QUANTIFYING FOREST STAND DIVERSITY USING STAND STRUCTURE

by

Christina Lynn Staudhammer

B.Sc., University of California (Davis), 1990

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF FORESTRY

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 1999

© Christina Lynn Staudhammer, 1999
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Forestry

The University of British Columbia
Vancouver, Canada

Date 18 December 99
Abstract

Stand structure is an important component of the overall description of biological diversity in a stand. The diversity of tree sizes within a stand affects tree growth and yield, and is highly correlated with the biodiversity of a stand.

Four measures of stand diversity based on the stand distribution of basal area by height, diameter, and species were proposed, assuming a baseline of maximum diversity corresponding to a uniform distribution. First, the Extended Shannon Index (ESI), a measure based on Shannon’s Index, was derived. Second, a measure based on fitting the univariate and bivariate distributions of diameter and height was investigated. The third measure, STRI, was derived as a modified R-squared, based on the fit of the distribution to that of a uniform distribution. The fourth measure, STVI, was derived based on comparing the variance of a distribution to that of a uniform distribution.

The four measures were evaluated with simulated data and inventory data from the Malcolm Knapp Research Forest (MKRF). The ESI and STVI resulted in logical orderings of both the simulated data and the MKRF data. The advantages of the ESI include its known sampling distribution and accepted use in forestry; however, the ESI depends on allocating data to arbitrary classes. The STVI does not require classifying data; however, its sampling distribution is unknown.
Since the measures derived require only normally collected inventory plot data, they are relatively inexpensive to use. However, a thorough investigation of the properties of each should be undertaken. Since evaluating structural diversity is becoming an increasingly important part of forest assessment, a quantitative measure is needed to measure diversity. These measures provide a starting point toward finding an inexpensive, practical measure of structural diversity which gives reasonable results.
# Table of Contents

**Abstract** ................................................................................................................................. ii

**Table of Contents** ....................................................................................................................... iv

**List of Tables** ............................................................................................................................... vi

**List of Figures** ............................................................................................................................... vii

**Acknowledgement** ....................................................................................................................... viii

**Chapter 1: Introduction** ............................................................................................................... 1

**Chapter 2: Common Diversity Measures** ................................................................................... 5

2.1 Shannon's Index ........................................................................................................................... 6

2.2 Simpson's Index ........................................................................................................................... 7

2.3 The Q Statistic ............................................................................................................................. 8

2.4 Relationships Among Indices ...................................................................................................... 9

2.5 Using Indices to Make Inferences About Populations ................................................................. 11

2.6 Discussion of Common Diversity Measures in Use ................................................................. 12

**Chapter 3: Structural Diversity Measures** .................................................................................. 16

3.1 Measures of Vertical Diversity ................................................................................................... 18

3.2 Measures of Horizontal Diversity .............................................................................................. 20

3.3 Other Structural Diversity Measurements .................................................................................. 22

3.4 Discussion of Measures .............................................................................................................. 24

**Chapter 4: Possible Measures of Stand Diversity Using Stand Structure** ................................... 26

4.1 Desirable Characteristics of a Measures of Stand Diversity Using Stand Structure .................. 27

4.2 Possible Alternative Measures of Stand Diversity Using Stand Structure ............................... 29

4.2.1 Shannon's Index Extended to Diameter, Height, and Species ............................................... 29

4.2.1.1 Post-hoc Method ............................................................................................................... 30

4.2.1.2 Combination Method ....................................................................................................... 31

4.2.1.3 Properties of the Extended Shannon's Index .................................................................... 32

4.2.2 Structural Index Based on Curve Fitting ................................................................................ 33

4.2.2.1 Univariate Curve Fitting ................................................................................................. 35

4.2.2.2 Bivariate Curve Fitting .................................................................................................... 36

4.2.2.3 Properties of a Structural Index Based on Curve Fitting ................................................ 37

4.2.3 Structural Index Based on R^2 (STRI) .................................................................................. 38

4.2.3.1 STRI Univariate Method ................................................................................................. 39

4.2.3.2 STRI Bivariate Method ................................................................................................. 41
4.2.3.3 Properties of STRI ......................................................... 42
4.2.4 Structural Index based on Variance Comparisons (STVI) ..................................................... 43
  4.2.4.1 STVI Univariate Method ..................................................... 44
  4.2.4.2 STVI Bivariate Method ..................................................... 49
  4.2.4.3 Properties of STVI ......................................................... 52

CHAPTER 5: EVALUATION OF ALTERNATIVE MEASURES OF STRUCTURAL DIVERSITY ........................................... 54

  5.1 SIMULATED TREE DATA ......................................................... 54
    5.1.1 Methods ......................................................................... 54
      5.1.1.1 Baseline Maximal Diversity ................................................ 54
      5.1.1.2 Simulated Stands of Varying Diversity ................................ 55
    5.1.2 Results and Discussion ....................................................... 62
      5.1.2.1 Shannon's Index Extended to Diameter, Height, and Species ........ 62
      5.1.2.2 Structural Index based on Curve Fitting ............................... 65
      5.1.2.3 Structural Index based on R² (STRI) ....................................... 71
      5.1.2.4 Structural Index based on Variance Comparisons (STVI) ............ 72
    5.1.3 Overall Discussion of Proposed Indices Computed with Simulated Data 76

  5.2 MALCOLM KNAPP RESEARCH FOREST DATA ............................... 76
    5.2.1 Methods ......................................................................... 77
    5.2.2 Results and Discussion ....................................................... 80
      5.2.2.1 Shannon's Index Extended to Diameter, Height, and Species ........ 80
      5.2.2.2 Structural Index based on Variance Comparisons (STVI) ............ 84
    5.2.3 Discussion of Extended Shannon Index and Univariate STVI Computed with MKRF Data ...................... 88

CHAPTER 6: CONCLUSION ........................................................................... 89

LITERATURE CITED ................................................................................. 95

APPENDIX I .................................................................................................. 101
List of Tables

Table 2.1. Summary of performance and characteristics of selected diversity measures. ........................................................................................................... 14
Table 5.1. Plot type abbreviations for simulated cases........................................... 60
Table 5.2. Extended Shannon indices computed using simulated stand data......... 64
Table 5.3. Resulting coefficients and RMSE for the univariate Weibull fitted to dbh using simulated stands. ................................................................. 66
Table 5.4. Resulting coefficients and RMSE for the univariate Weibull fitted to height using simulated stands. ................................................................. 67
Table 5.5. RMSE and resulting values of coefficient $a_3$ for the bivariate Weibull fitted to simulated stands................................................................. 69
Table 5.6. Resulting coefficients $b_1$ and $b_2$ for the bivariate Weibull fitted to simulated stands................................................................. 69
Table 5.7. Resulting coefficients $c_1$ and $c_2$ for the bivariate Weibull fitted to simulated stands................................................................. 70
Table 5.8. Univariate (STRI$_{d+h}$) and bivariate (STRI$_{dh}$) STRI using simulated stands. ................................................................................................... 71
Table 5.9. Univariate (STVI$_{dbh}$, STVI$_{height}$, and STVI$_{d+h}$) and bivariate (STVI$_{dh}$) STVI using MKRF data................................................................. 85
Table 5.10. Average sample variances for simulated cases compared to the variance of a theoretical uniform distribution................................. 73
Table 5.11. Data for height-dbh equation development by species......................... 78
Table 5.12. Descriptive statistics of selected MKRF plots...................................... 80
Table 5.13. Extended Shannon indices computed with MKRF data...................... 82
Table 5.14. Univariate (STVI$_{dbh}$, STVI$_{height}$, and STVI$_{d+h}$) and bivariate (STVI$_{dh}$) STVI using MKRF data................................................................. 85
Table 5.15. Sample variances and average sample variances by species for chosen MKRF stands compared to the variance of a theoretical uniform distribution... 87
List of Figures

Figure 4.1. STRI and three equal dbh ranges. .................................................................44
Figure 4.2. Comparison of STVI values with \( m=1 \) and varying values of \( p_1 \) and \( p_2 \). 47
Figure 4.3. Comparison of STVI values with \( p_1=p_2=2 \) and varying values of \( m \) ....48
Figure 5.1. Possible combinations of dbh versus height. .............................................57
Figure 5.2. Graphical description of simulated dbh and height values. ......................59
Acknowledgement

I would like to extend thanks to the many people that helped me to successfully complete this thesis, especially my supervisor Dr. Valerie LeMay, and my thesis committee, Drs. Tony Kozak, Guillaume Therien, Rick Routledge (Department of Statistics, Simon Fraser University), and Peter Marshall. Each and everyone one of them contributed significantly to my education and my experience at U.B.C., and I will not forget their generosity.

I would also like to thank Dr. LeMay for helping me to find funding for this project and securing the data necessary to complete this project.

Last, I would like to thank my family and friends for their amazing support during this time in my life, and extend an especially warm thanks to William Kovacs, whose love of the forest was my first inspiration. Thanks for letting me grow.
Chapter 1

Introduction

Over the past two decades, the focus of the sustainable forestry debate in North America has been increasingly drawn toward the concept of diversity. The maintenance of forest diversity has been linked to robust bird and insect populations, and is seen as an indicator of the wellbeing of a biological system (Magurran 1988). Interest in forest diversity has prompted both the Canadian Council of Forest Ministers (Canadian Standards Association 1996) and the United States federal government (U.S. federal law 36 Code of Federal Regulations Sec. 217.27) to require forest managers to conserve biological diversity at natural levels.

The term ‘biodiversity’ has gained wide acceptance in describing biological diversity, encompassing the total variety of a biological system, including species diversity of plants, animals, fungi and bacteria, as well as the genetic and individual variation that exists within those species (Dobson 1996). Biodiversity may be evaluated at many levels: over an entire ecosystem, in a forest stand, or within a species at the genetic level. At the stand level, biodiversity has been defined to include not only species diversity, but also the variation of size, age, and genetic composition of the stand (Lähde et al. 1999). Thus, the definition of stand diversity may include not only the composition of species, but also the size distribution of individuals of those species.
Along with governmental and non-governmental directives to conserve biological diversity comes the necessity to measure biological diversity. For example, U.S. federal law (36 C.F.R. Sec. 219.26) requires that inventories include “quantitative data making possible the evaluation of diversity”. The evaluation of diversity, however, has long been a subject of heated debate. Many measures of diversity, in the form of indices, are in use in forestry and ecology; however, none are without critics (Magurran 1988). It has even been asserted that since there is no precise accepted definition of biological diversity, there can be no effective quantitative measure (Hurlbert 1971, Silbaugh and Betters 1995).

Several traditional biodiversity indices, such as Simpson’s index (Simpson 1949) and Shannon’s index (Shannon and Weaver 1949) have gained wide acceptance in forestry as quantitative measures of species diversity (Swindel et al. 1984, Magurran 1988, Lewis et al. 1988, McMinn 1992, Silbaugh and Betters 1995). These indices rely on the distribution of individuals by species to arrive at a quantitative value of biological diversity. However, because of the variation that exists in tree size, traditional diversity indices are not entirely suited to the measurement of forest stand diversity (Lähde et al. 1999). Furthermore, constraints of time and money limit the forest manager to measuring only a subset of species represented in a single sample plot of forest.

Surrogate measures for ecological composition are unavoidable, as the number of existing species in forested ecosystems is vast, and there are inherent difficulties in measuring them. To help remedy this, the use of indicator species has been proposed
Desirable criteria for species used as biodiversity indicators include (Noss 1990):

- Wide distribution across regions
- Potential to act as surrogate for other species
- Inexpensive and easy to monitor
- Pre-existing data available
- Biology of species reasonably known
- Sensitive to forest management disturbances
- Sufficient sample size attainable
- Population dynamics stable

Since trees meet these criteria, a useful measure might be derived using data from forest inventories or other regular forest surveys (LeMay et al. 1997, Lähde et al. 1999).

Stand structure has been defined in terms of the mixture of tree sizes and species, without explicit recognition of other flora and fauna (Buongiorno 1995). The diversity of stand structure indicated by the mixture of tree size and species has been shown to be an important element of total biological diversity (MacArthur and MacArthur 1961, Willson 1974, Franzreb 1978, Temple et al. 1979, Aber 1979, Ambuel and Temple 1983, Freemark 1986), with the highest diversity found in stands where there are multiple species and multiple tree sizes (Buongiorno et al. 1994). It has been argued that managing forests for biodiversity can be accomplished by managing for structural diversity (Önal 1997). Thus, a useful measure of the total diversity of a forest stand may be found by measuring the structural diversity of a stand.
Stand structure can be measured based on the distribution of tree ages and sizes (Lähde et al. 1999). Although age is difficult to measure quickly and cheaply, tree size may be readily measured by diameter at breast height (dbh) and height. An index of stand structure as a function of species composition, dbh, and height should be correlated with overall forest diversity, and would be a function of tree variables found in forest inventory samples.

The objective of this study is to present possible measures of stand diversity that quantify the diversity in a stand based on both species and size diversity, using commonly measured tree attributes. Several measures are suggested and discussed. The measures were then evaluated using simulated stand data, and applied to the University of British Columbia’s Malcolm Knapp Research Forest (MKRF), located in coastal British Columbia. The measures were evaluated using objective criteria and conclusions were drawn as to the performance of the suggested measures.
Chapter 2
Common Diversity Measures

The most common diversity measures in use are species diversity measures. Species diversity measures can be applied at various spatial and temporal scales, incorporating landscape patterns, as well as stand level variation (LeMay et al. 1997). Biodiversity is often broken up into three spatial components. The $\alpha$-component of the diversity of organisms inhabiting a variety of habitats is loosely defined as the average diversity within a single habitat (Routledge 1980). The $\beta$-diversity can be defined as the amount of turnover between habitat types (Routledge 1979). $\Gamma$-diversity describes the variation in $\alpha$ and $\beta$ -diversity as one moves across regional landscapes and along major environmental gradients (Kimmins 1997). This paper will focus on $\alpha$-diversity, that is, diversity at the stand level.

Species diversity can be broken into two elements. Species richness refers to the number of species. Species evenness, or species equitability, refers to how equally abundant the species are. From these definitions, species diversity can be classified into three categories (Magurran 1988):

1. Species richness indices are a measure of the number of species in a defined sampling unit.
2. Species abundance models describe the distribution of species as a function, or parameter of an associated function.
3. “Diversity indices” or “heterogeneity indices” are indices based on proportional abundances of species, combining richness and evenness into a single index.
There are many measures of species diversity (see Magurran 1988). The diversity measures chosen for discussion in this thesis are Simpson’s Index of Concentration, Shannon’s Information Index (which is often incorrectly referred to as the Shannon-Weaver or Shannon-Wiener index (Magurran 1988)), and the $Q$ statistic. Based on information theory, Simpson’s Index and Shannon’s Index are perhaps the most popular diversity measures. Both are nonparametric indices; that is, neither relies on any assumptions about the underlying species abundance distribution. The $Q$ statistic is a diversity measure derived from an abundance model.

2.1 Shannon’s Index

Shannon’s index is a measure of how uncertain we are that an individual picked at random from an infinitely large community will be of a certain species. The more uncertainty one has about the identity of an individual, the higher the diversity of the community. As well as assuming that the population is ‘indefinitely large’ and that individuals are selected at random, the calculation of this index requires that all species be represented in the sample. The failure to represent all species can be a source of considerable error (Magurran 1988). Shannon’s index, $H'$, is defined as follows (Shannon and Weaver 1949):

$$H' = \sum_{i=1}^{S} p_i \ln p_i$$  [2.1]

where: $p_i =$ proportion of individuals in the $i$th species; and $S$ is the number of species in the population.
When the population proportions are unknown, Shannon’s Index is calculated with the maximum likelihood estimator (MLE) of $p_i$:

$$p_i = \frac{n_i}{N}$$

[2.2]

where: $n_i$ = the number of individuals sampled of species $i$, and $N$ = total number of individuals sampled.

In practice, the terms $p_i$, $n_i$, and $N$ are not necessarily computed with the number of individuals, as individual trees, plants, or animals can vary greatly in size. The proportion of a species has been measured in a variety of ways, including: number of individuals (Franzreb 1978, Swindel et al. 1991, Niese and Strong 1992, Condit et al. 1996), basal area (McMinn 1992, Harrington and Edwards 1995), stems per ha (McMinn 1992, Harrington and Edwards 1995), foliar cover (Swindel et al. 1984, Lewis et al. 1988, Qinghong 1994, Corona and Pignatti 1996), crown cover (Corona and Pignatti 1996), and biomass (Swindel et al. 1984, Swindel et al. 1991).

Shannon’s Index is often applied in the form of a relative index. This relative index is the ratio of absolute diversity to the maximum diversity possible. This index can be taken as a measure of evenness (Pielou 1977):

$$E = \frac{H'}{H'_\text{max}} = \frac{H'}{\ln S}$$

[2.3]

2.2 Simpson’s Index

Simpson’s index of concentration, $\lambda$, is the probability that two individuals selected at random from an infinite population are of the same species (Simpson 1949). Simpson’s
index differs from Shannon’s in that Simpson’s index decreases as diversity increases. Thus, Simpson’s Index is often expressed as 1/λ or exp(λ). Simpson’s index is defined as follows (Simpson 1949):

\[ \lambda = \sum_i p_i^2 \]  

where: \( p_i \) = proportion of individuals in the \( i \)th species.

When a sample is chosen at random from a population, Simpson’s index can be computed with the following unbiased estimator (Simpson 1949):

\[ \lambda = \sum_i \frac{n_i(n_i-1)}{N(N-1)} \]  

where: \( n_i \) is the number of individuals of the \( i \)th species; and

\[ N = \text{total number of individuals sampled.} \]

As with Shannon’s Index, the terms \( p_i, n_i, \) and \( N \) are not necessarily computed with the number of individuals. The proportion of a species has been measured in a variety of ways, including: number of individuals (Swindel et al. 1984, Condit et al. 1996), basal area (Harrington and Edwards 1995), stems per ha (Harrington and Edwards 1995), foliar cover (Swindel et al. 1984, Lewis et al. 1988), crown cover (Swindel et al. 1991), and biomass (Swindel et al. 1984, Swindel et al. 1991).

2.3 The \( Q \) Statistic

The \( Q \) statistic can be considered to be a species richness index, but it is a statistic derived from the species abundance distribution of a particular population. The \( Q \) statistic
measures the inter-quartile slope of the cumulative species abundance curve, providing an indication of the diversity of the community without weighting toward the very abundant or very rare species (Kempton and Wedderburn 1978):

\[
Q = \frac{0.5n_{R1} + \sum_{r=1}^{R2-1} n_r + 0.5n_{R2}}{\log(R2 / R1)} 
\]

[2.6]

where: \( n_r \) is the total number of species with abundance \( r \);

\( R1 \) and \( R2 \) are the 25% and 75% quartiles, respectively; and

\( n_{R1} \) and \( n_{R2} \) are the number of individuals in the classes where \( R1 \) and \( R2 \) fall, respectively.

