A Quantitative Method

For Building Systems of Classification

Ian S. Moss[1], R.P.F.
May 18, 2003

The aim of statistics is to reduce the abundance of situations and to provide a summarized description of them that can be remembered and used as a basis for action.

Alain Desrosieres[2], 1998

Introduction

This paper outlines the quantitative methods used to derive a stand structure classification (ForesTree Dynamics Ltd. 2003) using data collected from over 400 plots representing Douglas-fir and lodgepole pine stands throughout the Lignum Limited Innovative Forest Practices Agreement Area (IFPA). The IFPA consists of several large parcels of land covering a total area of approximately 600,000 hectares. These lands are located within the Central Interior of British Columbia within the vicinity of Williams Lake, 100 Mile House, Alexis Creek and Horselfy. They occur predominantly within the Sub-Boreal Spruce Pine (SBPS), Interior Douglas-fir (IDF) and Sub-Boreal Spruce (SBS) Biogeoclimatic Zones.

The primary purpose of developing a stand structure classification is fascilitate more precise communication amongst foresters, biologists and ecologists on the subject of differences between and similarities amongst different kinds of stands. For even aged stands, characteristics are easily gauged in terms of dominant tree age, height, stand density and basal area or quadratic mean diameter, and these attributes can be conveniently represented in Stand Density Management diagrams for monocultures. In natural forests, and managed forests that have retained natural characteristics, such definitions are plagued by the occurrence of irregularities in stand structures, and uncertainty about when an even-aged stand can no longer be classified as such. The alternative classifications, multi-layered, irregular or complex stands, are vague and difficult to quantify. It would be better if we could be more precise, and still succinct in describing differences in stand structures, particularly as they pertain to distributions in tree sizes along one axis and tree species along another.

The challenges in producing a system for classifying stand structures are:

---

[1] ForesTree Dynamics Ltd., 1458 Kingsview Road, Duncan, B.C. CANADA V9L 5P1.
Phone: 250 709-2199. email: moskyn@shaw.ca

[2] The politics of large numbers. P. 13.

1. The resultant units of classification must have internal consistency – that is the means by which stands are assigned to classes ensures that stands within a class are more nearly similar to each other than they are to other members outside the class.

2. Practitioners must easily recognize the resultant units, at least within one unit of differentiation. If the differences cannot be easily recognized then the result will be more rather than less confusion.

3. The resultant units must be sufficient in number to allow for a wide variety of interpretations for a wide variety of purposes. Criteria 2 and 3 capture the dilemmas of "lumpers" and "splitters".

4. Stand structures invoke the notion of both spatial and temporal continuums so that there is no truly optimal subdivision. In this context, the act of classification is an act of a setting a standard for communication for the purposes of convenience and consistency.

5. There should be one definitive method for classifying plots or whole stands such that disagreements around the classification of a particular instance can be resolved definitively. This is a (quantitative) standard for comparison that many systems of classification used in forestry fail to meet.

6. The attributes used to define stand structures at the top end of the classification should be easy to measure or assess making the classification generally applicable to a broad scale (i.e. inventory) of application. This is important for management interpretations that extend from forest, to landscape, watershed, stand and even tree level interpretations. Inconsistencies in management interpretations and decisions across these scales often arise from inconsistencies in the way in which stands are described depending on the scale.

7. The classification should be open to improvement with knowledge of more subtle differences that have significance for interpretation and management. Such differences are best incorporated by: a) creating further subdivisions at the bottom end of the classification hierarchy using enhanced criteria for subdivision, b) redefining the locations of the splits at the upper end using the same criteria, or c) both of the above.

17 Stand Structure Types were defined using the methods described herein. These types include, young and old stands, as well as even and uneven-aged stands. They are defined on the basis of the similarities amongst plots and differences between them with respect to their cumulative distributions of numbers of trees and basal areas per hectare, starting from the largest diameter trees and working toward the smallest diameters trees. For this reason, the methods used herein have been referred to as the (reverse) Cumulative Distribution Approach to Stand Structure Classification.

