These questions were addressed using data collected from a subalpine wet meadow of south-coastal British Columbia.

Methods

Study area

The study site is a subalpine wet meadow located at an elevation of ca. 1700 m above sea level in the Black Tusk recreation area of Garibaldi Provincial Park (Fig. 1). Total mean annual precipitation recorded over a thirty year period (1951-1980) at a nearby weather station (Alta Lake, 648 m in elevation) was 1415 mm, almost half of which was snow (657 mm; Environment Canada 1980). Average monthly temperatures are mild during June-August (12.5 °C - 15.3 °C), although snow cover may persist well into the growing season. Freezing temperatures and deep snow are common from October to April (0.6 °C av-
Field sampling

Within the study area, percent coverage of plant species was estimated visually in 225, 0.5 m x 0.5 m quadrats systematically located at 5 m intervals in a 40 m x 120 m sampling grid. The sampling grid was positioned with the long axis running upslope, along the drainage gradient away from the lake’s edge. Quadrat elevations relative to the lowest point in the sampling grid were determined using a standard surveyor’s level. A total of sixty-five soil sample units were collected from the centre of every other quadrat throughout the sampling grid. Soil sample units were taken to a depth of ca. 15 cm, air dried at approximately 20°C (Jackson 1958, Davidescu and Davidescu 1982), and analyzed for pH, electrical conductivity (Rhoades 1982), carbon content (Leco analyzer), and texture (hydrometer method). These environmental variables were selected because they have been associated with subalpine plant communities throughout the Pacific coastal region of northwestern North America (Brink 1959, Brooke et al. 1970, Kuramoto and Bliss 1970, Douglas and Bliss 1977, Evans and Fonda 1990). Further, they have been shown to be influenced by soil moisture patterns under conditions similar to those in the present study area (Isard 1986, de Molenaar 1987).

Assembly of coarser sampling scales

Coarser scales of observation were constructed by combining neighbouring quadrats in the field sampling grid and calculating mean percent coverage values of species for larger composite sampling units (Gauch 1982). In this fashion, five aggregation designs of the basic quadrats (termed ‘agg’ scales), each representing a different scale of observation for the same vegetation, were generated. The agg1 scale consisted of the 225 basic quadrats (0.5 m x 0.5 m) used in the field sampling grid (i.e., no aggregation). The agg4 and agg9 scales consisted of non-overlapping aggregations of the basic quadrats in groups of four and nine, located within areas of 5 m x 5 m and 10 m x 10 m, respectively. Additionally, two orientations of a rectangular composite sampling unit, combining the quadrats in groups of six within a 5 m x 10 m area, were constructed: agg6a was assembled by aggregating quadrats in rectangular-shaped groups, oriented across the drainage gradient (i.e., perpendicular to the long axis of the sampling grid); agg6b was assembled in the same way but oriented parallel to the length of the sampling grid. In aggregating the basic quadrats to form coarser sampling scales, no attempt was made to nest the smaller scale sampling units at agg4 and agg6a/b within the larger units at the agg9 scale. Thus, the different sampling scales did not represent a spatial hierarchy in a strict sense. On the other hand, the reassembly of basic quadrats at the coarser sampling scales was considered advantageous for making cross-scale comparisons of the vegetation patterns.

Cross-scale comparison of plant community patterns

Delineation of plant community patterns within the sampling grid was performed using MISSQ (Ward 1963) with chord distance based on species coverage data as the quantitative measure of dissimilarity among quadrats (Or-loci 1967). MISSQ was selected from among other clustering methods because it produces homogeneous groups with minimum within-group and maximum between-group variability. Five MISSQs were performed, one on each of the vegetation data sets arising from the five aggregation designs. Sample sizes available for analysis declined with increasing scale given the aggregations of basic quadrats (for agg1, n = 225; for agg4, n = 48; for agg6a, n = 36; for agg6b, n = 32; for agg9, n = 24).

Evaluations of the total vegetation and environmental variation explained by the MISSQ results were determined using the method outlined by Jongman et al. (1987), as implemented in the forward selection procedure in CANOCO (ter Braak 1990). In this procedure, four samples were designated by dummy variables (0/1 values) indicating group membership of the respective sampling units. The dummy variables were then treated as nominal explanatory variables in a multiple least-squares regression procedure for determining the fraction of total species and environmental variation explained by the respective MISSQ groups. Only the top nine dendrogram levels, producing from two to ten clusters (referred to as levels C2-C10), were evaluated for their effectiveness in accounting for overall variation in the data. Dendrogram levels beyond the ninth were not examined given the increase in pattern detail and limited practicality for field interpretation of the clustering results. Estimates of the fraction of total species variation explained by the MISSQ groups, regardless of sampling scale at which the MISSQs
were performed, were made with reference to the species composition data at the aggl scale. This method avoided the problem of basing the evaluations on different vegetation data sets, thus maintaining consistency in making cross-scale comparisons.