This approach takes into account the distribution of species abundances, but does not actually entail fitting a model.

### 2.4 Relationships Among Indices

Hill’s (1973) indices provide an elegant relationship between \( S, \lambda, \) and \( H' \). Hill’s indices were derived from the observation that these indices can be arranged by their propensity to weight toward rare species or common species (i.e., toward species richness or evenness).

Hill (1973) defined a diversity index as “reciprocal mean proportional abundance”, with the \( a \)th order of diversity given by:

\[
\widetilde{N}_a = \left( \sum p_i^a \right)^{(1-a)} \quad \text{for some } a > 0 
\]

[2.7]

where: \( p_i \) = proportion of individuals in the \( i \)th species.

It has been proven that if an index of diversity, \( \widetilde{N} \), satisfies the following properties, it must be one of Hill’s indices (Routledge 1979):
H1: $\bar{N}(1/S, 1/S, \ldots, 1/S) = S$. For a population with a given number of species, the measure of diversity will be at a maximum when all the species are present in equal proportions (or with maximum evenness).

H2: $\bar{N}(p_1, p_2, \ldots, p_S) < S$

H3: $\bar{N}$ is a continuous function in each $p_i$, for $p_i$ in $[0, 1]$.

H4: Species diversity = genus diversity times average species diversity within genera, where the average should satisfy:

1. Consistency: If all the genera are equally diverse, the average is $\bar{N}$.

2. Strict monotonicity: If the diversity of any genus is increased, the average increases.

3. Quasi-linearity: Averages can be performed in stages.

4. Homogeneity: If the diversity of each genus was increased by a factor $c$, then the overall diversity should be increased by $c$.

5. Symmetry: The average will be independent of the ordering of the genera.

$S$, Simpson’s Index, and Shannon’s Index satisfy the above properties, as:

\[
\begin{align*}
\bar{N}_0 &= S \\
\bar{N}_1 &= \exp(H') \\
\bar{N}_2 &= \frac{1}{\lambda}
\end{align*}
\]

[2.8]

Several authors (Routledge 1980, Kobayashi 1981) advocate the use of Shannon’s and Simpson’s indices in Hill’s form, as they are more stable and less sensitive to sample size.
2.5 Using Indices to Make Inferences about Populations

Since diversity indices attempt to compress a multidimensional description into a single number, they can be ambiguous when making comparisons between populations (Hurlbert 1971). However, the effects of such ambiguity can be greatly reduced by restricting the use of indices to comparisons of similar collections of species (Zahl 1977). The comparability of two communities will depend on: 1) taxonomic definition, 2) temporal scale, and 3) the region or space that the community occupies (Pielou 1975).

Differences in estimated diversity from samples of two different populations can result from three effects (Smith and Grassle 1977):

1. true differences in population diversity;
2. sample size dependent bias of the diversity estimator; or
3. random sampling error.

It may be difficult to separate these effects, particularly when the sample size is small and the bias is large (Smith and Grassle 1977).

The traditional method for testing for significant differences in diversity indices is a $t$-test, here shown with Shannon’s Index (Hutcheson 1970):

$$ t_{(d.f.)} = \frac{H'_1 - H'_2}{(\text{Var} H'_1 + \text{Var} H'_2)^{1/2}} $$ \hspace{1cm} [2.9]

$$ \text{d.f.} = \frac{(\text{Var} H'_1 + \text{Var} H'_2)^2}{(\text{Var} H'_1)^2 / N_1 + (\text{Var} H'_2)^2 / N_2} $$ \hspace{1cm} [2.10]
This test, although conservative, may be used only if the populations are normally (or nearly normally) distributed (Hutcheson 1970), the population is randomly sampled, and the total number of species is known (Heltshe and Bitz 1979, Routledge 1980, Pielou 1966).

2.6 Discussion of Common Diversity Measures in Use

Diversity indices attempt to describe both the richness and evenness of a community. A major criticism of diversity indices is that they confound a number of variables that characterize community structure in combining richness and evenness (Ludwig and Reynolds 1988):

1. the number of species;
2. the relative species abundances; and
3. the homogeneity and size of the area sampled.

Despite these criticisms, they are widely used in forestry (e.g., Swindel et al. 1984, Lewis et al. 1988, McMinn 1992, Qinghong 1994, Corona and Pignatti 1996).

Pielou (1975) gives two reasons to be cautious in the use of Shannon’s index. First, it has been shown that Shannon’s Index, as it is commonly computed with equation [2.1], is biased. Although this bias can be removed, it requires that $S^*$, the number of species in the population be known; $S$ calculated from a sample is merely a lower bound for $S^*$. Secondly, using the MLE of $p$, makes an implicit assumption that the sample units are a random sample from a community (Pielou 1975). Although random locations of plots can be assured, the patterns of the organisms’ locations may not be random. In natural communities, organisms are more likely to be located in clumps or patches, and thus, this assumption will not be met.
One may adjust for these shortcomings by using a re-sampling technique, such as Pielou’s pooled quadrat method (Pielou 1966) or a jackknife estimator (Routledge 1980). However, there is a potential for severe bias using the pooled quadrat method (Routledge 1980). When a small number of quadrats is sampled, the jackknife has been found to give better estimates of $H'$ than traditional computational methods or bootstrapping methods (Smith 1984).

The relativized Shannon index, $E$, is often used as a measure of evenness in a community. However, $E$ contains the term $S$, which often is not known. Estimates of $S$ from a sample can vary greatly as a result of sample size or stochastic variation, greatly affecting the computation of $E$. Evenness will be overestimated when $S$ is underestimated and thus, this relative index can be misleading (Peet 1974). Species evenness should only be compared between communities when collection sizes are equal (Kobayashi 1981).

The major criticism levelled at Simpson’s Index is its insensitivity to rare species (Hurlbert 1971, Pielou 1975, Magurran 1988). However, this criticism does not recognize that diversity is an aspect of community structure, and that structurally, rare species are minor components of their community (Hurlbert 1971). If this viewpoint is taken, Simpson’s index may be a better representation of community composition than other diversity indices.

The choice of a diversity index may depend on the characteristic of a community that one wishes to measure and how proportions of each species ($p_i$) are computed. Simpson’s index
is most affected by changes in the abundances of the commonest species (dominance), whereas Shannon’s index is more affected by rare species (richness). However, this may be compensated for at least in part by choosing basal area or stems per ha to compute $p_i$. In forest stands, the use of basal area has been recommended, as it gives a better measure of the occupation of a stand (McMinn 1992). Furthermore, one may prefer to use an index with a biological interpretation. Although Simpson’s index can be described as the probability of inter-specific encounter, Shannon’s index lacks such an interpretability.

Kempton and Wedderburn (1978) recommend the $Q$ statistic in preference to Simpson’s or Shannon’s index for two reasons. First, $\lambda$ and to a lesser extent $H'$, are strongly influenced by the abundance of the commonest species, while $Q$ depends on a wide range of species with medium abundance. Second, $Q$ shows a smaller variability than $\lambda$ or $H'$ when testing replicate samples.

There is much debate on the discriminant abilities of diversity indices. Magurran (1988) offers a summary of the performance and characteristics of some selected diversity statistics. Table 2.1 contains a partial list.

### Table 2.1. Summary of performance and characteristics of selected diversity measures (Magurran 1988).

<table>
<thead>
<tr>
<th>Index</th>
<th>Type of index</th>
<th>Discriminant Ability</th>
<th>Sensitivity to Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$ statistic</td>
<td>Richness</td>
<td>Good</td>
<td>Low</td>
</tr>
<tr>
<td>$S$ (species count)</td>
<td>Richness</td>
<td>Good</td>
<td>High</td>
</tr>
<tr>
<td>Shannon</td>
<td>Richness</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Simpson</td>
<td>Dominance</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td>Shannon Evenness</td>
<td>Evenness</td>
<td>Poor</td>
<td>Moderate</td>
</tr>
</tbody>
</table>
Authors vary on their preference of indices. Simpson’s is preferred over Shannon’s by Peet (1974) and Routledge (1979), while Kempton and Wedderburn (1978) favour the $Q$ statistic. Magnussen and Boyle (1995) indicate that Shannon’s index is more efficient than the Simpson Index in terms of number of sample plots required; however, this is contrary to results found by Routledge (1979).

Species diversity measures can be applied at various spatial and temporal scales; however, this review focused on diversity at the stand level. Stand level species diversity can measure species richness, species evenness, or both. Diversity indices may be more practical, as they combine richness and evenness into a single index, and there is no need to fit a model to abundance data. Regardless of the diversity index chosen, a consistent approach should be used in order to facilitate comparisons and provide meaningful diversity measures.
Chapter 3

Structural Diversity Measures

Structural diversity can be as important to overall species diversity as is the diversity of plant species in the forest plant community (Kimmins 1997). This viewpoint is supported by research into avian and insect diversity. Stand structure can also have a profound impact on stand growth efficiency, and measurements of structure can play important roles in predicting the growth of trees.

Birds have long been observed to frequent more structurally diverse habitats, depending on certain structures or seral stages for “niches” or “ways of life” (Whittaker 1972). Spatial complexity was cited as a key element of the maintenance of bird species diversity in northeastern and north central mixed hardwood-conifer forests (Temple et al. 1979), while the vegetation profile was found to be more important to bird species diversity than the number of plant species present in the southwestern United States (Franzreb 1978). In northern California, Moen and Gutierrez (1997) reported that spotted owls were found to use areas of high structural diversity as roosts and nest sites. Shrubs, snags, and downed wood, which occur in a wide range of forest ages, were found to be an important part of western North American bird habitat, and bird species abundance was found to be highly related to the surrounding habitat pattern (Vernier et al. 1997).

In Australian Eucalypt forests, more species of arthropods and a greater abundance of individuals were found on the most productive plants and where the physical and structural
complexity was high (Recher et al. 1996). Furthermore, data also showed that bird species diversity was correlated with the structural attributes of the vegetation and the number of plant genera in the understory. These data led Recher et al. to conclude that the high structural and floristic complexity of the forests which are rich in this regard were correlated with abundant and diverse arthropod fauna.

Bird diversity, as well as arthropod diversity, has been shown to be significantly correlated with mature (and more structurally diverse) forests and site productivity (Recher et al. 1996). Since both structural and floristic diversity determine the energy rates, nutrient flux, water, and hence, available resources in the forest (Aber 1979), species' abilities to coexist will be affected by structural and/or floristic diversity and will determine species richness and diversity (Recher et al. 1996).

Stand structure can have an impact on stand growth. Gaps in the forest cover can allow for rapid regeneration, while large leaf areas and tree crowns can intercept sunlight and precipitation (Spies 1998). There are numerous studies that show a variety of patterns of growth relating to structural complexity, depending on the species involved (see Oliver and Larson 1996). In general, mixed species stands are considered to have a positive effect on growth, as they tend to slow or prevent the spread of pathogens (Hunter 1990). Multi-cohort stands can have higher growth rates than that of single-cohort stands in situations where lower-strata trees do not die readily (Oliver and Larson 1996). On the other hand, some studies show that yield can be less in these types of stands, but tree size and quality can be higher (Oliver and Larson 1996).
Many stand structure diversity measures have been developed. Most measurements of stand structure are 'ad hoc' methods; that is, they are derived from modifying an existing diversity measurement. However, structural indices can be functions of many different stand variables. Since most use either height or diameter size densities, the indices have been classed as vertical, horizontal, or "other".

3.1 Measures of Vertical Diversity

Pioneering research on the relationship of structural diversity to songbird populations led to the first attempts at a structural diversity index. MacArthur and MacArthur (1961) constructed foliage height profiles by measuring the amount of vegetation at sixteen heights, from six inches to 60 feet above ground. The foliage height diversity index, $FHD$, has the same form as Shannon's Index and is calculated as follows:

$$FHD = -\sum_{i} p_i \ln p_i$$  \[3.11\]

where: $p_i$ = the proportion of total foliage in the $i$th layer.

MacArthur and MacArthur (1961) used this index in various analyses and found that knowledge of plant species diversity did not provide additional knowledge of bird species diversity. Linear regression of bird species diversity (calculated with Shannon’s Index) versus $FHD$ yielded a significant relationship. Moreover, the variability in bird species diversity not accounted for by the variation in foliage height diversity was not accounted for by variations in plant species diversity or latitude (MacArthur and MacArthur 1961).
Subsequent researchers into avian community structure in the United States, Germany, and Britain have continued to use MacArthur and MacArthur's (1961) approach, with slight modifications to the calculation of $p_i$. In a variety of midwestern habitats, ranging from grasslands to forest, Willson (1974) calculated $p_i$ using presence/absence in three defined vertical layers. In an analysis of bird communities in Germany, Erdelen (1984) calculated $p_i$ using foliage area in twelve defined vertical layers. The $FHD$ of British forests were evaluated by Ferris-Kaan (1998) with $p_i$ calculated as the percent foliar cover in four defined vertical height bands.

In other research, MacArthur and MacArthur's (1961) original approach has been modified. In a study of northern hardwood forests, Aber (1979) used that original approach, as well made a second parallel analysis, dividing the forest canopy into four strata and calculating $p_i$ as the relative distribution of foliage by strata. The diversity of vegetation in southern Wisconsin forests was evaluated by Ambuel and Temple (1983) in two ways. First, $FHD^1$ was evaluated similarly to MacArthur and MacArthur's (1961) original approach, but using Simpson's Index as a base:

$$FHD_{AT} = -\sum_i 1/p_i^2$$

where: $p_i$ = the proportion of total foliage in vertical layer $i$.

Second, a measure of forest profile diversity, $FPD$ (Ambuel and Temple 1983) was also introduced:

$$FPD = -\sum p_i \ln p_i$$

---

^1 To avoid confusion between the definition of $FHD$ used in MacArthur and MacArthur (1961) and that of Ambuel and Temple (1983), the latter will be referred to as $FHD_{AT}$. 

19
Ambuel and Temple (1983) designated sixteen profile types, using the presence/absence of vegetation in each possible combination of four height zones.

Willson (1974) and Erdelen (1984) confirmed MacArthur and MacArthur's (1961) results, finding that $FHD$ was highly correlated to bird species diversity. However, these relationships did not hold when only forested plots were included in the analyses. When Aber (1979) used MacArthur and MacArthur's (1961) layer definitions, $FHD$ correlated strongly with stand age. On the other hand, there was not a significant relationship with stand age when $FHD$ was calculated with four strata. Results of Ambuel and Temple's (1983) study revealed a strong relationship between $FHD_{AT}$ and the presence of long-distance migrants.

### 3.2 Measures of Horizontal Diversity

The first measures of horizontal diversity emerged from forest planning models. Distributions of diameters are used as inputs to some natural resources mathematical programming models, and so can be used to juxtapose the evaluation of diversity with economic returns in a forest. Gove et al. (1991) measured diameter diversity using the dichotomous rarity measures presented by Patil and Taillie (1979, 1982), with nine two-inch diameter classes in place of species. In this approach, rarity is measured as:

$$
\Delta_p(\pi) = \frac{1 - \sum_i \pi_i^{\beta+1}}{\beta}, \quad \beta \geq -1
$$

[3.14]

where: $\pi_i$ = the proportion of individuals in diameter class $i$; and
\[ s = \text{the number of diameter classes.} \]

These rarity indices are synonymous with Hill’s indices: when \( \beta = -1 \), \( \Delta_\beta (\pi) \) is the species count; when \( \beta = 0 \), \( \Delta_\beta (\pi) \) is Shannon’s index; when \( \beta = 1 \), \( \Delta_\beta (\pi) \) is Simpson’s index (Gove et al. 1991).

Later researchers used Shannon’s Index to predict the effects of some management regimes on the diversity of tree sizes. Buongiorno et al. (1994) used the diameter distribution as a measurement of stand structure for several reasons. First, because it plays a key role in the structure of forest stands, it is a key element in the biological diversity of stands. Second, since diameter – height relationships are well defined and tree height is positively correlated with \( FHD \), the diameter distribution is a good proxy for \( FHD \). Third, a measure using diameters is more economically feasible than \( FHD \).

Buongiorno et al. (1994) measured diameter diversity in northern hardwoods in Wisconsin using Shannon’s Index with \( p_i = \text{the proportion of trees in the } i \text{th two-inch diameter classes} \). In the French Jura Mountains, Buongiorno et al. (1995) used a similar technique with five-cm diameter classes.

A “normalized absolute deviation” (NAD) index was used by Önal (1997) to measure divergence of a diameter distribution from an arbitrarily specified target distribution:

\[
NAD = 2(1 - \beta) - \sum_{i=1}^{s} |y_i - \beta| \sum_{j=1}^{s} y_j \left/ \sum_{j=1}^{s} y_j \right. \\
= 2(1 - \beta) - \sum_{i=1}^{s} |p_i - \beta| \\
\]

\[3.15\]
where: \( y_i \) is the number of individuals in species \( i \);

\[
p_i = \frac{y_i}{\sum_i y_i};
\]

\( \{ \beta_i \} \) is a specified distribution with \( \beta_i > 0 \) and \( \sum_i \beta_i = 1 \);

\( n \) is the number of classes; and

\[
\beta = \min(\beta_i).
\]

This measure exhibits properties similar to that of Shannon's index when the target distribution is uniform (i.e., \( \beta = 1/n \)). Like Shannon's index, the maximum value of \( \text{NAD} \) is reached when the proportion in each class is equal, \( \text{NAD} \) is scale neutral, and the value of \( \text{NAD} \) is independent of class ordering \(^2\) (Önal 1997). Önal (1997) used simulations to show that there is a strong significant quadratic relationship between \( \text{NAD} \) and Shannon's Index.

### 3.3 Other Structural Diversity Measurements

Various other measures of structural diversity have been developed in the literature. However, none has gained widespread acceptance. The following methods are only a sampling of the various methods developed.

Freemark and Merriam (1986) introduced an index of spatial variability. Their habitat heterogeneity index \((HH)\) is a measure of plant species and forest structure diversity derived from Shannon's index (after Orloci 1970) \(^3\):

\[
HH = -\sum_{i} \sum_{j} X_{ij} \ln(X_{ij} / \bar{X}_i) \tag{3.16}
\]

---

\(^2\) See section 2.4: H1, point 4 in H4, and point 5 in H4, respectively.
where: $c =$ number of plots;

$r =$ number of classes; and

$X_{ij}$ is the proportion of individuals in the $i$th class of the $j$th plot.

Equation [3.16] differs from Shannon’s index in that the denominator contains the average value of the class, not the total value of the class. $HH$ was computed separately for eight components: tree density, tree dbh class, canopy closure, foliar cover in five defined vertical bands, average canopy height, herb height, percent litter, and percent bare ground. Since components varied in units, magnitude, and number of classes, $HH$ was scaled to facilitate comparability and combinability. To arrive at a total value of forest $HH$, the components were summed across forest type. Using this index, the authors concluded that the more heterogeneous habitats in the agricultural landscape near Ottawa had more bird species (Freemark and Merriam 1986).