The stand structure classification was initially built to be independent of species composition, but such differences could be easily incorporated into the system through extension of the methods described herein. The advantage of not including them is that the classification remains stable regardless of species composition; species composition then becomes an independent source of classification that when combined with stand structure classification provides a full definition of stand types.

Methods

The methods used herein are non-parametric using local neighbourhood search, steepest descent, swapping techniques that are consistent with the principles of using graph theoretical, dynamic programming techniques to solve combinatorial problems. Such procedures are commonly used when the time to search for optimum combinations increases exponentially (rather than as a polynomial) with the number of cases (Papadimitriou and Steiglitz, 1998). These procedures produce results that are close to optimum, but not necessarily optimal. A number of techniques can be used to more approach an optimal solution (including running the algorithm more than once) but these are described herein. The exact procedures used were developed independent of the work of others. Given the substantial amount of literature on problems of optimization and dynamic programming, the procedures developed herein may already have been produced by others. I am simply not aware of such instances.

The following procedures were used to build the stand structure classification:

1. Construct (reverse) cumulative distributions by plot (observation) for each of trees per hectare and basal area per hectare starting with a tree of diameter at breast height (dbh) of 140 cm and progressing toward a tree of 0 cm in 1 cm decrements.

2. Normalize the cumulative distributions to numbers between 0 and 1: Rank the total basal area per hectare greater than 0 cm dbh from 0 (0 m^2 ha^{-1} by definition) to N observations. Convert the Rank to a Percentile (where N = 1 by definition). Do the same for trees per hectare. Convert the cumulative distributions at each diameter threshold to percentiles using table interpolation with respect to each of basal area per hectare and trees per hectare to derive a percentile.

3. Construct a distance matrix for every plot pair by summing up the differences in trees per hectare and basal area per hectare at each 1 cm interval from 0 to 140 cm. Find the Euclidean distance between each pair given the sums of the normalized differences in trees per hectare and basal area per hectare. Ensure that the resultant matrix is transitive. The matrix is symmetric.

4. Randomly assign each of the plots to 1 of M numbers of groups, where M = 17 (in this case).

5. Sum the distances between each pair of plots to estimate the Total (T) distance.

6. Sum the distances for each plot, to the remaining plots in each group to estimate the Within Group (WG_{ij}) distance (i plot, in j group).

7. Estimate the Between Group (BG) distance for each plot ($T – WG_{i,j}$).

8. For each plot estimate the change in Between Group (ΔBG_i) distance if the plot is transferred to one of the remaining X groups ($\Delta BG_i = WG_{i,j}$).

9. For each plot estimate the Within Group ($WG_{i,j=1..M}$) distances associated with the transfer into each of j = 1 to M groups. The WG distance for the group to which the plot is currently assigned remains unchanged.

10. For each plot-x-group combination calculate the Ratio, R, where $R_{i,j} = WG_{i.j.} : \Delta BG_i$.

11. Find the minimum, $RMIN_{i,j}$ for i = 1…N, j = 1 …M.

12. If R_{min} is ≥ 1 then Stop. Note that by definition, the transfer of plot from its current group to the same group produces a value, R = 1 (see steps 8 and 10).

13. Transfer plot “i” to group “j” for $R_{i,j} = RMIN_{i,j}$.

14. Go to Step 6 (note that the matrix update procedure could be made more efficient by simply updating the matrix where changes have been invoked).

Results

The complete results are too extensive to present herein. These include:

1. A complete enumeration of the plot data used in the analyses.
2. The resultant ranking and normalization figures used to transform the initial plot estimates of trees per hectare, basal area per hectare greater than or equal to a given diameter threshold.
3. The final distance matrix used to derive the classification.

The nature of the algorithm used to derive the classification is discussed in more detail below.

The algorithm was run once, taking 526 iterations to classify 422 plots into 17 groups before coming to a stop. The run-time was over 8 hours. The RMIN ratio fluctuated considerably from one iteration to the next (Figure 1) but showed a steady increase over the course of the entire process. However the total within group distance showed a steady (monotonic) decline, with the biggest gains made at the beginning of the simulation (Figure 2). Similarly (and as expected) between group distances showed a steady increase (Figure 3). The algorithm had been run repeatedly on a precursor dataset with the same trends being exhibited as those described above, but with some random starts producing better results (lower total within group distances, higher between group distances) than others.