The vegetation classifications at each sampling scale also were assessed for their ability to explain variation in the environmental data. In this case, the problem of missing soil data for alternating quadrats throughout the sampling grid was addressed by applying punctual kriging (Gamma Design Software 1991) to estimate values of the five soil variables (pH, conductivity, carbon content, percent sand, and percent clay) for the unsampled quadrats. The kriged estimates were determined through application of semivariance and anisotropic best fit model analyses based on measured values in the local neighborhood. The complete environmental data set used for evaluating the vegetation classifications thus consisted of elevation plus the measured and kriged estimates of five soil variables for the 225 quadrats of the sampling grid.

Sampling grid maps, showing the vegetation-types corresponding to the different combinations of sampling scales and dendrogram levels, were drawn to facilitate visual comparison of the plant community patterns. A corresponding map for the environmental data displayed variation within the sampling grid along the first axis from PCA of the six environmental variables.

Results

A total of 35 vascular plant species and one, common species of sphagnum moss were recorded within the sampling grid. The most frequently recorded species were Carex nigricans (90% frequency), C. spectabilis (50%), Erigeron peregrinus (48%), Caltha biflora (44%), Epilobium anagallidifolium (43%), and Juncus sp. (40%). Sixteen species occurred in fewer than 10% of the quadrats. Two species of low shrubs, Cassiope mertensiana (24%) and Phyllodoce mertensiana (18%), occurred with intermediate frequency.

For all sampling scales, the variation in species composition explained by the MISSQ results rose with increasing levels of refinement in the dendrograms (i.e., from C2 to C10; Fig. 2a). The steepest rise occurred from C2 to C3. In general, the variation explained was consistently highest at the aggl scale suggesting that vegetation pattern tends to be fine grained. Total vegetation variation explained by the classifications remained low (<35% variance explained by ten clusters at the aggl scale) indicating the presence of considerable noise in the vegetation data; thus, boundaries between plant communities in the meadow may not be clearly defined.

In contrast to the vegetation results, the coarser sampling scales (agg4-agg9) explained a greater proportion of environmental variation than the aggl scale implying that environmental heterogeneity is coarse-grained (Fig. 2b). Also, the total environmental variation explained (ca. 75% by dendrogram level C10) was higher than that for vegetation. Because of the smaller number of environmental variables analyzed (i.e., 6 environmental variables versus 36 plant species) and the relatively consistent changes in environmental conditions along the elevation gradient, these data may have contained less noise. The steep rise in variance explained from two to three groups for most sampling scales suggested the occurrence of three general environmental zones within the sampling grid. The agg6b sampling design yielded the highest levels of environmental variation explained over the range of three to five groups suggesting a concordance between vegetation and environmental pattern in a direction parallel to the main drainage gradient.

Figure 2. Variation explained in species composition (a) and environmental data (b) by vegetation patterns corresponding to dendrogram levels C2-C10 from MISSQ. Symbols denote sampling scales on which the vegetation classifications were based: aggl (●), agg4 (○), agg6a (△), agg6b (●), agg9 (□).
The sampling grid maps of the MISSQ results indicated that similar vegetation patterns were detected across a range of sampling scales (Fig. 3). Greater pattern detail was evident at the agg1 scale, but the general features were similar to those at the coarser sampling scales. At the three-group level, three vegetation zones were evident along the length of the sampling grid: an upper zone (labeled 1 in Fig. 3a-e) characterized by sedges (Carex nigricans, C. spectabilis) and forbs (Eriogonum peregrinum, Lupinus latifolius, Valeriana sitchensis); a middle zone (2) characterized by low shrubs (Cassiope tetragona, Phyllodoce empetriformis) and herbs (Caltha biflora, Carex nigricans, Eriogonum peregrinum, Luetkea pectinata); and a lower zone (3) dominated by a sedge (Carex nigricans), with interspersed forbs (Caltha biflora, Leptrhenna pyrolifolia).

At the five-group level, two sub-groups of the upper zone (labeled 1 and 2 in Fig. 3f-j) and lower zone (4 and 5) were recognized. The middle zone (3) was the same as that at the three-group level. The general configuration of the pattern boundaries and the disjunction within the upper zone suggested that processes determining vegetation pattern become more complex with increasing distance from the lake.