Some researchers have included a measure of structure in Shannon’s and Simpson’s indices by modifying the way $p_i$ is calculated. Cain and Shelton (1995) computed the $p_i$’s with importance values:

$$IV = \text{relative density} + \text{relative height} + \text{relative frequency}$$

where: “relative” refers to each species’ contribution to the total for all species.

Then, $p_i = \text{proportion of } IV \text{ in species } i$. The resulting values for both Shannon’s and Simpson’s indices were then used to compare the diversity of woody understory growth after disturbance in the southeastern United States.

---

3 Although this measure of diversity is presented as a forest-level measure (whereas the other measures presented are stand level measures), the measure could easily be modified to a stand level one by removing the second summation sign.
A variety of tree and stand variables have been used to arrive at an index of structural diversity. One such recent example (Lähde et al. 1999) uses seven indicator variables by species: stem distribution by diameter size, basal area of growing stock, volume of standing dead trees, volume of fallen dead trees, undergrowth density, occurrence of "special trees", and volume of charred wood. All variables were delineated into classes, e.g., trees were classed into three diameters groups and two basal area groups. Based on the values of the indicator variables, stands were given diversity scores by species, which were then combined into an index for the entire stand. The index was found to differentiate between forest development types and site types fairly well (Lähde et al. 1999).

3.4 Discussion of Measures

Reviews of bird communities and vegetation stand structure subsequent to MacArthur and MacArthur (1961) emphasized that the location of plots greatly affects the discriminant abilities of the foliage height diversity index. However, all but one of the studies cited was not directly comparable to MacArthur and MacArthur's original work. Willson (1974) and Erdelen (1984) assert that the mere existence of a tree layer in the canopy determines bird diversity, regardless of the foliage volume or distribution among layers; however, their results are not truly comparable as classes were not defined in exactly the same manner. Aber (1979) used four strata in the calculation of FHD, finding it uncorrelated with stand age. Again, the calculation of this FHD was quite different from MacArthur and MacArthur's since Aber's strata were of different sizes and at different heights, depending on
the sample site. This result is only comparable to MacArthur and MacArthur's results if stand age is seen as a proxy for bird diversity.

A variety of techniques were used to assess vegetation cover and FHD, and results depended, at least in part, on the selection of class boundaries. Height and diameter are continuous variables, and as such, the division of data into classes will necessarily depend on the distribution of the data. Since many studies are incomparable, caution should be used in drawing conclusions (Erdelen 1984). Further, no study compared the use of vertical and horizontal stand structure indices, and thus, no comparison may be drawn.

Many stand structure measures have been developed to characterize both vertical and horizontal stand structure. However, most measurements of stand structure are derived by modifying existing diversity measurements. Unfortunately, no one method has been used consistently, which makes comparisons and evaluations difficult. Furthermore, results depend on splitting continuous variables into classes. This class division could have a marked effect on results. Some researchers have also used variables such as volume of coarse woody debris or crown ratio, which are not always measured in common forest surveys, and thus, these measures would be difficult to use from an economic standpoint. A more practical measure of stand structure will be one that uses readily available data, does not depend on subjective division of continuous variables into classes, and directly correlates to overall stand biological diversity.
Chapter 4

Possible Measures of Stand Diversity Using Stand Structure

An index of diversity is needed which will characterize the diversity of stand structure. Stand structure diversity has been defined to include the richness of tree species and sizes (Lähde et al. 1999). Further, structural diversity increases as the number of species and number of sizes increases (Buongiorno et al. 1994). This situation may be described using a probability distribution where every value over a specified range is represented equally. That is, every possible value of dbh and height may be found in a stand with equal probability. The probability distribution of maximal diversity under these circumstances is described mathematically by the uniform distribution. Low diversity stands would be defined as those with few species and few tree sizes, and minimum diversity would occur where all trees in a stand are of single tree species and single size. This distribution would be a spike at a single point.

In an effort to incorporate stand structure into a diversity measure, several possible indices were derived. For all of these indices, the standard for maximum diversity was a uniform distribution. However, this assumption may be challenged since there is no quantitative definition of high structural diversity. Although a highly diverse stand is well characterized with this distribution, proposed measures could be modified with relative ease for use with another distribution.
There are several characteristics which a stand diversity index should possess. The alternative measures will be evaluated with a goal of meeting all or most of these criteria.

4.1 Desirable Characteristics of Measures of Stand Diversity Using Stand Structure

An index of forest stand structure should have the following characteristics:

1) The index should measure both richness and evenness of structural and species diversity. A diversity index of this type would take into account the number of individuals of different species and size, as well as the distribution of individuals over those species and sizes.

2) The index should use commonly measured tree attributes only. This would enable forest managers to use the index with inventory data or other common forest sample data without additional expense.

3) The index should not rely on combining data into diameter or height classes. Many of the indices presented in Chapter 3 require that data be segregated into arbitrary classes. This practice is undesirable, as a change in class definitions could cause a change in index.

4) The index should not depend on sample size or plot size. Sample plots from forest inventories may consist of one plot or a cluster of several plots. In order to make comparisons between populations, the index should be insensitive to cluster size and plot size.

5) Horizontal and vertical diversity should contribute equally to the index. It has been shown that both tree diameter and height contribute to diversity by creating a variety of habitats. Thus, given two populations with equal horizontal diversity, the population with greater vertical diversity would have a higher index.
6) Basal area should be used to measure proportions of trees in species and size distributions. Although some authors (for example, Lewis et al. 1988, McMinn 1992) have used stems per ha, LeMay et al. (1997) recommended basal area, as it better represented the proportion of resource use.

7) The index should take on values between zero and $S$. An index value of $S$ should be associated with stands of high diversity, and an index value of zero should be associated with stands of low diversity.

8) The index should measure the deviation between a particular community and that of a maximally diverse community.

   a) Given a set range, the index should be tree size insensitive, or location invariant. For example, a population with small trees evenly distributed over a ten cm diameter range would have the same index value as a sample with large trees evenly distributed over a ten cm diameter range (assuming all other things being equal). This would ensure that larger individuals would not drive the index.

   b) The index should be related to the range of values in diameter or height for continuously occupied ranges. In other words, a population with a larger diameter and/or height range that occurs evenly over that range would yield a larger diversity index than a population with values occurring over a smaller range, or over only parts of the range (as in a bimodal distribution). This would ensure that stands that exhibit greater size diversity would have higher index values.

   c) The index associated with a bimodally distributed population with two separate and distinct canopy layers with ranges of $Range_1$ and $Range_2$, should be at least as great as that of a unimodal distribution with a continuous range $= Range_1 + Range_2$. For
instance, an old growth stand with regenerating trees in the understory would be considered at least as diverse as a stand with a wide single-canopy layer.

If these characteristics are met, the index should be easy to use and interpret, and the maximum value of the index will result for uniform stands with the greatest range. Some of the criteria may not be met, and evaluations with data analysis should reveal whether these criteria are truly necessary.

4.2 Possible Alternative Measures of Stand Diversity Using Stand Structure

Several alternative measures were derived based on existing measures, such as Shannon’s Index, or based on techniques familiar to foresters and biologists, such as curve fitting and variance measurements.

4.2.1 Shannon’s Index Extended to Diameter, Height, and Species

Shannon’s index (Shannon and Weaver 1949) has been computed in the forestry and ecology literature with a species distribution, a diameter distribution, or a height distribution. However, Shannon’s index computed with diameter, height, and species was not found in the literature. The index could be used separately for each of these variables, producing $H'_d$, $H'_h$, and $H'_s$, respectively, and then these indices could be combined to produce a stand index, $H'_{d+h+s}$. Alternatively, data could be classed simultaneously into diameter, height and species classes, producing one index, $H'_{dhu}$. The former method will be referred to as the ‘Post-hoc Method’, and the latter will be referred to as the ‘Combination Method’.
4.2.1.1 Post-hoc Method

The procedure for computing the post-hoc Shannon's index for a sample plot is:

1) Tree data are combined into diameter classes using dbh. The proportion of trees in the \( i \)th diameter class, \( p_i \), is the percentage of basal area in that diameter class. Then, Shannon's Index is computed with \( p_i \) and equation [2.1] to yield \( H'_d \).

2) Tree data are combined into height classes. The proportion of trees in the \( j \)th height class, \( p_j \), is calculated as the percentage basal area in that height class. Then, Shannon's Index is computed with \( p_j \) and equation [2.1] to yield \( H'_h \).

3) Tree data are combined by species. The proportion of trees in the \( k \)th species, \( p_k \), is calculated as the percentage basal area in that species. Then, Shannon's Index is computed with \( p_k \) and equation [2.1] to yield \( H'_s \).

4) The final index is computed as the average of the diameter, height, and species indices, as it is a convenient way to keep the index on a similar scale of values as the original Shannon's Index:

\[
H'_{d+h+s} = \text{average} (H'_d, H'_h, H'_s) = \frac{H'_d + H'_h + H'_s}{3}
\]  

[4.17]

The minimum value of the Extended Shannon Index under the post-hoc method is zero. This minimum value results when only one species, one diameter class, and one height class are present in a sample. The maximum value of the Extended Shannon Index under the post-hoc method is reached when \( H'_d \), \( H'_h \), and \( H'_s \) are at their maximum values. The maximum value of the \( H'_s \) is \( \ln(S) \) (Magurran 1988). This maximum is achieved when each species is represented equally within a plot (uniformly distributed). Similarly, if all diameter and
height classes are represented equally within a plot, the maximum values of $H'_d$ and $H'_h$ are
$\ln(n_1)$ and $\ln(n_2)$, respectively, where $n_1$ is the number of possible dbh classes and $n_2$ is the
number of possible height classes. The overall maximum of the Extended Shannon Index
under these circumstances is the average of these maximum values:

$$\max(H'_{d+h+s}) = \frac{\ln(S) + \ln(n_1) + \ln(n_2)}{3}$$  \hspace{1cm} [4.18]

4.2.1.2 Combination Method

The procedure for computing the combination index for a sample plot representing one
forest stand is to first combine the data simultaneously by species, diameter class, and height
class. The proportion of basal area in the $i$th diameter, $j$th height, and $k$th species class, $p_{ijk}$, is
then calculated. Then, Shannon's Index is computed with $p_{ijk}$ and summed over all classes to
yield $H'_{dhs}$.

As with the post-hoc method, the minimum value of the Extended Shannon Index under
the combination method is zero. This value results when there is only one species in a plot,
and this species occurs in only one diameter-height class combination. The maximum value
of the Extended Shannon Index under the combination method will be reached when all
diameter and height classes, as well as species are represented equally within a plot
(uniformly distributed). This maximum value is:

$$\max(H'_{dhs}) = \ln(S \times n_1 \times n_2) = 3 \times \max(H'_{d+h+s})$$  \hspace{1cm} [4.19]
4.2.1.3 Properties of the Extended Shannon's Index

The Extended Shannon Index, in both the post-hoc and combination versions, measures richness and evenness of stand structure, giving equal weight to horizontal and vertical diversity. The indices use basal area and require only commonly measured tree attributes in their computation. Shannon's Index also has been shown to be relatively insensitive to sample size (Magurran 1988), and attains its highest values when distributions are perfectly even (i.e., uniform). The Extended Shannon indices should be insensitive to tree size, as indices are weighted only on the proportion of basal area occurring in a particular class.

Both forms of the Extended Shannon Index possess the desirable characteristics listed as items 1, 2, 4, 5 and 6 (Section 4.1). Item 8 cannot be considered without further using data. Item 7 is not satisfied because the maximum values of these indices are functions of the number of possible diameter and height classes. This dependence results in indices that have a lower bound of zero, but are not bounded by $S$. Item 3 is not met because the indices rely on pre-defined classes. This is undesirable, as this creates issues such as where to put class boundaries. A change in class boundaries or an increase in the number of classes would invariably change the value of the indices. Even if set classes were used, there would be pronounced differences simply due to geography. For instance, in Coastal British Columbia, trees are larger on average and have a wider range of sizes than in the interior of the province. More classes would be represented in Coastal areas, resulting in larger indices. On the other hand, more species may be represented in sample plots from Interior areas of the province, which would also result in larger index values. For this reason, care should be taken in comparing diversity indices (of any type) from different geographic areas.
To alleviate problems associated with classes, an integral form of the Extended Shannon Index could be presented. However, this option would rely on finding functional forms for species, diameter, and height distributions. While a functional form may be found, it may not have a closed form, or may not be flexible enough to fit the variety of distributions that may be found. This issue will be addressed in the next section.

4.2.2 Structural Index based on Curve Fitting

When a continuous variable, such as height or dbh, is translated into a class variable, information about the distribution of that variable is invariably lost. In order to avoid using diameter and height classes, which may affect the index, an index based on a continuous distribution is desirable. Foresters and biologists have used continuous distributions to describe the frequency distributions of diameter and height both univariately and bivariately. These distributions have included the Beta, Johnson’s SB, and Weibull, with varying degrees of success (Hafley and Schreuder 1977).

In fitting such a distribution, the resulting functional form may be used in an integral form of Shannon’s Index, or the \( Q \) statistic could be used to measure diversity based on the derived curve. However, if data are sparse, a derived curve may cover a range much greater than that of the data. The use of the Extended Shannon Index or the \( Q \) statistic in this type of situation would involve making an inference about the population using the fit of a distribution with a restricted range. This may be inappropriate, and could produce misleading results. Hence, this option was not further investigated.
Alternatively, in fitting a regression curve, the regression coefficients may indicate the centre, shape, or scale of a distribution. For instance, the two coefficients in the function describing a univariate normal distribution indicate the mean and variance of the distribution. A known distribution could be fit to the data and its regression coefficients could be analyzed as indicators of diversity.

Of the candidate distributions, the Beta distribution offers the most flexibility in terms of skewness, kurtosis and shape of the curve (Hafley and Schreuder 1977). However, the Beta is a complex function, which involves the computation of Gamma function values and may be extremely difficult to fit.

Johnson's $S_B$ and its bivariate form, Johnson's $S_{BB}$ have been shown to perform consistently better than the Weibull or Beta distributions in fitting diameter and height data (Schreuder and Hafley 1977). However, this probability density function does not have a closed-form cumulative density function (Schreuder and Hafley 1977), and thus, cannot be used without allocating data to classes or making an approximation with numerical integration.

The Weibull distribution has been used extensively in forestry applications (Clutter et al. 1983). It has advantages in that it is fairly flexible in shape and much simpler than some other functions. Although the univariate and bivariate Weibull distributions can closely approximate a normal curve under certain conditions (Bailey and Dell 1973), approximations
to univariate and bivariate uniform distributions are only adequate over a portion of the range of the data.

A desirable distribution for investigation would be one that is flexible in shape, but relatively easy to fit. Considering the advantages and disadvantages, the Weibull distribution (in both univariate and bivariate forms) was selected as the base curve for fitting.

4.2.2.1 Univariate Curve Fitting

The cumulative univariate Weibull probability functions for diameter and height based on basal area are as follows:

\[ F_1(dbh) = 1 - \exp\left(-\left(dbh/b_1\right)^{c_1}\right) \] \[ F_2(\text{height}) = 1 - \exp\left(-\left(\text{height}/b_2\right)^{c_2}\right) \]

where: \(b_1, b_2, c_1,\) and \(c_2\) are coefficients to be estimated; and \(dbh\) and \(\text{height}\) are the independent regression variables.

Equations [4.20] and [4.21] would be fit for \(dbh\) and \(\text{height}\) separately by species using non-linear fitting methods. A good fit would be indicated by a small root mean square error (RMSE), computed as:

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - m}} \]

where: \(y_i\) is the dependent variable, the cumulative distribution of the sample \((F_1(dbh))\), or \(F_2(\text{height}))\);
\[ \hat{y}_i = \text{the predicted value of the cumulative distribution under the assumed theoretical distribution;} \]
\[ n = \text{the number of observations; and} \]
\[ m = \text{the number of coefficients in the regression model.} \]

The estimated coefficients could be analyzed for patterns relating to diversity and a coefficient chosen as a representative diversity index. By species, the diversity index would be the average of the dbh and height diversity indices. The total diversity index for a plot would be computed as the sum of all species diversity indices. Thus, if two plots have the same structure, the plot with more species would have a higher diversity index.

### 4.2.2.2 Bivariate Curve Fitting

The bivariate Weibull has been derived in many forms (see Lu and Bhattacharyya 1990); however, based on pilot investigations, the best performing form was given by Hafley and Schreuder (1976):

\[
G(\text{dbh}, \text{height}) = F_1(\text{dbh}) \times F_2(\text{height}) \times [1 + a_3 \times (1 - F_1(\text{dbh})) \times (1 - F_2(\text{height}))]
\]

[4.23]

where: \( F_1(\text{dbh}) \) and \( F_2(\text{height}) \) are defined as above; and

\[ a_3, \text{along with } b_1, b_2, c_1, \text{and } c_2 \text{ from [4.20] and [4.21], are coefficients to be estimated.} \]

\( F(\text{dbh, height}) \) is computed as the cumulative basal area at each dbh - height combination represented in the plot. Equation [4.23] would be fit for dbh and height simultaneously, by
species, using non-linear fitting methods. Starting values for $b_1$, $b_2$, $c_1$, and $c_2$ could be obtained from the univariate regressions. The resulting estimated coefficients would then be analyzed for patterns relating to diversity. If possible, a coefficient would be chosen as a representative diversity index. As in the univariate case, the total diversity index for a plot would then be computed as the sum of all species diversities. Thus, all things being equal, if two plots have the same structure, one with more species would have a higher diversity index.

4.2.2.3 Properties of a Structural Index Based on Curve Fitting

Both the univariate and bivariate methods possess the desirable characteristics listed as items 1, 2, 3, 5, and 6 (Section 4.1). An index derived from the Weibull distribution would measure richness and evenness using basal area and commonly measured tree attributes. The index would not rely on combining data into classes, and would account for vertical and horizontal diversity equally. However, items 4, 7, and 8 cannot be considered without further evaluation using data.