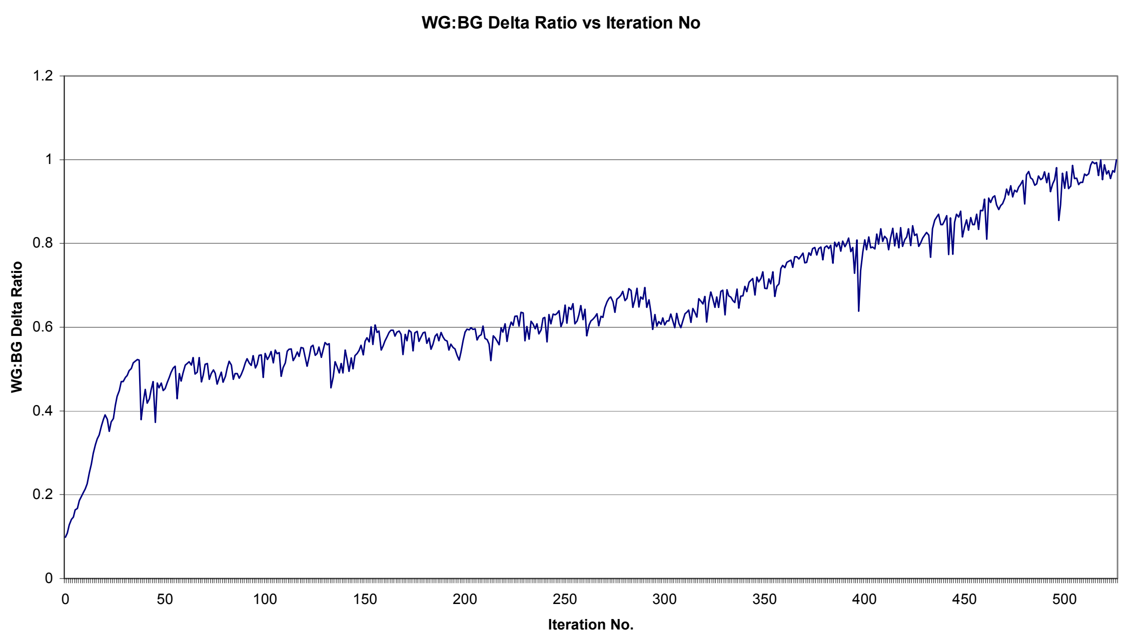

Figure 1. RMIN ratio versus iteration number.