A surface map of sampling unit first axis scores from the PCA performed on the environmental data indicated a complex gradient running along the length of the sampling grid (Fig. 4). The first PCA axis explained 60% of the variance in the environmental data, and was positively correlated, in a direction toward the lake, with increasing pH (Pearson r = 0.83) and percent clay (0.89), and negatively correlated with elevation (-0.94), electrical conductivity (-0.83), and percent sand (-0.74). The second PCA axis was correlated with percent carbon (-0.91); however, the small eigenvalue for the second axis (19% of variance explained) indicated that its usefulness for interpretation was limited. Apart from a general conformity between vegetation and environmental zones along the length of the sampling grid, a precise matching of the environ-
The results of this study suggest there may be merit in aggregating small quadrats to form larger composite units prior to statistical analysis where the objective is pattern detection (sensu Kenkel et al. 1989). Regardless of dendrogram level used to delineate plant communities, a notable effect of sampling scale was the pronounced decline in apparent patchiness of the vegetation between the aggregate scale and all higher scales (Fig. 3). By aggregating adjacent quadrats, much of the fine-scale patchiness disappeared, and a spatial pattern emerged that remained relatively stable with further increases in sampling scale. Whereas the use of small quadrats (i.e., 0.5 m x 0.5 m) facilitated the recording of accurate data on species abundance in the field, the analysis of larger composite sampling units led to a clearer synthesis of the major vegetation patterns within the sampling grid. As noted by Kenkel et al. (1989), specific decisions relating to sampling will depend on study objectives. Nonetheless, we feel that the exploratory approach described here may aid in the recognition of plant communities at coarser scales that were otherwise noisy and more difficult to interpret at finer scales.

Analysis of larger sampling units also revealed that abiotic factors affecting vegetation heterogeneity may be operating at spatial scales considerably larger than the sizes of individual plants. In a related study of vegetation pattern in a New Zealand subalpine meadow, Weir and Wilson (1987) also reported successful detection of mesoscale pattern using small quadrats (0.25 m²). In conjunction with our findings, these results point to a strong identity of subalpine meadow vegetation types at very fine scales. On a cautionary note, the extensive use of kriged values in our analysis of environmental data may have contributed to the higher variance explained at the coarser sampling scales. These kriged estimates, despite being of the same grain size as the observed data (Legendre and Legendre 1998), tend to smooth over the actual variation at fine scales. Nonetheless, the relatively continuous nature of the major environmental gradient along the length of the sampling grid suggests that fine scale variation in abiotic conditions is not pronounced.

The slight increase in environmental variation explained by the vegetation classification at the aggregate sampling scale is in agreement with early claims that rectangular quadrats on sloping terrain should be positioned at right angles to observed vegetation zones (Clapham 1932, Bormann 1953). This conclusion, however, remains speculative owing to the limited treatment of rectangular quadrat orientation in this study. In our study site, vegetation pattern may have been strongly influenced by changes in abiotic conditions along the length of the sampling grid...
merit in the units pattern of density, a no-decline aggregating adjacent relaxed scale. 0.5 m abundance vegetation by sampling, we aid scales represent the relatively low variation in species composition explained by cluster analysis suggested that plant community boundaries are not sharp in the subalpine meadow study site. In part, this problem arises from the tendency for natural vegetation zones to be quite noisy, with boundaries rarely being clearly defined (Minchin 1987, Fortin 1999). Using the MISSQ results for their ability to explain variation in the environmental data provided a basis for selecting a classification that was consistent with a potential causal mechanism underlying the vegetation pattern (i.e., species responses to abiotic gradients). While the precise location of boundaries cannot be determined with this method, it does provide guidelines for determining the optimum number of vegetation units, irrespective of details of the boundaries of such units. This method, however, is not intended to manipulate scale with the express purpose of trying to maximize the degree to which vegetation patterns and environmental variables are correlated. Rather, it should be viewed as a heuristic process for gaining further insight into potential determinants of vegetation pattern which may, in turn, lead to more focused hypotheses about the causes of variation.

A multi-scaled approach such as described here offers the potential for different perspectives to be gained on the underlying processes that determine plant community structure. From a management viewpoint, this could lead to increased options for decision making. For example, in the subalpine meadow, if the maintenance of broad scale vegetation patterns were of prime concern, then interest should center on the three general zones (lower, middle, and upper) represented in the sampling grid. Alternately, if a finer scale of resolution addressing particular features of vegetation heterogeneity is required, then a higher level of resolution in the dendrogram corresponding to the finest sampling scale may provide a more suitable template. Consideration of the scales over which physical processes influence plant communities provides a basis for improved understanding of complex ecosystems.

Acknowledgements: This research was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada to G.E.B., and The Vladimir J. Krajina Scholarship in Plant Ecology to P.B.D. We thank the British Columbia Ministry of Parks for granting permission to work in Garibaldi Provincial Park. Dr. L.M. Lavkulich and Mr. B.W. von Spindler, Department of Soil Science, University of British Columbia, provided advice and facilities for the laboratory analysis of soil sample units. We are especially grateful to Dr. Wei Zhang for assistance with data analysis. Final revisions to the manuscript were guided by comments from Dr. Janos Podani and three anonymous reviewers.

References


Evans, F.C. 1952. The influence of size of quadrat on the distributional patterns of plant populations. Contributions from the
Laboratory of Vertebrate Biology of the University of Michigan 54:1-15.