Both the univariate and bivariate methods presented above rely on the fit of the data to a known distribution. To assure that no inappropriate inferences are made, this method should be used only if samples yield similar fit statistics, such as small residual error and bias. A more significant problem may arise in that no coefficient may be adequate as a measure of diversity.
4.2.3 Structural Index Based on \( R^2 \) (STRI)

The coefficient of determination, or \( R^2 \), is the measure of how much of the variation in a regression line (or curve) is explained by the independent variable(s). Thus, a function representative of maximum diversity, such as the uniform, could be put into the form of a cumulative distribution function (cdf), and compared to a cumulative empirical sample distribution. If the cumulative empirical sample is regressed on the function of the theoretical cdf, the resulting \( R^2 \) will give a measure of the deviation of the sample from maximum diversity. A high \( R^2 \) (close to one) would indicate that the empirical sample came from a distribution that was similar to that of a high diversity population; a low \( R^2 \) (close to zero) would indicate that the sample came from a population with a distribution that deviated substantially from that of a high diversity population. \( R^2 \) is traditionally computed as follows:

\[
R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total corrected sum of squares}} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]  

where: \( y_i \) is the dependent variable, the cumulative distribution of the sample (\( F(dbh) \), \( F(height) \), or \( F(dbh, height) \));

\( \hat{y}_i \) = the predicted value of the cumulative basal area distribution under the assumed theoretical distribution;

\( \bar{y} \) = the mean value of the cumulative basal area distribution of the sample; and

\( n \) = the number of observations.

If the theoretical distribution is taken to be a uniform distribution, then the regression form would necessarily go through the origin. This changes the formula for \( R^2 \) to:
\[ R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total uncorrected sum of squares}} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i)^2} \]  

Although, [4.25] is algebraically correct, it no longer has the same interpretation as [4.24] (Kozak and Kozak 1995). However, this $R^2$ could be used as a measure of diversity, as it describes the amount of dispersion between some theoretical distribution of maximal diversity and a sample. $R^2$ gives a numerical value between zero and one in most cases, though it may fall below zero in cases where the shape of the empirical distribution is much steeper than that of the theoretical. Since this is a variation on the $R^2$ value, this measure will be known as STRI.

4.2.3.1 STRI Univariate Method

For the univariate case, dbh and height are regressed separately by species. The cumulative empirical basal area distributions of dbh and height for species $k$ are labelled as: $F_1(dbh)_k$ and $F_2(height)_k$. If it is assumed that a uniform distribution represents the highest possible diversity in a community, then the theoretical predicted values of the cumulative distribution functions of $dbh$ and $height$ for any species are represented by:

\[ \hat{F}_1(dbh) = \frac{dbh - a_4}{b_4 - a_4}, \quad a_4 \leq dbh \leq b_4 \]  

\[ \hat{F}_2(height) = \frac{height - a_5}{b_5 - a_5}, \quad a_5 \leq height \leq b_5 \]

where: $a_4$ and $a_5$ are the minimum values of dbh and height, respectively; and $b_4$ and $b_5$ are the maximum values of dbh and height, respectively.

Then, $\text{STRI}_{dbh}$ and $\text{STRI}_{height}$ for species $k$ are:
\[
\text{STRI}_{\text{dbh}_k} = 1 - \frac{\sum_{i=1}^{n_1} \left( F_1(\text{dbh}_i)_k - \hat{F}_1(\text{dbh}_i)_k \right)^2}{\sum_{i=1}^{n_1} \left( F_1(\text{dbh}_i)_k \right)^2} \quad [4.28]
\]

\[
\text{STRI}_{\text{height}_k} = 1 - \frac{\sum_{j=1}^{n_2} \left( F_2(\text{height}_j)_k - \hat{F}_2(\text{height}_j)_k \right)^2}{\sum_{j=1}^{n_2} \left( F_2(\text{height}_j)_k \right)^2} \quad [4.29]
\]

where: \( \text{dbh}_i \) corresponds to the \( i \)th value of dbh;

\( \text{height}_j \) corresponds to the \( j \)th value of height;

\( n_1 \) is the number of observations with measured dbh; and

\( n_2 \) is the number of observations with measured height.

To arrive at a measure of structural diversity for species \( k \), \( \text{STRI}_{\text{dbh}_k} \) and \( \text{STRI}_{\text{height}_k} \) are averaged to produce \( \text{STRI}_{(d+h)_k} \):

\[
\text{STRI}_{(d+h)_k} = \frac{\text{STRI}_{\text{dbh}_k} + \text{STRI}_{\text{height}_k}}{2} \quad [4.30]
\]

An overall measure of diversity for a sample plot will be \( \text{STRI}_{d+h} \), the sum of the values of all \( \text{STRI}_{(d+h)_k} \) over a plot:

\[
\text{STRI}_{d+h} = \sum_{k=1}^{S} \text{STRI}_{(d+h)_k} \quad [4.31]
\]

The maximum value of the index is \( S \), the number of species. This maximum would occur if both of the cumulative distributions for dbh and height match that of the uniform for each species. The minimum value of this index is negative, and would occur if dbh or height were a single value at either the minimum or maximum value of the possible range.
4.2.3.2 STRI Bivariate Method

For the bivariate case, dbh and height are considered simultaneously by species. The cumulative empirical sample distribution of dbh and height for species \( k \) is labelled as: 

\[ G(dbh, height)_k \]. If the assumption is made that a bivariate uniform distribution represents the highest possible diversity in a community, then the theoretical predicted cumulative basal area distribution function of \( dbh \) and \( height \) for species \( k \) is:

\[
\hat{G}(dbh, height)_k = \left[ \frac{dbh - a_6}{b_6 - a_6} \right] \times \left[ \frac{height - a_7}{b_7 - a_7} \right], \quad \begin{cases} 
 a_6 \leq dbh \leq b_6 \\
 a_7 \leq height \leq b_7 
\end{cases} \tag{4.32}
\]

where: \( a_6 \) and \( a_7 \) are the minimum possible values of dbh and height, respectively;

and

\( b_6 \) and \( b_7 \) are the maximum possible values of \( dbh \) and \( height \), respectively.

Then, \( STRI_{dbh} \) for species \( k \) is:

\[
STRI_{dbh} = 1 - \frac{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (G(dbh_i, height_j)_k - \hat{G}(dbh_i, height_j)_k)^2}{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (G(dbh_i, height_j)_k)^2} \tag{4.33}
\]

where: \((dbh_i, height_j)\) corresponds to the \( i \)th value of \( dbh \) and \( j \)th value of \( height \).

An overall measure of diversity for a sample plot is \( STRI_{dbh} \), the sum of the values of all \( STRI_{dbh} \) over a plot:

\[
STRI_{dbh} = \sum_{k=1}^{S} STRI_{dbh_k} \tag{4.34}
\]

The maximum value of \( STRI_{dbh} \) is \( S \). The maximum would occur if the distribution of the sample was a bivariate uniform for each species in the sample. The minimum value of this index is negative, and would occur in cases where the surface of the empirical joint distribution is much steeper than that of the bivariate uniform distribution.
4.2.3.3 Properties of STRI

Both the univariate and bivariate STRI possess the characteristics 1, 2, 3, 5 and 6 listed in Section 4.1. An index derived from the \( R^2 \) would measure richness and evenness using basal area and commonly measured tree attributes. The index would not rely on combining data into classes, and would account for vertical and horizontal diversity equally.

Item 4 will not be met, as \( R^2 \) is dependent on sample size. If \( n \) is equal to the number of coefficients in the regression equation, then \( R^2=1 \); as \( n \) gets large (greater than 30), \( R^2 \) decreases, regardless of the fit of the data (Neter et al. 1996).

Items 7 and 8 will only partially be met. While the index would be correlated to the range of values in dbh and/or height and would adequately describe bimodal distributions, the index will be sensitive to tree size. For example, consider a forest type where the maximum range of dbh values is 0 to 100 cm, and where there are three stands which have three univariate uniform dbh distributions, covering three equal portions of this 100 cm dbh range: 0.0 to 33.3 cm, 33.3 to 66.7 cm, and 66.7 to 100 cm. If the three distributions have the same proportions of basal area equally spaced over their respective ranges, the denominators of each of the three STRI's derived from the distributions will be equal. The numerator of the univariate STRI is dependent on the area between the empirical and theoretical cdf’s; it is the squared distance between the two curves. Graphically, this is represented by the shaded areas in Figure 4.1.
The grey region and the blue region (corresponding to the small stand and the large stand) have equal areas. However, it can be proven mathematically that the yellow region (corresponding to the mid-range stand) has a smaller area than that of the large and small stands. Thus, the STRI associated with a uniform distribution centred in the middle of the theoretical range will be higher than that of uniform distributions centred either at the low end or high end of the range.

4.2.4 Structural Index based on Variance Comparisons (STVI)

The distribution of diameters and/or heights in a structurally diverse forest may be described in terms of variances. The uniform distribution yields a very high variance, though not the highest possible variance for any distribution over a given range. The diversity of a stand may be evaluated on the measured stand variance relative to that of a uniform distribution.
4.2.4.1 STVI Univariate Method

The variance of an empirical dbh or height distribution can be given by:

\[ S^2 = \frac{\sum_{i=1}^{n} w_i \times (x_i - \bar{x})^2}{\sum_{i=1}^{n} w_i} \]  

where:
- \( x_i = \text{dbh}_i \) or \( \text{height}_i \);
- \( \bar{x} \) is the mean of \( \text{dbh} \) or \( \text{height} \);
- \( w_i \) is the weight of the \( i \)th cell, defined as proportion of basal area of the plot represented by the \( i \)th dbh or height point; and
- \( n \) is the number of dbh or height points.

This estimate of the variance differs from that of the usual estimator used in forestry, which is weighted by trees per hectare. This weighting is consistent with the other proposed measures. The variances of univariate uniform distributions of dbh and height weighted by basal area are given by:

\[ S^2_{\text{dbh}_{ui}} = \frac{(b_8 - a_8)^2}{12} \]  
\[ S^2_{\text{height}_{ui}} = \frac{(b_9 - a_9)^2}{12} \]

where:
- \( a_8 \) and \( a_9 \) are the minimum values of dbh and height, respectively; and
- \( b_8 \) and \( b_9 \) are the maximum values of dbh and height, respectively.

The maximum possible variance of a distribution occurs when the distribution is maximally bimodal. For the univariate case with dbh alone, this occurs when half of the basal area is at \( a_8 \) and half the basal area is at \( b_8 \). In this situation, the variance estimator weighted by basal area is given by:
Similar results can be obtained for height by substituting coefficients $a_9$ and $b_9$ for $a_s$ and $b_s$, respectively.

Assuming that a uniform distribution represents a maximally diverse population, the difference between the variance of the dbh distribution from an extremely diverse population ($S_{dbh_{max}}^2$) and that of the uniform distribution ($S_{dbh_u}^2$) will be close to zero. However, the difference between the variance of a low diversity dbh distribution ($S_{dbh_{low}}^2$) and that of a uniform is not as predictable. If the distribution consists of trees that are all of the same dbh, $S_{dbh_{low}}^2$ will be close to zero and $S_{dbh_{low}}^2 - S_{dbh_u}^2$ will be large and negative (at most, $-S_{dbh_u}^2$). If the distribution is maximally bimodal, then $S_{dbh_{low}}^2$ will be very large and $S_{dbh_{low}}^2 - S_{dbh_u}^2$ will be large and positive (at most, $S_{dbh_{max}}^2 - S_{dbh_u}^2$). In order to use these results to yield an index that measures the difference between the variance of a sample plot and that of the maximally diverse case, both instances must be considered. The following formula can be used to define a diversity index, $STVI_{dbh}$ for a species $k$:

$$STVI_{dbh_k} = \begin{cases} 
1 - \left( \frac{S_{dbh_k}^2 - S_{dbh_u}^2}{S_{dbh_{low}}^2} \right)^{p_1} & \text{when } S_{dbh_k}^2 \leq S_{dbh_{low}}^2 \\
1 - \left( \frac{S_{dbh_k}^2 - S_{dbh_u}^2}{m \times S_{dbh_{max}}^2 - S_{dbh_{low}}^2} \right)^{p_2} & \text{when } S_{dbh_k}^2 > S_{dbh_{low}}^2 
\end{cases}$$ [4.39]

where: $S_{dbh_k}^2$ is the variance of $dbh$ for species $k$;
$p_1$ and $p_2$ are constants $> 0$; and

$m$ is a constant $\geq 1.0$.

The constants $p_1$ and $p_2$, define the shape of the curve relating the value of the index to the sample variance: when $p_1$ (or $p_2$) < 1, the curve is concave upward; when $p_1$ (or $p_2$) = 1, the curve is segmented linear; when $p_1$ (or $p_2$) > 1, the curve is concave downward (Figure 4.2). If $p_1 = p_2 > 1$, then a smooth, continuous function results.

Figure 4.2. Comparison of STVI values with $m=1$ and varying values of $p_1$ and $p_2$.

The coefficient $m$ controls the value of the index when the distribution is maximally bimodal (i.e., equation [4.38], for dbh). If $m = 1$, then the index will be zero for a maximally bimodal distribution; as $m$ gets larger, the index value increases for the maximally bimodal case (Figure 4.3).
The values for $p_1$, $p_2$, and $m$ may be chosen by placing three constraints on the index to yield certain index values under defined conditions. First, when a stand is uniform over half the maximum possible range, the value of the index was constrained to equal 0.5. Second, when a stand is distributed bimodally, such that half of its values are uniformly distributed over the lower quartile of the maximum possible range and half of its values are uniformly distributed over the upper quartile of the maximum possible range, the index was constrained to equal 0.5. Third, when a stand is distributed such that its variance is at the maximum value, the index was constrained to equal 0.1. Algebraically, these constraints are:

$$0.5 = \text{STVI}_{\text{dbh}_i} = 1 - \left( \frac{S_{\text{dbh}_i}^2 - S_{\text{dbh}_{5:15}}^2}{S_{\text{dbh}_i}^2} \right)^{p_1} \quad [4.40]$$

$$0.5 = \text{STVI}_{\text{dbh}_i} = 1 - \left( \frac{S_{\text{dbh}_{\text{max}}}^2 - S_{\text{dbh}_i}^2}{m \times S_{\text{dbh}_{\text{max}}}^2 - S_{\text{dbh}_i}^2} \right)^{p_2} \quad [4.41]$$
and \[ 0.1 = STVI_{dbh_k} = 1 - \left( \frac{S_{dbh_{max}}^2 - S_{dbh_{i}}^2}{m \times S_{dbh_{max}}^2 - S_{dbh_{i}}^2} \right)^{p_2} \]  

where: \( S_{dbh_{0.50}}^2 \) is the variance associated with the half-range uniform distribution described in the first constraint; and 

\( S_{dbh}^2 \) is the variance associated with the bimodal distribution described in the second constraint.

After algebraic manipulations, [4.40], [4.41], and [4.42] simplify to:

\[
p_1 = \frac{\ln(1/2)}{\ln(3/4)} \approx 2.4094 \]  

\[
p_2 = \frac{\ln(5/9)}{\ln(3/8)} \approx 0.5993 \]  

and 

\[
m = \frac{1}{3} \left[ 1 + \exp \left( \frac{\ln(0.9) \times \ln(0.75) - (\ln(2))^2}{\ln(9/5)} \right) \right] \approx 1.1281 \]

With these choices of \( p_1, p_2, \) and \( m, \) plots with variances close to that of the uniform would have index values close to one. Plots that are of low diversity and unimodal (e.g., the variance is close to zero) would have index values close to zero. Plots that are of low diversity and bimodal (e.g., variance is close to the maximally bimodal case) would have index values close to 0.1.
Similarly, a diversity index, \( STVI_{\text{height}} \), based on height could be developed. To arrive at a measure of structural diversity for species \( k \), \( STVI_{\text{dbh}_k} \) and \( STVI_{\text{height}_k} \) are then averaged to produce \( STVI_{(d+h)_k} \):

\[
STVI_{(d+h)_k} = \frac{STVI_{\text{dbh}_k} + STVI_{\text{height}_k}}{2}
\]  \[4.46\]

An overall measure of diversity for a sample plot is labelled as \( STVI_{d+h} \), the sum of the values of all \( STVI_{(d+h)_k} \) over a plot:

\[
STVI_{d+h} = \sum_{k=1}^{S} STVI_{(d+h)_k}
\]  \[4.47\]

The maximum value of \( STVI_{d+h} \) is \( S \). The maximum occurs when the univariate distributions of dbh and height are uniform for each species in the plot. The minimum value is zero, and results in cases where each of dbh and height only exist in a single value for each species represented in the plot.

### 4.2.4.2 STVI Bivariate Method

The variance of an empirical bivariate distribution is usually described with the generalized variance (Johnson and Wichern 1998), which is equal to the determinant of the variance-covariance matrix:

\[
S_{\text{dbh, height}}^2 = \det \begin{bmatrix}
S_{\text{dbh}}^2 & \text{cov}(\text{dbh, height}) \\
\text{cov}(\text{dbh, height}) & S_{\text{height}}^2
\end{bmatrix}
\]  \[4.48\]

where: \( S_{\text{dbh}}^2 \) and \( S_{\text{height}}^2 \), are given by equation [4.36]; and

\[
\text{cov}(\text{dbh, height}) = \frac{\sum_{i=1}^{n} w_i \times (\text{dbh}_i - \overline{\text{dbh}}) \times (\text{height}_i - \overline{\text{height}})}{\sum_{i=1}^{n} w_i}
\]  \[4.49\]
where: \( w_i \) is the weight of the \( i \)th cell, defined as proportion of basal area of the plot represented by the \( i \)th \((dbh, height)\) pair; and

\( n \) is the number of \((dbh, height)\) pairs.

The variance of the bivariate uniform distribution of \( dbh \) and height is given by the determinant of its variance-covariance matrix:

\[
S_U^2 = \det \begin{bmatrix}
S^2_{dbh} & \text{cov}(dbh, height)_U \\
\text{cov}(dbh, height)_U & S^2_{height}
\end{bmatrix}
\]  

[4.50]

where: \( S^2_{dbh} \) and \( S^2_{height} \) are given in [4.36] and [4.37]; and \( \text{cov}(dbh, height)_U = 0 \).

Thus, [4.50] simplifies to:

\[
S_U^2 = \det \begin{bmatrix}
\frac{(b_8 - a_8)^2}{12} & 0 \\
0 & \frac{(b_9 - a_9)^2}{12}
\end{bmatrix} = \frac{(b_8 - a_8)^2 \times (b_9 - a_9)^2}{144}
\]  

[4.51]

As in the univariate case, the maximum possible variance occurs when the population is maximally bimodal. This occurs when one quarter of the basal area is at \((a_8, b_8)\), one quarter is at \((a_8, b_9)\), one quarter is at \((a_9, b_8)\), and one quarter is at \((a_9, b_9)\). In this situation, the variance of \( dbh \) is as in [4.37] and the variance of height can be computed similarly. The covariance of \( dbh \) and height is identical to that of the uniform:

\[
\text{cov}(dbh, height)_{\max} = 0
\]  

[4.52]

The maximum variance of a joint distribution of \( dbh \) and height is given by the determinant of its variance-covariance matrix:
The formula given as [4.39] for STVIdbh can now be modified to use the generalized variance to arrive at an index of diversity for a species $k$ using a bivariate distribution:

$$STVI_{dbh} = \begin{cases} 
1 - \left( \frac{S_U^2 - S_k^2}{S_U^2} \right)^{p_1}, & \text{when } S_k^2 \leq S_U^2 \\
1 - \left( \frac{S_k^2 - S_U^2}{m \times S_{max}^2 - S_U^2} \right)^{p_2}, & \text{when } S_k^2 > S_U^2 
\end{cases}$$

[4.54]

where: $S_k^2$ is the generalized variance for species $k$;

$p_1$ and $p_2$ are constants $> 0$; and

$m$ is a constant $\geq 1.0$.