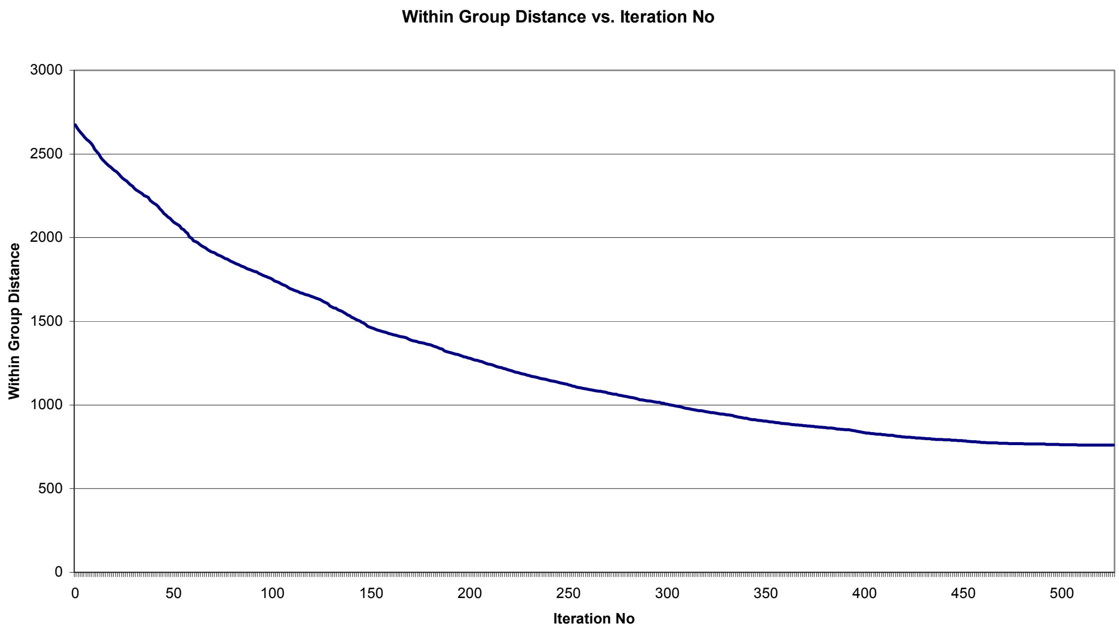

Figure 2. The sum of all within group distances versus iteration number.

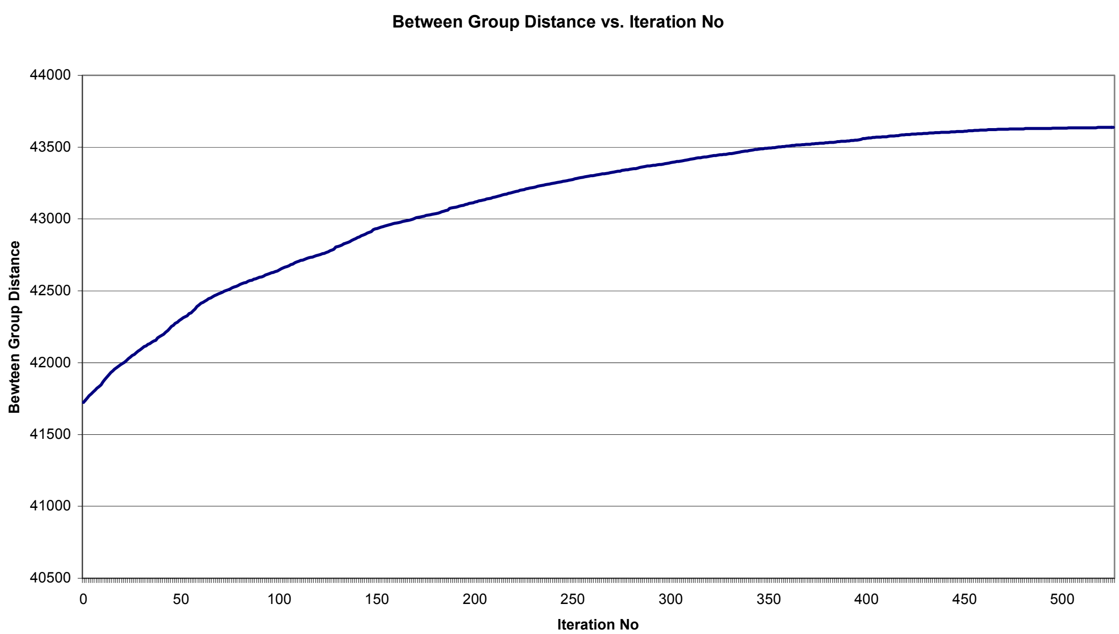

Figure 3. Between group variation versus iteration number.

Discussion

The Realism of Aggregates

… how can collective objects – aggregates of individuals – be made to hold? Extending far beyond the history of statistics, it spans the history of philosophy and social sciences, from Occam's opposition to realism and nominalism to Dumont's opposition to holism and individualism. I have chosen here to follow it through debates on statistical mean, especially as formulated by Quetelet, centering on the all-unifying idea of constant cause. The use of this tool by doctors and hygienists is shown that beyond ontological controversies, the controversies of aggregation, whose various justifications and supports depend on circumstance, find their meaning within the framework of the practices they account for. When the actors can rely on the objects thus constructed, and these objects resist the tests intended to destroy them, aggregates do exist – at least during the period and in the domain in which these practices and tests succeed.

Alain Desrosieres[3], 1998

The criteria for success in this case were:

1. Those observers in the field easily recognize the differences between the aggregates.

2. That the groups be internally consistent with respect to their characteristics.

3. That each group represents a reasonable proportion of the population, such that "outliers" are accommodated within a larger group rather than separated into groups on their own, and dense portions of population-space are subdivided into smaller groups rather than aggregated into one big group.