As in the univariate case, the powers, $p_1$ and $p_2$, define the shape of the curve, and the coefficient $m$ controls the value of the index when the distribution has the maximum variance. Imposing the same constraints on the index as in the univariate case, the same values of $p_1$, $p_2$, and $m$ would be used in the bivariate case. Using these values of $p_1$, $p_2$, and $m$, sample plots with variances close to that of the uniform would have index values close to one. Sample plots that are of low diversity would have index values close to zero if they were unimodal or 0.1 if they are bimodal. An overall measure of diversity for a sample plot is labelled as STVIdbh, the sum of the values of all STVI_{dbh} over a plot:

$$STVI_{dbh} = \sum_{k=1}^{S} STVI_{dbh}$$

[4.55]
The maximum value of STVI\textsubscript{dbh} is \( S \). The maximum occurs when the joint distribution of \( \text{dbh} \) and height is uniform for each species in the sample. The minimum value is zero, and results when there are data at only one \( \text{dbh} \) – \text{height} combination for each species represented in the sample.

4.2.4.3 Properties of STVI

The STVI in both the univariate and bivariate forms possesses the characteristics presented as items 1, 2, 3, 5, 6, and 7 (Section 4.1). Both forms of the STVI would measure richness and evenness using basal area and commonly measured tree attributes. The indices would not rely on combining data into classes, and would account for vertical and horizontal diversity equally. In order to satisfy criteria 4, the indices should not depend on sample size; however, this criteria would need to be evaluated using data. Since the indices are based on the variances of the \( \text{dbh} \) and height distributions, weighted by basal area, they should be insensitive to nominal tree size with respect to the distribution of basal area. However, since basal area is used as a weight, larger trees are given more weight. The indices should also be correlated to range, as the variance of distributions with wider ranges tends to be larger. Furthermore, providing that \( p_1, p_2, \) and \( m \) are well chosen, the index value for a bimodal distribution with two distinct ranges should be the same as that of a unimodal distribution with a continuous range equal to the sum of the bimodal ranges. Thus, the desirable characteristics listed as items 8a, 8b, and 8c (Section 4.1) should be met; however, further evaluation is necessary.
The STVI may be the most flexible of the proposed indices presented, as the shape of the curve and its response to bimodal distributions may be modified to fit the users' expectations. Furthermore, this index does not require fitting a distribution to the data, making its application more simple. However, a disadvantage of these indices are their lack of theoretical background. Existing indices, such as Shannon's Index, have been well studied and have well-defined sampling distributions. The sampling distributions of these proposed indices will not be known unless further research is conducted.
Chapter 5  
Evaluation of Alternative Measures of Structural Diversity

Evaluations of the alternative measures with respect to item 8 (Section 4.1) were performed using simulated stands and using data collected during the 1995 Malcolm Knapp Research Forest (MKRF) inventory. Methods for generating the simulated data and collecting the MKRF data are presented, followed by results in using the proposed indices.

5.1 Simulated Tree Data

Proposed measures were evaluated using simulated stands of varying structural diversity. Simulated stands are an invaluable tool in testing indices in that the shape of the distribution of dbh and height can be controlled, allowing for the evaluation of well-defined cases, which may be difficult to find in natural data sets.

5.1.1 Methods

A baseline of maximally diverse structure was first chosen, then simulated stands were compared to this baseline.

5.1.1.1 Baseline Maximal Diversity

High diversity stands were defined as those stands that have a large number of tree sizes and species present, and maximum diversity would occur when the maximum number of tree sizes and species are present. Mathematically, this is described with a uniform distribution:

\[ f(X) = \frac{1}{\max(X) - \min(X)} , \min(X) \leq X \leq \max(X) \]  

[5.56]
where: $X$ represents the variable of interest, which may be dbh or height;

\[
\min(X) \text{ is the minimum value of } X; \text{ and} \\
\max(X) \text{ is the maximum value of } X.
\]

The uniform distribution may also be extended to the bivariate case, for simultaneous consideration of dbh and height:

\[
g(X, Y) = \frac{1}{\max(X) - \min(X)} \cdot \frac{1}{\max(Y) - \min(Y)} \cdot \begin{cases} 
\min(X) \leq X \leq \max(X) \\
\min(Y) \leq Y \leq \max(Y)
\end{cases}
\]

where: $X$ and $Y$ represent the variables of interest, which may be dbh and height;

\[
\min(X) \text{ and } \min(Y) \text{ are the minimum values of } X \text{ and } Y, \text{ respectively; and} \\
\max(X) \text{ and } \max(Y) \text{ are the maximum values of } X \text{ and } Y, \text{ respectively.}
\]

By convention, if $X$ is distributed as a univariate uniform variate over a range $r_1$ to $r_2$, it is denoted as $X \sim U[r_1, r_2]$. If $X$ and $Y$ are distributed jointly as bivariate uniform variates with ranges for each of $X$ and $Y$ of $r_{11}$ to $r_{12}$ and $r_{21}$ to $r_{22}$, respectively, their joint distribution is denoted as $(X, Y) \sim U[(r_{11}, r_{12}), (r_{21}, r_{22})]$.

Using a bivariate uniform distribution over the possible range of diameter and height as the standard for maximum diversity, $(dbh, height) \sim U[(\min(dbh), \max(dbh)), (\min(height), \max(height))]$.

**5.1.1.2 Simulated Stands of Varying Diversity**

Simulated stands were used to represent the most diverse case (the baseline), as well as other cases representing various degrees of diversity. Trees of various sizes can be generated
assuming a variety of distributions. There is evidence to suggest that the distribution of sizes of real tree data may approximate known distributions. For instance, the distribution of stems per hectare of single-cohort, single-species stands have been characterized as following the normal distribution, whereas well-differentiated stands can be bimodal or skewed (Oliver and Larson 1996). A skewed distribution may result when intermediate or suppressed trees die readily, or in very old or very young stands (Oliver and Larson 1996). In mixed species stands where tolerant suppressed trees form lower strata, a bimodal distribution may result (Oliver and Larson 1996).

In order to evaluate the possible structural diversity measures, stands were simulated that represent different degrees of diversity in terms of the spread and shape of the distribution of tree sizes by basal area. To represent these distributions, notation similar to that introduced for the uniform will be used. If $X$ is distributed as a univariate normal variate with mean $\mu$ and standard deviation $\sigma$, it is denoted as $X \sim N[\mu, \sigma]$. The bivariate normal notation for $X \sim N[\mu_1, \sigma_1]$ and $Y \sim N[\mu_2, \sigma_2]$ is $(X, Y) \sim N[(\mu_1, \sigma_1), (\mu_2, \sigma_2)]$. Given set ranges of dbh and height, Figure 5.1 shows the possible combinations of values of dbh versus height that may exist in a simulated sample plot.
Simulated plots were generated under ten descriptions:

1. Very diverse: \((dbh, height) \sim U[(r_{11}, r_{12}), (r_{21}, r_{22})]\)

2. Moderately diverse and nominally large, with a range two-thirds that of case one and centred at the upper end: \((dbh, height) \sim U[((2r_{11}+r_{12})/3, r_{12}), ((2r_{21}+r_{22})/3, r_{22})]\)

3. Moderately diverse and nominally small, with a range two-thirds that of case one and centred at the lower end: \((dbh, height) \sim U[(r_{11}, (2r_{12}+r_{11})/3), (r_{21}, (2r_{22}+r_{21})/3)]\)

4. Not very diverse and nominally large, with a range one-third that of case one and centred at the upper end: \((dbh, height) \sim U[((2r_{12}+r_{11})/3, r_{12}), ((2r_{22}+r_{21})/3, r_{22})]\)

5. Not very diverse and nominally mid-range, with a range one-third that of case one and centred at the mean of case 1: \((dbh, height) \sim U[((2r_{11}+r_{12})/3, (2r_{12}+r_{11})/3), ((2r_{21}+r_{22})/3, (2r_{22}+r_{21})/3)]\)

6. Not very diverse and nominally small, with a range one-third that of case one and centred at the lower end: \((dbh, height) \sim U[(r_{11}, (2r_{12}+r_{11})/3), (r_{21}, (2r_{21}+r_{22})/3)]\)
7. Moderately diverse and normally distributed, with a mean equal to that of case one and a standard deviation of 20% of the mean: 

\[(dbh, height) \sim N[((r_{1l} + r_{12})/2, (r_{1l} + r_{12})/10, (r_{2l} + r_{22})/2, (r_{2l} + r_{22})/10)]\]

8. Not very diverse and normally distributed, with a mean equal to that of case one and a standard deviation of 10% of the mean: 

\[(dbh, height) \sim N[((r_{1l} + r_{12})/2, (r_{1l} + r_{12})/20, (r_{2l} + r_{22})/2, (r_{2l} + r_{22})/20)]\]

9. Moderately diverse and bimodally distributed, with half the basal area distributed as in case four and half the points distributed as in case six.

10. Not very diverse and maximally bimodal, with half the basal area at \((r_{1l}, r_{12})\), and half the basal area at \((r_{12}, r_{22})\).

Theoretically, all of the above simulated plots are possible; however, there are many values that are biologically impossible. To avoid unreasonable dbh-height combinations, and to ensure that the marginal distributions of dbh and height were uniform (where applicable), restrictions were applied. For instance, in simulated case one, univariate uniform distributions of dbh and height were desired. In order to accomplish this without producing unreasonable \((dbh, height)\) pairs, the uniform random variables for dbh and height were generated in twelve very small ranges. This resulted in \(dbh \sim U[r_{1l}, r_{12}]\) and \(height \sim U[r_{2l}, r_{22}]\). Although the joint distribution of \((dbh, height)\) is not a bivariate uniform distribution, the resulting simulated plots were more biologically reasonable. Also, for the normal cases (seven and eight), randomly generated \(dbh/height\) ratios might be outside of a reasonable range. Thus, dbh and height values were generated with restrictions on the \(dbh/height\) ratio.
Figure 5.2 shows the possible values of dbh and height for each of the numbered simulated plots in graphical form.

The dbh and height ranges for the simulations were selected to roughly correspond with the ranges found in the MKRF (presented in Section 5.2). Trees grown in this region can be extremely large, reaching heights over 60 m and dbh over 120 cm in some species. Thus, the
simulated ranges were wide, with \( r_{11} = 0.1 \text{ cm} \), \( r_{12} = 120 \text{ cm} \), \( r_{21} = 0.1 \text{ m} \), and \( r_{22} = 60.0 \text{ m} \).

Species distributions were not simulated in these plots; however, species distributions were evaluated for the MKRF inventory presented later (Section 5.2). For ease of presentation, the simulated cases will be referred to with plot type abbreviations (Table 5.1):

<table>
<thead>
<tr>
<th>Simulated Case No.</th>
<th>Plot Type</th>
<th>Plot Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High diversity, with basal area uniformly distributed</td>
</tr>
<tr>
<td>2</td>
<td>Med-Lrg</td>
<td>Medium diversity, with basal area skewed to large trees</td>
</tr>
<tr>
<td>3</td>
<td>Med-Sml</td>
<td>Medium diversity, with basal area skewed to small trees</td>
</tr>
<tr>
<td>4</td>
<td>Low-Lrg</td>
<td>Low diversity, with basal area skewed to large trees</td>
</tr>
<tr>
<td>5</td>
<td>Low-Mid</td>
<td>Low diversity, with basal area skewed to mid-sized trees</td>
</tr>
<tr>
<td>6</td>
<td>Low-Sml</td>
<td>Low diversity, with basal area skewed to small trees</td>
</tr>
<tr>
<td>7</td>
<td>Med-Norm</td>
<td>Medium diversity, with basal area normally distributed</td>
</tr>
<tr>
<td>8</td>
<td>Low-Norm</td>
<td>Low diversity, with basal area normally distributed</td>
</tr>
<tr>
<td>9</td>
<td>Med-Bim</td>
<td>Medium diversity, with basal area bimodally distributed</td>
</tr>
<tr>
<td>10</td>
<td>Low-Bim</td>
<td>Low diversity, with basal area bimodally distributed</td>
</tr>
</tbody>
</table>

The "High diversity" case represents the baseline of maximum diversity as closely as possible, where the entire range of both dbh and height is represented equally in terms of basal area. In terms of stems per hectare, this would be similar to a reverse-J distribution. Because smaller trees represent larger numbers of trees per hectare than larger trees, they are inflated more than large trees when converting from basal area to stems per hectare, resulting in a distribution that is reverse-J shaped. Cases that are of "Medium diversity" have approximately two-thirds the range of data as the maximum diversity case. The "Low diversity" cases have very small ranges, where basal area represents one-third or less of the possible dbh and height ranges.
Simulated dbh and height values were generated using random number generators in SAS (SAS Institute Inc. 1988). For the uniform distributions (simulated cases one through six and nine), the SAS function RANUNI was used to generate random numbers. First, a random number was generated and scaled to conform to the dbh range for the plot type. Then, a second random number was generated and scaled to conform to the corresponding height range. This process was repeated over all dbh-height ranges within a particular plot. For example, there were twelve dbh-height ranges for simulated case 1, and four dbh-height ranges for simulated case four. Equal amounts of basal area were generated separately over all dbh-height ranges for a particular plot.

The SAS function RANNOR was used to generate normally distributed numbers. For the two normally distributed cases (seven and eight), first, the dbh was generated with an underlying mean and variance. Then, the height was generated with an underlying mean and variance. The (dbh, height) pair was then tested for reasonableness by computing the dbh/height ratio. If the ratio did not fall in the range (0.8, 2.4), additional heights were generated until the ratio did fall in the range. This range was again based on the MKRF data; approximately 90% of dbh/height ratios fell in the range from 0.8 to 2.4. To test the normality of the simulated distributions, the SAS procedure PROC UNIVARIATE (SAS Institute Inc 1989) was used. Dbh and height distributions were not significantly different from the normal at $\alpha = 0.05$. Because of the restrictions in generating the data, the joint distribution of dbh and height likely only approximated a bivariate normal distribution.
Each simulated plot consisted of 60 cells (dbh-height combinations), with each cell contributing equally to the total basal area of the plot. This ensured a uniform distribution of basal area over the plot. Each simulated plot type was generated 100 times. Since there are ten plot types, there are 60,000 simulated cells in all. This method of simulation produced tree size distributions which were univariately uniform by basal area (or normal); however, because of the restrictions made to ensure biologically reasonable tree sizes, the stands are not jointly uniform. The method of simulating normal stands was slightly less restrictive and produced stands that may be approximately jointly normal.

5.1.2 Results and Discussion

The indices presented in Section 4.2 were computed with the simulated data. The results were examined and discussed using criteria 8 listed in Section 4.1.

5.1.2.1 Shannon’s Index Extended to Diameter, Height, and Species

The procedure for computing the Extended Shannon Index, using the post-hoc method was as follows:

1. Tree data were classed in ten-centimetre diameter classes, using the diameter outside bark at breast height (1.3 m above ground; dbh): 0 to 10 cm through 110 to 120 cm. Then, Shannon’s Index was computed with $p_i$, the proportion of basal area in the $i$th diameter class, and equation [2.1] to yield $H'_d$.

2. Tree data were classed into five-metre height classes: 0 to 5 m through 55 to 60 m. Then, Shannon’s Index was computed with $p_j$, the proportion of basal area in the $j$th height class, and equation [2.1] to yield $H'_h$.  

62
3. The final index, $H'_{d+h}$, was computed as the average of the diameter and height indices.

The procedure for computing the Extended Shannon Index under the combination method was slightly different. The data were classed simultaneously by diameter and height, using the same classes as with the post-hoc method. Then, Shannon’s Index was computed with $p_{ij}$, the proportion of basal area in the $i$th diameter, $j$th height class, and summed over all classes to yield $H'_{dh}$.

The results of calculating the two types of Shannon's indices with the simulated data are shown in Table 5.2. The average, minimum, maximum, and standard deviation show the spread of the index over the sample plots of the same type. The range of values represents the variability in the generation of the data. The simulated stands were ranked by average index values, with the highest index receiving a rank of one and the lowest receiving a rank of ten. Index values within 0.01 received the same rank.
Table 5.2. Extended Shannon indices computed using simulated stand data.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Average $H'_{d+h}$</th>
<th>Minimum $H'_{d+h}$</th>
<th>Maximum $H'_{d+h}$</th>
<th>Standard Deviation $H'_{d+h}$</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>2.485</td>
<td>2.476</td>
<td>2.506</td>
<td>0.006</td>
<td>1</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>2.079</td>
<td>2.073</td>
<td>2.127</td>
<td>0.008</td>
<td>2</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>2.079</td>
<td>2.077</td>
<td>2.124</td>
<td>0.009</td>
<td>2</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>1.391</td>
<td>1.381</td>
<td>1.466</td>
<td>0.014</td>
<td>6</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>1.392</td>
<td>1.385</td>
<td>1.448</td>
<td>0.014</td>
<td>6</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>1.392</td>
<td>1.386</td>
<td>1.447</td>
<td>0.014</td>
<td>6</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>1.494</td>
<td>1.339</td>
<td>1.643</td>
<td>0.069</td>
<td>5</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>0.950</td>
<td>0.796</td>
<td>1.079</td>
<td>0.068</td>
<td>9</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>2.082</td>
<td>2.073</td>
<td>2.112</td>
<td>0.010</td>
<td>2</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.693</td>
<td>0.693</td>
<td>0.693</td>
<td>0.000</td>
<td>10</td>
</tr>
</tbody>
</table>

The indices associated with the combination index tended to be larger than that of the post-hoc index, as did the standard deviations. This was not unexpected, as the simultaneous relegation of data to dbh and height classes necessarily produced more dbh-height combinations, and hence produced a larger index. Both types of indices were ranked in the same manner, with two exceptions. Whereas the post-hoc index ($H'_{d+h}$) ranked the medium diversity normal stand (Med-Norm) with the three low-diversity stands (Low-Lrg, Low-Med, Low-Sml), the combination index ($H'_{dh}$) ranked Med-Norm higher than that of the high diversity stand (High). The low-diversity normal stand (Low-Norm) also ranked higher when the combination method was used. The normal stands may have ranked higher under the combination method because they were generated in such a way as to be more diverse in a bivariate sense. That is, under the normal simulation method, stands had larger bivariate ranges than those that were generated as uniform distributions, and this could have resulted in illogical orderings.
With respect to the criteria listed in Section 4.1, both methods of computing the Extended Shannon Index produced indices that possess the characteristics listed as 8a, 8b, and 8c. The indices are tree size insensitive, are correlated with the range of diameter and height values, and rank bimodally distributed stands in a logical way. The post-hoc method may have performed better than the combination method in that it produced a more logical ordering of normally distributed stands; however, this may have been a consequence of the method of random number generation.

5.1.2.2 Structural Index based on Curve Fitting

The Weibull function was first fit univariately and separately to the cumulative distribution functions of dbh and height by plot. The regressions were performed using the SAS procedure PROC NLIN and the Marquardt method (SAS Institute Inc. 1989).

The fit statistics of the equations were first examined to ensure that the Weibull provided a good relationship. As shown in Table 5.3, the RMSE was low (below 0.05) for every case except the medium-diversity bimodal (Med-Bim), where it was nearly twice as high. The poor fit of the Weibull function to a bimodal stand was expected, since it has been shown that the Weibull function has only one maximum (Bailey and Dell 1973). The perfect fit (RMSE equal to zero) of the extremely bimodal case was also expected, since the two-parameter regression was fit with only two unique points, which reduced the regression to the solution of two equation in two unknowns.
Table 5.3. Resulting coefficients and RMSE for the univariate Weibull fitted to dbh using simulated stands.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Average RMSE</th>
<th>Average $b_1$</th>
<th>Average $c_1$</th>
<th>Minimum $b_1$</th>
<th>Minimum $c_1$</th>
<th>Maximum $b_1$</th>
<th>Maximum $c_1$</th>
<th>Standard Deviation</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.041</td>
<td>70.43</td>
<td>1.576</td>
<td>68.86</td>
<td>1.515</td>
<td>71.26</td>
<td>1.663</td>
<td>0.475</td>
<td>3</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>0.030</td>
<td>89.15</td>
<td>3.338</td>
<td>87.64</td>
<td>3.189</td>
<td>90.46</td>
<td>3.488</td>
<td>0.451</td>
<td>2</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>0.043</td>
<td>46.49</td>
<td>1.565</td>
<td>45.32</td>
<td>1.430</td>
<td>47.67</td>
<td>1.671</td>
<td>0.487</td>
<td>5</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>0.034</td>
<td>104.67</td>
<td>8.589</td>
<td>103.25</td>
<td>7.698</td>
<td>105.42</td>
<td>9.512</td>
<td>0.412</td>
<td>1</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>0.032</td>
<td>64.52</td>
<td>5.051</td>
<td>63.35</td>
<td>4.574</td>
<td>65.48</td>
<td>5.604</td>
<td>0.450</td>
<td>5</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>0.045</td>
<td>23.52</td>
<td>1.581</td>
<td>22.30</td>
<td>1.342</td>
<td>24.48</td>
<td>1.808</td>
<td>0.448</td>
<td>9</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>0.027</td>
<td>64.05</td>
<td>5.803</td>
<td>60.53</td>
<td>4.116</td>
<td>67.76</td>
<td>7.508</td>
<td>1.608</td>
<td>6</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>0.030</td>
<td>62.08</td>
<td>11.603</td>
<td>59.85</td>
<td>9.255</td>
<td>63.87</td>
<td>15.136</td>
<td>0.841</td>
<td>7</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>0.079</td>
<td>67.76</td>
<td>0.977</td>
<td>66.13</td>
<td>0.935</td>
<td>69.09</td>
<td>1.027</td>
<td>0.610</td>
<td>4</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.000</td>
<td>0.23</td>
<td>0.436</td>
<td>0.23</td>
<td>0.436</td>
<td>0.23</td>
<td>0.436</td>
<td>0.000</td>
<td>10</td>
</tr>
</tbody>
</table>

The resulting average coefficients $b_1$ and $c_1$ were then examined (Table 5.3). The value of the coefficient $b_1$ was roughly correlated to the centre of the distribution. The low-diversity large tree (Low-Lrg), medium-diversity large tree (Med-Lrg), and very diverse (High) cases had the largest values of $b_1$. However, the lowest value of $b_1$ occurred for the low-diversity bimodal distribution (Low-Bim), whereas the distribution centred at the smallest dbh value was the low-diversity small tree case (Low-Sml). There was some overlap in the range of values of $b_1$, as the most diverse case (High) and the medium-diversity bimodal case (Med-Bim) produced similar values of $b_1$, as did the low-diversity mid-sized tree stand (Low-Mid), and the two normally distributed stands (Med-Norm and Low-Norm).

The value of the coefficient $c_1$ was between 1.58 and 3.4 for the more diverse stands (High, Med-Lrg, and Med-Sml). For the most part, a much larger $c_1$ resulted for the
nominally larger and less diverse stands (Low-Lrg and Low-Norm), and a smaller \( c_1 \) resulted for the nominally smaller and less diverse stand (Low-Sml). However, the coefficient \( c_1 \) was very small for the two bimodal stands (Med-Bim and Low-Bim). There was considerable overlap in the values of \( c_1 \), with three groups (High, Med-Low, and Low-Sml; Low-Mid and Med-Norm; and Low-Lrg and Low-Norm) that produced similar values.

The resulting average coefficients, \( b_2 \) and \( c_2 \) for height are listed in Table 5.4, along with the average RMSE. The RMSE and resulting coefficient \( b_2 \) tracked with the results for dbh, as did the cases with overlapping coefficients. The Low-Lrg, Med-Lrg, and High cases had the largest values of \( b_2 \); the lowest values of \( b_2 \) occurred in the Low-Bim case. The value of the coefficient \( c_2 \) was between 1.56 and 3.34 for the more diverse stands (High, Med-Lrg, and Med-Sml); however, the value of \( c_2 \) also fell in this range for the Low-Sml case. As in the univariate dbh case, for the most part, a larger \( c_2 \) resulted for the nominally larger and less diverse stands (Low-Lrg and Low-Norm), and the coefficient \( c_2 \) was very small for the bimodal cases.

Table 5.4. Resulting coefficients and RMSE for the univariate Weibull fitted to height using simulated stands.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Average RMSE</th>
<th>( b_2 )</th>
<th>( c_2 )</th>
<th>Minimum ( b_2 )</th>
<th>( c_2 )</th>
<th>Maximum ( b_2 )</th>
<th>( c_2 )</th>
<th>Standard Deviation</th>
<th>( b_2 )</th>
<th>( c_2 )</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.041</td>
<td>35.20</td>
<td>1.564</td>
<td>34.61</td>
<td>1.481</td>
<td>35.72</td>
<td>1.648</td>
<td>0.229</td>
<td>0.031</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>0.030</td>
<td>44.56</td>
<td>3.339</td>
<td>43.98</td>
<td>3.068</td>
<td>45.10</td>
<td>3.489</td>
<td>0.218</td>
<td>0.071</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>0.042</td>
<td>23.17</td>
<td>1.571</td>
<td>22.71</td>
<td>1.437</td>
<td>23.62</td>
<td>1.733</td>
<td>0.213</td>
<td>0.048</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>0.035</td>
<td>52.34</td>
<td>8.494</td>
<td>51.70</td>
<td>7.700</td>
<td>52.91</td>
<td>9.234</td>
<td>0.238</td>
<td>0.319</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>0.033</td>
<td>32.26</td>
<td>5.087</td>
<td>31.65</td>
<td>4.585</td>
<td>32.82</td>
<td>5.681</td>
<td>0.221</td>
<td>0.208</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>0.044</td>
<td>11.77</td>
<td>1.596</td>
<td>11.27</td>
<td>1.391</td>
<td>12.41</td>
<td>1.813</td>
<td>0.234</td>
<td>0.072</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>0.029</td>
<td>34.29</td>
<td>7.425</td>
<td>32.67</td>
<td>5.612</td>
<td>35.79</td>
<td>9.681</td>
<td>0.713</td>
<td>0.860</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>0.032</td>
<td>31.51</td>
<td>12.970</td>
<td>30.74</td>
<td>10.19</td>
<td>32.46</td>
<td>19.616</td>
<td>0.342</td>
<td>1.655</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>0.079</td>
<td>33.83</td>
<td>0.982</td>
<td>32.96</td>
<td>0.934</td>
<td>34.89</td>
<td>1.040</td>
<td>0.428</td>
<td>0.024</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.000</td>
<td>0.21</td>
<td>0.490</td>
<td>0.21</td>
<td>0.490</td>
<td>0.21</td>
<td>0.490</td>
<td>0.000</td>
<td>0.000</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
The investigation of the univariate Weibull distributions did not lead to a useful index of diversity. No coefficient was found to possess the desirable characteristics presented as item 8 (Section 4.1), in that the coefficients produced were tree size sensitive and did not differentiate between stands of low, medium, and high diversity.

The results of bivariate Weibull curve fitting follow in Tables 5.5 to 5.7. The bivariate Weibull models were fit to the joint cumulative distribution function of dbh and height. As in the univariate case, the regressions were performed using the SAS procedure PROC NLIN and the Marquardt method (SAS Institute Inc. 1989). Starting values for $b_1$, $b_2$, $c_1$, and $c_2$ were obtained from the univariate regressions. The RMSE was first analyzed to ensure that the Weibull provided a good fit to the data. As shown in Table 5.5, RMSE values were low (less than 0.05) for all but the bimodal stands. The upper limiting value for the coefficient $a_3$ (1.00) resulted for all stands except those based on the normal distribution (Table 5.5). Convergence to this maximum value may have been a result of lack of fit. Because data were generated under univariate uniform conditions, but not uniform bivariate conditions, the bivariate Weibull function may not have been appropriate for the uniformly generated stands.
Table 5.5. RMSE and resulting values of coefficient $a_3$ for the bivariate Weibull fitted to simulated stands.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>RMSE</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.045</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>0.047</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>0.046</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>0.053</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>0.049</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>0.043</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>0.029</td>
<td>0.923</td>
<td>-0.704</td>
<td>1.000</td>
<td>0.217</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>0.029</td>
<td>0.645</td>
<td>-0.237</td>
<td>1.000</td>
<td>0.375</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>0.059</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.076</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

As in the univariate case, values of $b_1$ and $b_2$ were roughly correlated to size (Table 5.6); the simulated stands with the largest average tree size (Med-Lrg and Low-Lrg) had the largest values for $b_1$ and $b_2$. As in the univariate results, the coefficient ranges for the High, Med-Norm, Low-Norm, Low-Mid, and Med-Bim cases overlapped.

Table 5.6. Resulting coefficients $b_1$ and $b_2$ for the bivariate Weibull fitted to simulated stands.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Average $b_1$</th>
<th>Minimum $b_1$</th>
<th>Maximum $b_1$</th>
<th>Standard Deviation $b_1$</th>
<th>Rank $b_1$</th>
<th>Average $b_2$</th>
<th>Minimum $b_2$</th>
<th>Maximum $b_2$</th>
<th>Standard Deviation $b_2$</th>
<th>Rank $b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>61.52</td>
<td>30.78</td>
<td>58.08</td>
<td>29.71</td>
<td>63.51</td>
<td>32.08</td>
<td>0.97</td>
<td>0.49</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>84.10</td>
<td>41.90</td>
<td>81.39</td>
<td>40.46</td>
<td>86.87</td>
<td>43.41</td>
<td>1.37</td>
<td>0.67</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>40.87</td>
<td>20.16</td>
<td>38.86</td>
<td>18.99</td>
<td>43.01</td>
<td>21.10</td>
<td>0.92</td>
<td>0.42</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>102.46</td>
<td>51.16</td>
<td>99.27</td>
<td>49.57</td>
<td>104.94</td>
<td>52.48</td>
<td>1.51</td>
<td>0.72</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>62.20</td>
<td>31.08</td>
<td>58.69</td>
<td>29.51</td>
<td>64.97</td>
<td>32.49</td>
<td>1.38</td>
<td>0.69</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>20.85</td>
<td>10.39</td>
<td>17.91</td>
<td>9.40</td>
<td>23.42</td>
<td>11.61</td>
<td>1.16</td>
<td>0.52</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>64.00</td>
<td>33.99</td>
<td>60.21</td>
<td>31.80</td>
<td>67.86</td>
<td>35.81</td>
<td>1.78</td>
<td>0.89</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>62.25</td>
<td>31.52</td>
<td>59.56</td>
<td>30.73</td>
<td>64.26</td>
<td>32.51</td>
<td>0.96</td>
<td>0.40</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>56.16</td>
<td>27.92</td>
<td>46.29</td>
<td>21.65</td>
<td>68.64</td>
<td>33.71</td>
<td>5.02</td>
<td>2.53</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.00</td>
<td>0.00</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Values of the coefficients $c_1$ and $c_2$ are shown in Table 5.7. The more diverse cases had average values of $c_1$ and $c_2$ in the range 1.3 to 3.2. However, the Low-Sml case also had values of $c_1$ and $c_2$ in this range. As in the univariate case, values of $c_1$ and $c_2$ were affected by the range and centre of the distribution, with higher values of $c_1$ and $c_2$ associated with larger ranges and mean values. The bivariate values of $c_1$ and $c_2$ also were prone to large ranges, and many of the values of these coefficients overlapped from case to case.

Table 5.7. Resulting coefficients $c_1$ and $c_2$ for the bivariate Weibull fitted to simulated stands.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Average $c_1$</th>
<th>Minimum $c_1$</th>
<th>Maximum $c_1$</th>
<th>Average $c_2$</th>
<th>Minimum $c_2$</th>
<th>Maximum $c_2$</th>
<th>Standard Deviation $c_1$</th>
<th>Standard Deviation $c_2$</th>
<th>Rank $c_1$</th>
<th>Rank $c_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1.371</td>
<td>1.356</td>
<td>1.254</td>
<td>1.483</td>
<td>1.452</td>
<td>0.038</td>
<td>0.041</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>3.181</td>
<td>3.214</td>
<td>2.923</td>
<td>2.880</td>
<td>3.541</td>
<td>3.585</td>
<td>0.139</td>
<td>0.154</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>1.341</td>
<td>1.366</td>
<td>1.207</td>
<td>1.206</td>
<td>1.494</td>
<td>1.512</td>
<td>0.059</td>
<td>0.061</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>8.462</td>
<td>8.472</td>
<td>7.016</td>
<td>6.711</td>
<td>10.142</td>
<td>10.621</td>
<td>0.796</td>
<td>0.888</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>4.900</td>
<td>4.989</td>
<td>4.054</td>
<td>3.879</td>
<td>5.948</td>
<td>6.350</td>
<td>0.446</td>
<td>0.506</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>1.350</td>
<td>1.378</td>
<td>1.055</td>
<td>1.127</td>
<td>1.639</td>
<td>1.767</td>
<td>0.115</td>
<td>0.129</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>5.999</td>
<td>7.285</td>
<td>4.036</td>
<td>5.298</td>
<td>8.994</td>
<td>10.855</td>
<td>0.795</td>
<td>1.079</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>0.905</td>
<td>0.912</td>
<td>0.791</td>
<td>0.790</td>
<td>1.073</td>
<td>1.071</td>
<td>0.060</td>
<td>0.065</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.436</td>
<td>0.490</td>
<td>0.436</td>
<td>0.490</td>
<td>0.436</td>
<td>0.490</td>
<td>0.000</td>
<td>0.000</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

The investigation of the bivariate Weibull distribution did not lead to a useful index of diversity. No coefficient was found to possess the desirable characteristics presented as item 8 (Section 4.1), in that the coefficients produced were tree size sensitive and did not differentiate between stands of low, medium, and high diversity.
5.1.2.3 Structural Index based on $R^2$ (STRI)

Both the univariate and bivariate forms of the STRI were computed with the maximum and minimum dbh and height values used to generate the simulated stands. Therefore, the theoretical distributions used in calculating the univariate STRI were:

\[ \hat{F}_1(dbh) = \frac{dbh - 0.1}{120.0 - 0.1}, \quad 0.1 \leq dbh \leq 120.0 \]  \[5.58\]

\[ \hat{F}_2(height) = \frac{height - 0.1}{60.0 - 0.1}, \quad 0.1 \leq height \leq 60.0 \]  \[5.59\]

The theoretical distribution used in calculating the bivariate STRI was:

\[ \hat{G}(dbh, height) = \left[ \frac{dbh - 0.1}{120.0 - 0.1} \right] \times \left[ \frac{height - 0.1}{60.0 - 0.1} \right], \quad \{ 0.1 \leq dbh \leq 120.0 \}, \quad \{ 0.1 \leq height \leq 60.0 \} \]  \[5.60\]

Results of the STRI methods follow in Table 5.8. Index values within 0.01 received the same rank.

Table 5.8. Univariate (STRI$_{d+h}$) and bivariate (STRI$_{dh}$) STRI using simulated stands.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Average STRI$_{d+h}$</th>
<th>Minimum STRI$_{d+h}$</th>
<th>Maximum STRI$_{d+h}$</th>
<th>Average STRI$_{dh}$</th>
<th>Minimum STRI$_{dh}$</th>
<th>Maximum STRI$_{dh}$</th>
<th>Standard Deviation</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.999</td>
<td>0.927</td>
<td>0.999</td>
<td>0.927</td>
<td>0.999</td>
<td>0.935</td>
<td>0.000</td>
<td>1</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>0.895</td>
<td>0.882</td>
<td>0.881</td>
<td>0.866</td>
<td>0.903</td>
<td>0.896</td>
<td>0.004</td>
<td>3</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>0.877</td>
<td>0.562</td>
<td>0.871</td>
<td>0.547</td>
<td>0.883</td>
<td>0.579</td>
<td>0.003</td>
<td>5</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>0.582</td>
<td>0.121</td>
<td>0.529</td>
<td>-0.018</td>
<td>0.640</td>
<td>0.259</td>
<td>0.022</td>
<td>9</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>0.892</td>
<td>0.778</td>
<td>0.881</td>
<td>0.761</td>
<td>0.905</td>
<td>0.794</td>
<td>0.005</td>
<td>3</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>0.548</td>
<td>0.168</td>
<td>0.539</td>
<td>0.159</td>
<td>0.561</td>
<td>0.179</td>
<td>0.005</td>
<td>10</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>0.877</td>
<td>0.794</td>
<td>0.855</td>
<td>0.726</td>
<td>0.899</td>
<td>0.830</td>
<td>0.009</td>
<td>5</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>0.821</td>
<td>0.685</td>
<td>0.792</td>
<td>0.638</td>
<td>0.846</td>
<td>0.721</td>
<td>0.009</td>
<td>7</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>0.972</td>
<td>0.894</td>
<td>0.967</td>
<td>0.874</td>
<td>0.976</td>
<td>0.908</td>
<td>0.002</td>
<td>2</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.800</td>
<td>0.571</td>
<td>0.800</td>
<td>0.571</td>
<td>0.800</td>
<td>0.571</td>
<td>0.000</td>
<td>8</td>
</tr>
</tbody>
</table>
The very diverse and medium-diverse bimodal stands (High and Med-Bim, respectively) ranked the highest under both the univariate and bivariate versions of this index. The lowest rank under both types of STRI occurred in the low diversity large and small tree stands (Low-Lrg and Low-Sml, respectively). As expected, higher values of STRI appeared to be correlated with distributions centred at the middle of the range. There was much overlap in STRI_{d+h} values for most of the middle-centred stands (Med-Sml, Med-Lrg, Low-Mid, and Low-Norm); however, the STRI_{dh} index values overlapped less. The Med-Lrg and Med-Sml cases produced very different STRI_{dh} index values, whereas the STRI_{d+h} values for these two stands were very close.