The meeting of these criteria was obtained by way of the methods used to derive the groups.

The first criterion depended on the number of groups that are chosen to represent the range of forest conditions, as well as on the degree to which the data[4] spans the breadth of outcomes exhibited in any population as a whole. In this context, the region chosen for developing the classification has a wide range of stand conditions that commonly occur within the landscape, and the data used to develop the classification included large numbers of both randomly located and subjectively located plots. In more uniform landscapes, certain stand conditions would not occur and therefore would not be integrated into the classification system using the same set of procedures.

[3] The politics of large numbers. P. 101.

The second criterion was enforced through construction of the distance matrix and the use of the local search algorithm. In particular, numerical integration was used across short intervals (1 cm diameter thresholds) to sum up the differences (distances) between (reverse) cumulative distributions for each of basal area per hectare and trees per hectare. Equal weight was given to these two dimensions through normalization. Differences in basal area per hectare tend to predominate in the analyses, with differences in trees per hectare becoming more important with small trees at the upper ends of the cumulative distribution. The two are of course highly correlated. The use of reverse cumulative distributions smothers the differences that exist between stands with trees in one diameter class but not the next. The extent to which such differences become important is adjudicated in this case by the size of the dataset and the chosen number of groups. The groups exhibit reasonable consistency (not maximally consistent given the nature of the algorithm) given the chosen number of groups used to represent the population.

The procedure to normalize the data by way of ranking the total basal area per hectare and total trees per hectare, has the effect of drawing outliers toward the centre and of increasing the dispersal of plots within dense portions of sample space. This contributes to the production of more groups where there are more samples and ensures that outliers are not assigned to their own groups.

The form of the algorithm also contributes to meeting the third criteria. Within group distances (and so too, the change in between group distances ΔBG) are calculated as the sum of the distances from the subject observation (that being considered eligible for moving to another group) to all of the other observations in the candidate group. This is in contrast to more standard approaches that rely on minimizing the within group *variation* relative to the mean distance for each pair of observations within a group. The net effect of this calculation, is that small groups tend to "attract" observations in the process of minimizing total within group distances and maximizing the change in between group distances (i.e. there is a greater potential for R_{ij} to be minimized). This "attraction" is diminished however, if the candidate group, albeit consisting of only a few observations, but with observations that are in fact far apart. The net effect is as follows: dense portions of the population sample space tend to accumulate a larger number of samples, but as they do so, tend to make more dispersed portions of the sample space more attractive. It is this compensatory process that tends to produce reasonable group sizes (along with the normalization procedure) for the purposes of covering the entire sample space. The process limits groups from getting too large (providing too general a coverage of the sample space) and from getting too small (providing delineations that are too specific for practical use).

Perhaps the foregoing algorithm falls within Kohonen's (1997) definition of "Unsupervised Classification" (we do not know the classes *a priori*) and more specifically "Simple Clustering" such that "Determination of the subsets A_j is a global optimization problem, whereby a set of simultaneous algebraic equations describing the optimality criterion must be solved by direct or iterative methods". The solution developed herein was predicated on a local search algorithm producing a good solution. It may be possible to convert this to a global optimization problem simply by minimizing the sum of within group distances, but this seems unlikely; it is more likely that this problem does not lend itself to a Polynomial-time solution, i.e. that the solution to the problem increases exponentially with the number of observations and with the number of groups up to a maximum[5].

The algorithm presented herein provides an approach to the development of classification that at the very least has not been commonly applied, and that may in fact be a novel contribution to the extensive set of procedures already available to approach these kinds of problems[6].