Due to the nature of the $R^2$ calculation, the middle-centred distributions, such as Low-Mid, Med-Norm, and Low-Norm, will always receive higher STRI values than skewed or bimodal distributions. Both types of STRI indices are not desirable in that they are tree size sensitive, are not well-correlated to range, and do not rank bimodal stands logically. Thus, the STRI does not possess the desirable characteristic presented as item 8 (Section 4.1).

5.1.2.4 Structural Index based on Variance Comparisons (STVI)

Both the univariate and bivariate STVI were computed with the maximum and minimum dbh and height values used in generating the simulated data. Thus, the theoretical variances for univariate uniform distributions of $dbh$, $height$, and the joint distribution of $(dbh, height)$ were:

$$S^2_{dbh_u} = (120.0 - 0.1)^2 / 12 = 1198$$  \[5.61\]

$$S^2_{height_u} = (60.0 - 0.1)^2 / 12 = 299$$  \[5.62\]
\[ S_U^2 = (120.0 - 0.1)^2 \times (60.0 - 0.1)^2 / 144 = 358,203 \] \[ 5.63 \]

The maximum variances used in the STVI calculations were:

\[ S_{dbh_{\text{max}}}^2 = (120.0 - 0.1)^2 / 4 = 3,594 \] \[ 5.64 \]

\[ S_{\text{height}_{\text{max}}}^2 = (60.0 - 0.1)^2 / 4 = 897 \] \[ 5.65 \]

\[ S_{\text{max}}^2 = (120.0 - 0.1)^2 \times (60.0 - 0.1)^2 / 16 = 3,223,829 \] \[ 5.66 \]

Results of calculating indices with the STVI follow in Table 5.9. Index values within 0.01 received the same rank. The index values associated with the univariate method were much larger than that of the bivariate method, as were the standard deviations.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Standard Deviation</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STVI&lt;sub&gt;dbh&lt;/sub&gt;</td>
<td>STVI&lt;sub&gt;dh&lt;/sub&gt;</td>
<td>STVI&lt;sub&gt;dbh&lt;/sub&gt;</td>
<td>STVI&lt;sub&gt;dh&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>0.946</td>
<td>0.032</td>
<td>0.894</td>
<td>1.000</td>
<td>0.024</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>0.772</td>
<td>0.014</td>
<td>0.747</td>
<td>0.798</td>
<td>0.011</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>0.761</td>
<td>0.014</td>
<td>0.730</td>
<td>0.787</td>
<td>0.011</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>0.248</td>
<td>0.003</td>
<td>0.223</td>
<td>0.271</td>
<td>0.010</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>0.252</td>
<td>0.004</td>
<td>0.226</td>
<td>0.279</td>
<td>0.011</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>0.251</td>
<td>0.004</td>
<td>0.231</td>
<td>0.274</td>
<td>0.009</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>0.230</td>
<td>0.020</td>
<td>0.170</td>
<td>0.325</td>
<td>0.031</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>0.065</td>
<td>0.002</td>
<td>0.047</td>
<td>0.082</td>
<td>0.009</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>0.620</td>
<td>0.047</td>
<td>0.600</td>
<td>0.646</td>
<td>0.009</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>0.086</td>
<td>0.000</td>
<td>0.086</td>
<td>0.086</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The highest values of the univariate STVI occurred for the very diverse case (High), followed by the medium-diverse uniform stands (Med-Lrg and Med-Sml). Under the
bivariate method, however, the highest rank occurred for the medium-diverse bimodal case (Med-Bim), followed by the very diverse case. As was the case with the Extended Shannon indices, the normally distributed cases ranked higher using the bivariate method. Under the univariate method, the indices for the medium–diverse normal stand (Med-Norm) were ranked far below the other medium-diverse stands. However, under the bivariate method, Med-Norm was ranked above the other medium-diverse stands. Both methods ranked the three low-diversity uniform stands (Low-Lrg, Low-Mid, and Low-Sml) in the same range, with the low-diversity normal stand (Low-Norm) slightly lower. Higher values of STVI were correlated with widely distributed stands. The overlap that occurred was mostly between stands of similar diversity (e.g., Low-Lrg, Low-Mid, and Low-Sml; and Med-Lrg, Med-Sml, and Med-Norm). However, the high diversity and medium-diverse bimodal cases also overlapped in range.

The univariate STVI method produced indices that possess the desirable characteristics listed under item 8 (Section 4.1). The index is tree size insensitive, is correlated to continuously occupied ranges, and gives bimodal stands logical index values. The bivariate index did not produce indices that possessed these characteristics, in that the indices did not correlate well with continuously occupied ranges and made illogical rankings with respect to the high and medium diversity uniform simulations. However, this may have been a function of the methods used to generate the simulated stands. Simulated uniform stands were generated such that they followed a univariate uniform distribution, and were thus well described by the univariate STVI. However, simulated uniform stands were not generated to
follow a bivariate distribution. One consequence of this is that the sample variance of the stands never reached the value of the theoretical uniform variance (Table 5.10).

Table 5.10. Average sample variances for simulated cases compared to the variance of a theoretical uniform distribution.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Dbh $S^2_{dbh}$</th>
<th>Rank</th>
<th>Height $S^2_{height}$</th>
<th>Rank</th>
<th>Joint (Dbh, Height) $S^2_{dbh, height}$</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1,219</td>
<td>3</td>
<td>306</td>
<td>3</td>
<td>4,815</td>
<td>2</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>549</td>
<td>4</td>
<td>137</td>
<td>4</td>
<td>2,141</td>
<td>4</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>539</td>
<td>5</td>
<td>133</td>
<td>5</td>
<td>2,075</td>
<td>5</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>133</td>
<td>9</td>
<td>34</td>
<td>8</td>
<td>520</td>
<td>8</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>137</td>
<td>7</td>
<td>34</td>
<td>7</td>
<td>523</td>
<td>7</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>136</td>
<td>8</td>
<td>34</td>
<td>6</td>
<td>527</td>
<td>6</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>143</td>
<td>6</td>
<td>26</td>
<td>9</td>
<td>2,959</td>
<td>3</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>36</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>256</td>
<td>9</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>1,768</td>
<td>2</td>
<td>441</td>
<td>2</td>
<td>7,052</td>
<td>1</td>
</tr>
<tr>
<td>Low-Bim</td>
<td>3,655</td>
<td>1</td>
<td>912</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

$S^2_U$ 1,198 299 358,203

When dbh and height are considered univariately, the high-diversity case has a variance near that of the uniform, and the two bimodal cases have variances greater than that of the uniform. However, when dbh and height are considered jointly (bivariately), all simulations have variances much less than that of the uniform. The medium-diverse bimodal case had the highest generalized variance of all the simulated cases, but this variance was not high enough to use the right-hand side of the STVI curve. Thus, the univariate STVI may be preferred over the bivariate STVI, in that it may better meet the criteria presented in Section 4.1.
5.1.3 Overall Discussion of Proposed Indices Computed with Simulated Data

The proposed indices were evaluated with respect to the criteria presented in Section 4.1. Although the Extended Shannon indices rely on combining data into classes, they ranked the simulated stands in a logical manner and possess the desirable characteristics presented as item 8 in Section 4.1. Curve fitting techniques may provide a method to derive indices without relying on class data; however, further analysis is required with more flexible curves, such as the Beta distribution, before deciding whether this approach shows promise. The STRI was derived to measure the deviation between empirical data and the uniform distribution; however, due to the properties of $R^2$, this index did not discriminate well between simulated cases of varying diversity. All of the desirable characteristics presented in Section 4.1 were met by the univariate STVI. The STVI in its bivariate form did not perform as well; however, this may have been a consequence of the methods used to generate the simulated data. Further investigation with real data is necessary to evaluate this measure.

The most desirable indices of those presented are the Extended Shannon indices and the STVI. The univariate STVI meets all of the criteria laid out in Section 4.1; however, it lacks a theoretical basis. While Shannon’s Index has a solid theoretical background, it may be undesirable because it relies on combining data into classes.

5.2 Malcolm Knapp Research Forest Data

Evaluation of the Extended Shannon indices and the STVI was also performed with data collected in a recent forest inventory of the University of British Columbia Malcolm Knapp Research Forest (MKRF). The approximately 5000 ha MKRF is located in the Coastal
Western Hemlock (CWH) biogeoclimatic (BEC) zone of southwestern British Columbia (BC), where the maximum tree size for most species is quite large, reaching over 1 m in dbh and over 50 m in height. The region encompasses low to middle elevations west of the Coastal Mountains and is the wettest BEC zone in BC, receiving 1000 to 4000 mm of precipitation annually (Meidinger and Pojar 1991). The climate is cool mesothermal, with a mean annual temperature of 8° C and mild winters. The most common species in the forest cover are Douglas-fir (*Pseudotsuga menzeisii* (Mirb.) Franco), western hemlock (*Tsuga heterophylla* (Raf.) Sarg.), western red cedar (*Thuja plicata* Donn), amabilis fir (*Abies amabilis* (Dougl.) Forbes), and Sitka spruce (*Picea sitchensis* (Bong.) Carr).

5.2.1 Methods

In 1995, sample plot data were collected using the Ministry of Forests, Vegetation Inventory Sampling Procedures (Resources Inventory Branch 1994) as the basis for plot establishment, plot size, and measurement procedures. Eighty-two clusters of plots were systematically located over the research forest using square spacing. At each grid intersection, a central plot was located, and then additional plot centres were located in each cardinal direction, 25 m from the main plot centre. If the total basal area of a stand was estimated to be less than 15 m²/ha, fixed radius plots were used; otherwise, variable radius plots were used. Because plots occurred throughout the whole of the MKRF, the data represent a variety of microsites and growing conditions.

For all trees above two cm dbh, the species, dbh (cm) and tree class (live or dead, standing or fallen) were obtained. For a subset of trees, the total height (m), crown class...
(dominant, codominant, intermediate, or suppressed), and height to live crown (m) were also measured. The 82 clusters yielded 327 plots containing 1,832 live trees. Height was measured on only 763 trees (44%). Height prediction models were developed to estimate the remaining heights.

The major species in the MKRF are western red cedar, Douglas-fir, western hemlock, and alder (*Alnus rubra* Bong. and *Alnus tenuifolia* Nutt.). Height prediction equations were fit to these four species groups separately. Models were not fit separately for other species because of few observations. An equation was developed for all other broad-leaved species by grouping them with the two species of alder. Heights for the remaining coniferous species were obtained by using the prediction equations for species with similar growth patterns, as shown in Table 5.11.

Table 5.11. Data for height-dbh equation development by species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of Trees</th>
<th>Grouped With (If Applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Fir (<em>Abies amabilis</em> (Dougl.) Forbes)</td>
<td>1</td>
<td>Douglas-fir</td>
</tr>
<tr>
<td>Sitka Spruce (<em>Picea sitchensis</em> (Bong.) Carr.)</td>
<td>1</td>
<td>Douglas-fir</td>
</tr>
<tr>
<td>Red Cedar (<em>Thuja plicata</em> Donn)</td>
<td>210</td>
<td>-</td>
</tr>
<tr>
<td>Alder (<em>Alnus rubra</em> Bong. and <em>A. tenuifolia</em> Nutt.)</td>
<td>28</td>
<td>-</td>
</tr>
<tr>
<td>Douglas-fir (<em>Pseudotsuga menzeisii</em> (Mirb.) Franco)</td>
<td>230</td>
<td>-</td>
</tr>
<tr>
<td>Other Broad-leaved Trees</td>
<td>22</td>
<td>Alder</td>
</tr>
<tr>
<td>Western Hemlock (<em>Tsuga heterophylla</em> (Raf.) Sarg.)</td>
<td>261</td>
<td>-</td>
</tr>
<tr>
<td>Lodgepole Pine (<em>Pinus contorta</em> Dougl.)</td>
<td>8</td>
<td>W. Hemlock</td>
</tr>
<tr>
<td>Yew (<em>Taxus Brevifolia</em> Nutt.)</td>
<td>3</td>
<td>Red cedar</td>
</tr>
</tbody>
</table>

Twelve models were evaluated by species (see Kovacs and LeMay 1999a, and Kovacs and LeMay 1999b). Based on these evaluations, the final height prediction equations by
species were selected. Where trees were not measured for height, the following height prediction equations were used:

**Alder:**

\[ \text{height} = 1.3 + (68.0529 + 0.019287 \times \text{TSTEMSHA} - 0.77315 \times \text{TBAHA}) \times \\
[1 - \exp((-0.00877 + 0.00000044 \times \text{TSTEMSHA} - 0.00061 \times \text{TBAHA}) \times \\
\text{dbh}^{0.64991 - 0.00008194 \times \text{TSTEMSHA} + 0.00554 \times \text{TBAHA}}] \]

[5.67]

**Cedar and yew:**

\[ \text{height} = 1.3 + (59.5789 + 12.5373 \times \text{REL_dbh} - 36.98 \times \text{REL_dbh}^2) \times \\
[1 - \exp((0.00071 - 0.03818 \times \text{REL_dbh} + 0.01551 \times \text{REL_dbh}^2) \times \\
\text{dbh}^{1.6577 - 2.02199 \times \text{REL_dbh} + 1.4599 \times \text{REL_dbh}^2})] \]

[5.68]

**Douglas-fir, true fir, and Sitka spruce:**

\[ \text{height} = 1.3 + (15.759 + 767.419 \times \text{REL_dbh} - 699.761 \times \text{REL_dbh}^2) \times \\
[1 - \exp((0.00025 - 0.00332 \times \text{REL_DBH} - 0.0118 \times \text{REL_dbh}^2) \times \\
\text{dbh}^{2.0754 - 3.3873 \times \text{REL_dbh} + 2.1785 \times \text{REL_dbh}^2})] \]

[5.69]

**Hemlock and lodgepole pine:**

\[ \text{height} = 1.3 + (-1.25839 + 163.146 \times \text{REL_dbh} - 126.716 \times \text{REL_dbh}^2) \times \\
[1 - \exp((0.00021 - 0.06091 \times \text{REL_DBH} + 0.05552 \times \text{REL_dbh}^2) \times \\
\text{dbh}^{2.34135 - 4.11239 \times \text{REL_dbh} + 3.31119 \times \text{REL_dbh}^2})] \]

[5.70]

**All other broad-leaved species:**

\[ \text{height} = 1.3 + (1.07897 - 1.69747 \times \text{REL_dbh} + 0.772644 \times \text{REL_dbh}^2) \times \\
(0.868644 + 0.276184 \times \text{REL_DBH} - 0.198905 \times \text{REL_dbh}^2) \times \\
\text{dbh}^{1.75194 - 1.17265 \times \text{REL_dbh} + 1.3268 \times \text{REL_dbh}^2})] \]

[5.71]

Since the analysis and presentation of all 82 cluster plots would be cumbersome, nine plots were selected to match as closely as possible the dbh and height distributions in simulated cases one through nine; a plot similar to plot ten was not found in the MKRF data. Descriptive statistics for the selected plots are given in Table 5.12.
Table 5.12. Descriptive statistics of selected MKRF plots.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>MKRF Species</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>57</td>
<td>52.80</td>
<td>3</td>
<td>51.5</td>
<td>11.70</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>09</td>
<td>84.80</td>
<td>3</td>
<td>66.2</td>
<td>11.13</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>46</td>
<td>34.55</td>
<td>3</td>
<td>42.9</td>
<td>8.82</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>58</td>
<td>58.24</td>
<td>3</td>
<td>99.3</td>
<td>7.33</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>49</td>
<td>56.17</td>
<td>3</td>
<td>40.8</td>
<td>6.97</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>52</td>
<td>12.47</td>
<td>3</td>
<td>15.9</td>
<td>3.79</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>15</td>
<td>44.84</td>
<td>3</td>
<td>42.2</td>
<td>6.07</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>77</td>
<td>33.74</td>
<td>3</td>
<td>30.2</td>
<td>4.47</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>45</td>
<td>95.42</td>
<td>3</td>
<td>68.3</td>
<td>16.74</td>
</tr>
</tbody>
</table>

Graphs of the distribution of basal area by plot and species are found in Appendix I.
Plots were chosen on the basis of dbh distribution and height distribution for all species combined, and thus, the distributions by species may differ substantially from the chosen plot type.

5.2.2 Results and Discussion

The Extended Shannon Index and the STVI were computed with the MKRF data by five-point cluster plot, using all live trees. The indices were examined using criteria 8 listed in Section 4.1.

5.2.2.1 Shannon’s Index Extended to Diameter, Height, and Species

The steps in computing the post-hoc Extended Shannon Index with the MKRF data were modified from Section 5.1.2.1:

1. Tree data were classed in ten-centimetre diameter classes, using the diameter outside bark at breast height (1.3 m above ground; dbh): 0 to 10 cm through 140 to 150 cm,
and 150+ cm. Then, Shannon's Index was computed with $p_i$, the proportion of basal area in the $i$th diameter class, and equation [2.1] to yield $H'_d$.

2. Tree data were classed into five-metre height classes: 0 to 5 m through 55 to 60 m, and 60+ m. Then, Shannon's Index was computed with $p_j$, the proportion of basal area in the $j$th height class, and equation [2.1] to yield $H'_h$.

3. The data were combined by species, and Shannon's Index was computed with $p_k$, the proportion of basal area in the $k$th species, and equation [2.1] to yield $H'_s$.

4. The final index was computed as the average of the diameter, height, and species indices and labelled $H'_{d+h+s}$.

The procedure for computing the Extended Shannon index under the combination method was also modified to include species. The data were classed simultaneously by species, diameter, and height, using the same classes as with the post-hoc method. Shannon's Index was then computed with $p_{ijk}$, the proportion of basal area in the $i$th diameter, $j$th height, and $k$th species class, and summed over all classes to yield $H'_{dhs}$.

The results of computing both types of Extended Shannon index follow in Table 5.13, along with Shannon's Index computed individually for the diameter, height, and species distributions. As was the case with the simulated data, the indices associated with the combination index tended to be larger than that of the post-hoc index. This was not unexpected, as the simultaneous relegation of data to species, dbh, and height classes necessarily produced more combinations, and hence produced a larger index.
Table 5.13. Extended Shannon indices computed with MKRF data.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>Plot #</th>
<th>$H'_d$</th>
<th>$H'_h$</th>
<th>$H'_s$</th>
<th>$H'_{d+h+s}$</th>
<th>$H'_{d+h+s}$</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>57</td>
<td>2.186</td>
<td>2.091</td>
<td>0.979</td>
<td>1.752</td>
<td>2.860</td>
<td>1 3</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>09</td>
<td>2.173</td>
<td>1.935</td>
<td>1.067</td>
<td>1.725</td>
<td>3.103</td>
<td>2 1</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>46</td>
<td>1.748</td>
<td>1.915</td>
<td>0.997</td>
<td>1.553</td>
<td>2.781</td>
<td>4 6</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>58</td>
<td>1.732</td>
<td>1.736</td>
<td>0.791</td>
<td>1.420</td>
<td>2.651</td>
<td>5 7</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>49</td>
<td>1.859</td>
<td>1.480</td>
<td>0.876</td>
<td>1.405</td>
<td>2.811</td>
<td>6 5</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>52</td>
<td>0.958</td>
<td>1.117</td>
<td>1.659</td>
<td>1.245</td>
<td>2.399</td>
<td>8 8</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>15</td>
<td>1.616</td>
<td>1.562</td>
<td>1.011</td>
<td>1.396</td>
<td>2.974</td>
<td>7 2</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>77</td>
<td>0.995</td>
<td>1.149</td>
<td>0.981</td>
<td>1.042</td>
<td>2.062</td>
<td>9 9</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>45</td>
<td>2.141</td>
<td>1.851</td>
<td>1.098</td>
<td>1.697</td>
<td>2.827</td>
<td>3 4</td>
</tr>
</tbody>
</table>

In general, both types of index ranked the plots in the same manner. Exceptions occurred for the medium-diversity small tree plot, the low-diversity large tree plot, and the medium-diversity normal plot. Although the plot chosen to represent high diversity was not the highest ranked under the combination method, the index values for this plot type were close to that of the highest ranked plot (2.86 versus 3.10). In general, the indices ranked the high and medium-diversity stands above the low-diversity stands, with one exception: under the combination method, the stand with low-diversity and mid-sized trees (Low-Mid) was ranked as more diverse than the stand with medium-diversity and small trees.

The low-diversity normal (Low-Norm) and the low-diversity small tree (Low-Sml) stands always ranked last; however, this was no doubt due to their very limited ranges in comparison to the other low-diversity stands. Although the dbh and height distributions were similar for the two stands, Low-Sml ranked higher due to its higher species diversity. All of the selected stands except Low-Sml have only three species and $H'_s$ ranged from 0.791 to
1.011. Low-Sml, however, has six species and $H'_s$ was 1.659. Species diversity doubtlessly played a larger role in the final value of the index for this stand than in the other selected stands.

The computation of the Extended Shannon indices with real data brought to light some compelling practical problems. Whereas dbh and height for the high-diversity stand were well distributed over most of the possible range, other stands (e.g., Med-Lrg and Med-Bim) had taller and larger dbh trees. Thus, there were more classes represented in some of the less equally distributed stands, and this may have had a compensating effect. Had the classes been chosen such that there was one class for dbh greater than 120 cm, indices for the plots labelled as Med-Lrg and Med-Bim would have been lower.

As was the case for the simulated data, the post-hoc method produced more logical rankings of plots. For the simulated data, this may have been a function of how the random plots were generated. For the MKRF data, this may be a result of how the plots were chosen. Both the simulated plot generation and the MKRF plot selection were performed by dbh and height dimension separately, and the post-hoc method relies on the dbh and height dimensions separately. The combination method depends on the three dimensional dbh-height-species distribution, and this multi-dimensional distribution was not controlled in the simulations or selected for in the MKRF data.
5.2.2.2 Structural Index based on Variance Comparisons (STVI)

The first step in computing both the univariate and bivariate STVI was to select the maximum and minimum dbh and height values. One option was to use the actual maxima and minima present in the MKRF. However, in the forests of Coastal British Columbia, the maximum dbh may be 2 metres or more. The practical use question is whether to use the maxima “commonly found” or to use the actual maxima found in the sample data. When computing the STRI, the actual minima and maxima must be used because failure to do so could produce unreasonable indices. For example, if a sample had dbh and height values present over a larger range than had been selected for the theoretical range, the computed STRI could be negative, even if the slope of the empirical data matched that of the theoretical uniform distribution. On the other hand, when computing the STVI, negative values could never result. However, if a sample was perfectly uniform with values outside of the range chosen for computing the STVI, the resulting index would be less than one, indicating a less diverse stand. Considering these arguments, the maxima and minima were chosen such that the trees that fell outside of the range were rare: dbh in the range 2 to 150 cm, height in the range 1.3 to 60 m. Thus, the theoretical variances for univariate uniform distributions of dbh, height, and the joint distribution of (dbh, height) used in the calculation of the STVI were:

\[
S_{dbh_u}^2 = (150.0 - 2.0)^2 / 12 = 1825 \quad [5.72]
\]

\[
S_{height_u}^2 = (60.0 - 1.3)^2 / 12 = 287 \quad [5.73]
\]

\[
S_U^2 = (150.0 - 2.0)^2 \times (60.0 - 1.3)^2 / 144 = 524,128 \quad [5.74]
\]

The maximum variances used in the STVI calculations were:
Results of calculating indices with the STVI follow in Table 5.14. As was the case in analyzing the simulated data, the bivariate STVI values were smaller than that of the univariate. The maximum value of STVI is \( S \). Hence, the maximum value for all the selected plots except Low-Sml is 3.00; the maximum possible value for Low-Sml is 6.00.

Table 5.14. Univariate (STVIdbh, STVI\( h \), and STVI\( d+h \)) and bivariate (STVIdh) STVI using MKRF data.

<table>
<thead>
<tr>
<th>Plot Type</th>
<th>MKRF Plot #</th>
<th>STVIdbh</th>
<th>STVI( h )</th>
<th>STVI( d+h )</th>
<th>STVIdh</th>
<th>STVI( d+h )</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>57</td>
<td>2.061</td>
<td>1.126</td>
<td>1.594</td>
<td>0.085</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Med-Lrg</td>
<td>09</td>
<td>1.090</td>
<td>1.984</td>
<td>1.537</td>
<td>0.498</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Med-Sml</td>
<td>46</td>
<td>0.733</td>
<td>0.391</td>
<td>0.562</td>
<td>0.009</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Low-Lrg</td>
<td>58</td>
<td>0.582</td>
<td>0.678</td>
<td>0.630</td>
<td>0.031</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Low-Mid</td>
<td>49</td>
<td>0.890</td>
<td>0.551</td>
<td>0.721</td>
<td>0.042</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Low-Sml</td>
<td>52</td>
<td>0.321</td>
<td>0.295</td>
<td>0.308</td>
<td>0.000</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Med-Norm</td>
<td>15</td>
<td>1.028</td>
<td>0.697</td>
<td>0.863</td>
<td>0.013</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Low-Norm</td>
<td>77</td>
<td>0.482</td>
<td>0.172</td>
<td>0.327</td>
<td>0.001</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Med-Bim</td>
<td>45</td>
<td>1.298</td>
<td>1.853</td>
<td>1.575</td>
<td>0.215</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The highest values of the univariate STVI occurred for the very diverse case (High), followed by the medium-diverse bimodal and medium-diverse large tree stands (Med-Bim and Med-Lrg). Under the bivariate method, however, the highest values occurred for the medium-diverse large tree stand (Med-Lrg), followed by the medium-diverse bimodal
stand (Med-Bim), and then the very diverse case (High). Both methods ranked the low-
diversity large tree and mid-sized tree stands (Low-Lrg and Low-Mid) higher than that of the
medium-diverse small tree stand (Med-Sml). Under the bivariate method, the medium-
diverse normal tree stand ranked lower than that of Low-Lrg and Low-Mid. Higher values of
STVI were roughly correlated with more widely distributed stands.

The univariate STVI may possess the desirable characteristics listed under item 8
(Section 4.1). With the exception of the medium-diverse small tree stand, the index appeared
tree size insensitive, was correlated to continuously occupied ranges, and gave bimodal
stands logical index values. The bivariate index does not possess these characteristics, in that
it did not correlate well with continuously occupied ranges and made illogical rankings with
respect to many of the stands. However, this may be a function of the methods used to
choose the representative stands. Stands were chosen by looking at the univariate
distributions of dbh and height, and were thus well described by the univariate STVI. Stands
were not, however, chosen by looking at the bivariate distributions of dbh and height by
species. One consequence of this is that the sample variance of the stands never reached the
value of the theoretical uniform variance (Table 5.15).
When dbh and height are considered univariately for all species combined, the high-diversity (High) and medium-diverse large (Med-Lrg) stands had variances near that of the theoretical uniform distribution, and the medium-diverse bimodal (Med-Bim) stand had a variance greater than that of the uniform. However, when dbh and height were considered univariately by species, all stands except Med-Lrg had average variances much less than that of the uniform. The influence of the distribution of dbh and height by species is apparent, in that the stand with the highest variance by species (Med-Lrg) has a high value for STVI<sub>d+h</sub>, and by far the highest value of STVI<sub>dh</sub>.

When dbh and height were considered jointly (bivariately), all stands had variances much less than that of the uniform. The Med-Lrg stand had the highest average generalized variance by species of all the selected stands, but this variance was not high enough to use the right-hand side of the STVI curve. Nonetheless, the STVI for Med-Lrg was by far the

87
highest of the selected stands. The univariate STVI may be preferred over the bivariate STVI, in that it may better meet the criteria presented in Section 4.1, and provides a more logical ranking of the stands.

5.2.3 Discussion of Extended Shannon Index and Univariate STVI Computed with MKRF Data

The Extended Shannon indices and the univariate STVI gave adequate results with respect to the criteria presented in Section 4.1. The indices were fairly tree size insensitive, gave logical values for bimodal stands, and were roughly correlated to the range of tree sizes. Although some illogical rankings occurred, these rankings may have resulted from the way plots were selected.

Practical problems may occur with either type of index. The computation of the Extended Shannon indices requires the selection of dbh and height class boundaries. The use of classes leads to loss of information about the distribution of values within classes, and the method of class selection may have a large impact on the final values of the indices. The computation of STVI requires the selection of maximum and minimum values of dbh and height; however, this selection may be done by choosing ranges for dbh and height such that only rare values occur outside the range.

While the STVI may be favoured in that there is no necessity to choose classes, an index based on Shannon's Index may be preferable because it has a solid basis in theory and a history of use in biology and forestry.
Chapter 6

Conclusion

Tree diversity plays an important role in forested ecosystems (Lähde et al. 1999). Previous research has shown that the diversity of tree species and size is correlated with the presence of avian and insect species (e.g., see Franzreb 1978, MacArthur and MacArthur 1961, and Willson 1974), which are in turn correlated with a variety of other indicators of overall ecosystem function (Recher et al. 1996). Stand structure can also impact stand growth and yield, as structure is an important factor in predicting the growth of trees (Oliver and Larson 1996).

Previous attempts have been made to incorporate tree size diversity into existing diversity indices; however, no commonly used index is desirable from both a practical and biological standpoint. The objective of this thesis was to develop a quantitative measure of diversity that included both the diversity of tree species and tree size. The index should measure richness and evenness using basal area and commonly measured tree attributes. The index should not rely on combining data into classes, and should account for vertical and horizontal diversity equally. The size of samples should not affect the index, and the index should be insensitive to tree size and correlated to the range of diameters and heights.

Four alternative measures were proposed. The Extended Shannon indices were derived based on Shannon’s Index, measuring the diversity of tree heights, tree
diameters, and tree species simultaneously. Curve fitting techniques were used to explore the use of the regression coefficients from fitted diameter and height distributions based on the Weibull function as indices of diversity. The STRI was derived to measure the diversity of diameter and height distributions by measuring the fit of the distribution to that of a uniform distribution. The variance of the distribution of diameters and heights was used to measure tree size diversity in the STVI.

In order to evaluate the proposed measures, a baseline of maximum diversity was chosen. This baseline was a uniform distribution of dbh and height with respect to basal area. Many patterns have been assumed to describe the distribution of tree sizes and ages. All-aged multi-cohort stands can be described with a reverse-J shape distribution with respect to stems per hectare (Oliver and Larson 1996), which would correspond to a uniform distribution of basal area per hectare. However, several studies have shown that while the stand in total may be reverse-J, the individual species may be distributed differently, e.g., normally (Oliver and Larson 1996). Therefore, the distribution of maximum diversity, may be a much more complex mixture of several distributions. However, allowing for this complexity would require a much more complex analysis, and would result in indices with difficult interpretations. In order to create an index which may be useful to forest managers, a simple approach seemed more appropriate.
All four proposed measures are based on the distribution of basal area per hectare of a stand. Although the use of stems per hectare is common in diversity indices, basal area was chosen, as it has been found to be a better representation of stand occupation, and may provide a more accurate assessment of diversity than stems (Solomon and Gove 1999). If the analysis were repeated with distributions based on stems per hectare, the indices would be weighted toward smaller trees, which represent a greater proportion of the stems per hectare in a plot.

The indices were evaluated with simulated data and inventory data. Simulated data were created to represent, as closely as possible, several different levels of structural diversity. To keep this phase of the analysis simple, species distributions were not simulated. Because simulated tree sizes were limited to biologically possible values, the baseline maximum case, with dbh and height distributed as a bivariate uniform, could not be simulated. Resulting simulated distributions were controlled in a univariate sense, but not in a bivariate sense. Nine plots were selected from the 82 MKRF inventory plots to test the indices in a less controlled sense. Because of the volume of data, these plots were chosen to match the simulated cases with respect to the univariate diameter and height distributions of all species combined. Thus, the selected plots may have had radically different bivariate distributions by species. The selection criteria for the MKRF plots and the simulation definitions may have impacted the results of testing the indices.
The four proposed measures were initially evaluated with the simulated data. The curve fitting technique did not yield an adequate index of diversity, perhaps because of the choice of the Weibull function as the base. Further analysis with a more flexible function, such as the Beta may yield better results. Although the concept of deviation from the uniform distribution provided a good starting point for a measure of diversity, the STRI proved to be too sensitive to tree size. A more appropriate avenue for investigation may come from the area of goodness of fit testing, whereby distributions of dbh and height (univariate or bivariate) may be evaluated against that of the uniform. Again, further analysis may yield an index based on this concept. The remaining two measures gave better results, and were investigated with both the simulated data and the MKRF data.

The Extended Shannon Indices were easily applied and adequately described the simulated stands and the MKRF data both in its univariate and bivariate forms. However, their computation requires that data be combined into classes. Classes may be chosen arbitrarily, and changing class boundaries or range sizes will invariably change the resulting indices. Application of the Extended Shannon Indices to data in different geographical regions may necessitate re-defining classes, making comparisons inappropriate. The idea of applying Shannon’s index to a continuous distribution via an integral form may hold promise; however, one must ensure that any functional description of the distribution of dbh and height provides an adequate fit to the data. This would certainly make the application of the index more complicated, but would provide a good alternative to arbitrary classes. Because
Shannon’s Index has been researched extensively, and the sampling properties of this index are well known, it may be worthwhile investigating this option further.

The STVI was formulated to compare the variance of a sample plot to that of a distribution representing a baseline of maximum diversity. It is a very flexible index, as it may be derived assuming almost any baseline distribution. It may also be modified to fit user defined criteria. While both forms of the STVI were easily applied and performed adequately using the simulated data, only the univariate form of the STVI was adequate with the MKRF inventory data. This may have been due at least in part to the way in which plots were selected from the MKRF, as bivariate distributions were not evaluated. However, it may be inferred from the results of the MKRF, that the bivariate standard of maximal diversity may never be met with real data. No stand in the MKRF had a generalized variance close to that of the uniform distribution, even when all species are combined. Given this practical reality, the univariate STVI may be a better measure of diversity than the bivariate STVI, unless the standard for maximum diversity is modified.

The STVI does not rely on combining data into arbitrary classes, making it an attractive option compared to the Extended Shannon Indices. However, applying the index to other geographical regions may involve changing the ranges needed as inputs for deriving the index. This may make comparisons between regions difficult. Furthermore, STVI lacks theoretical background, whereas Shannon’s Index benefits from a long history of use in forestry and ecology. Clearly, analysis of the sampling
distribution of the index and the sensitivity of the index to the input ranges is required to adequately evaluate the index.

As the concept of diversity is gaining importance in forestry, a quantitative measure of diversity is needed to provide objective criteria to evaluate changes in diversity. The alternative measures presented are only a step toward deriving an index that will adequately give a quantitative measure of stand diversity incorporating structure. However, to properly evaluate the measures presented, further analysis is necessary. Before either index can be used, more information is necessary to describe the sampling properties of the indices, and describe the relationships of these indices to other stand variables. Once these properties are known, standards should be determined to make the indices adaptable to different geographic areas and standards of maximum diversity.
References Cited


Publisher, Dordrecht, Netherlands.


Routledge, R. D. 1980. Bias in estimating the diversity of large, uncensused


Unpublished paper.


Appendix I
Plot: High

DBH Distribution

Species: All

Species: Red Cedar

Species: Douglas-fir

Species: Western Hemlock

Height Distribution

Species: All

Species: Red Cedar

Species: Douglas-fir

Species: Western Hemlock
Plot: Low-Med

DBH Distribution

Species: All

Species: Red Cedar

Species: Douglas-fir

Species: Western Hemlock

Height Distribution

Species: All

Species: Red Cedar

Species: Douglas-fir

Species: Western Hemlock
Plot: Med-Norm

**DBH Distribution**

- **Species: All**
  - Chart shows distribution of DBH classes with percent basal area per ha.

  - **Species: Red Cedar**
    - Chart shows distribution of DBH classes for Red Cedar.

  - **Species: Douglas-fir**
    - Chart shows distribution of DBH classes for Douglas-fir.

  - **Species: Western Hemlock**
    - Chart shows distribution of DBH classes for Western Hemlock.

**Height Distribution**

- **Species: All**
  - Chart shows distribution of height classes with percent basal area per ha.

  - **Species: Red Cedar**
    - Chart shows distribution of height classes for Red Cedar.

  - **Species: Douglas-fir**
    - Chart shows distribution of height classes for Douglas-fir.

  - **Species: Western Hemlock**
    - Chart shows distribution of height classes for Western Hemlock.
Plot: Low-Norm

DBH Distribution

Species: All

Height Distribution

Species: All

Species: Amabilis Fir

Species: Douglas-fir

Species: Western Hemlock

Species: Western Hemlock

Species: Douglas-fir
Plot: Med-Bim

**DBH Distribution**

Species: All

**Height Distribution**

Species: All

**DBH Distribution**

Species: Red Cedar

**Height Distribution**

Species: Red Cedar

**DBH Distribution**

Species: Douglas-fir

**Height Distribution**

Species: Douglas-fir

**DBH Distribution**

Species: Western Hemlock

**Height Distribution**

Species: Western Hemlock