How good are the results? The algorithm had initially been run six times using a dataset that was slightly different from that of the final one. It was found that different random starts produced different results, but one of the runs could be identified as having produced the best results. Due to time constraints, several different random starts were not deployed with the finally derived distance matrix. Nevertheless, the results were taken to be sufficiently robust to produce a reasonable basis for classification since the improvements using several runs were small relative to the improvements obtained with the least efficient run of the initial six. The adequacy of the final results can be assessed visually by comparing the cumulative distributions for trees per hectare and basal area per hectare for all the plots within each pair of groups. It is clear, that different random starts result in changes in the centre of mass for each group, but the average differences between the groups in terms of the cumulative distributions tend to be very similar. Therefore, the issue becomes focussed on adopting one of the outcomes as the standard for comparison while acknowledging that there exists another standard that in all likelihood will perform (slightly) better than the one adopted.

From a broad perspective, the classification produces a useful starting point for describing differences in stand structures. It enhances our ability to classify stands by providing a quantitative means to definitively assign plots or groups of plots to specific units of classification. This can be achieved by using the initial set of classified plots as a standard reference set, with which all other plots may be compared. Alternatively, the average representation of the cumulative distributions for each group could be adopted as the standard reference set. The advantage of the first procedure is that in ensures that the classification remains consistent throughout, that is that plots used to build the system of classification are not reassigned to a new group based on a redefinition of the reference set.

Ultimately, the classification lends itself to modification without loosing the utility of a standard reference set. A new unit could be added, perhaps for special purposes by defining a additional reference set of (trees per hectare and basal area per hectare)

---

5 The number of combinations according to the age-old formula (Denny & Gains, 2000) is as follows: $\frac{n!}{i!(n-i)!}$ where "n" is the number of observations and "i" is the number of groups.

6 For one such review see Dillon and Goldstein (1984) and Moss (2000).

cumulative distributions. This could be used to redefine certain portions of the larger standard reference set that in turn could be used for further practical uses. The procedures developed herein provide a useful starting point for designing and redesigning stand structure classifications for a wide variety of purposes.

The process used for developing the classification could be used to further subdivide the groups or develop a whole new classification that incorporates differences in species composition. This can be done subdividing the trees per hectare and basal area per hectare dimensions (with respect to diameter thresholds) by species composition as part of the process of developing the initial distance matrix. Accordingly other qualitative attributes (dead and alive trees) could also be added. There is no limit to the number of additional factors that can be added. However, it is better to add these in a stepwise manner. If say 10 factors are included in the process of deriving the initial matrix, then the within group consistency with respect to any one of them will tend to be reduced relative to the inclusion of fewer factors. Of course it is possible to reduce the impact of this statement by extending the number of groups required to adequately describe the sample space, and to then explore hierarchical means for combining the groups for more general applications.

Conclusion

In conclusion, a method for quantitatively deriving stand structure was described herein. The method is useful for defining groups that can be easily recognized in the field and tend to have a high degree of internal consistency once identified. It may be extended to include consideration of factors other than trees per hectare and basal area per hectare greater than or equal to a given diameter threshold. This includes consideration for qualitative factors such as tree species composition and alive and dead trees. The data and the procedure can be used as a standard reference set and mechanism respectively for classifying new plots and stands represented by many plots, thereby avoiding much of the confusion that arises in more qualitative approaches to this problem. Although an exhaustive search has not been undertaken to find out if such techniques have been used elsewhere for similar problems, it seems reasonable to assert that the methods described herein are original in their contribution to solving problems of classification.

References

Denny, Mark and Steven Gains. 2000. Chance in biology. Princeton University Press, Princeton, New Jersey, USA.

Desrosieres, Alain. 1998. The politics of large numbers. A history of statistical reasoning. Translated by Camille Nash. Harvard University Press, Cambridge, Massachusetts, USA.

Dillon, William and Mathew Goldstein. 1984. Multivariate analyses. Methods and applications. Wiley series in probability and mathematical statistics. John Wiley & Sons, New York, NY, USA.

Kohonen, Teuvo. 1997. Self-Organizing Maps. Second Edition. Springer-Verlag, New York, NY, USA.

Moss, I.S. 2000. A review of cluster algorithms. Unpublished. Prepared in fulfillment of Forestry 531, A course from the University of British Columbia Faculty of Forestry on Quantitative Methods. ForesTree Dynamics Ltd., Duncan, B.C.

Papadimitriou, Christos and Kenneth Steiglitz. 1998. Combinatorial optimization. Dover Publications, Mineola, New York, USA.