Describing Tree Size Diversity

Abdul Aziz Ali

Arbetsrapport 22 1997

Master of Science Thesis in Forest Biometry
Supervisor: Hans Nyquist

SVERIGES LANTBRUKSUNIVERSITET
Institutionen för skoglig resurshushållning
och geomatik
S-901 83 UMEÅ
Tfn: 090-16 58 25   Fax: 090-14 19 15, 77 81 16

ISSN 1401-1204
ISRN SLU-SRG-AR--22--SE

ABSTRACT

Management based modifications of forest structure require a precise description of the forest state both before and after a silvicultural interference. While the average values of variables such as diameter, height and basal area provide useful descriptions of the forest state, these variables do not give information about the structural variety of the forest on a tree to tree basis.

The aim of this study is to introduce and discuss a simple index that can be used to measure the dissimilarity in sizes between trees. The index, called the tree size diversity index d_i, is based on the coefficient of variation of the diameter sizes of two neighbouring trees. For a theoretical all-sized forest, d_i has an expected value of 0.5. This value is considered to be the standard or 'yardstick' for measuring tree size diversity. d_i always ranges from 0 to 1. It is close to zero when the trees are of similar size, and tends to one as the difference between the trees gets larger.

The tree size diversity index can also be used to test hypotheses concerning size segregation, whereby large trees may tend to occur in the neighbourhood of other large trees and small trees in the neighbourhood of other small trees. This can be accomplished by carrying out randomisation tests as discussed in the study.

The tree size diversity index therefore provides a variable that can be used to describe a particular forest state, that is a practical measure of control for assessing silvicultural interference, and is a nearest neighbour statistic for testing hypotheses concerning tree size segregation.

TABLE OF CONTENTS

1 Introduction

Biodiversity is used to describe the variety and relative abundance of living organisms at the genetic, population, species and ecosystem levels [1-3]. Species diversity has been the most widely studied variety of ecological diversity mainly because species are natural units of classification. Other types of ecological diversity include niche diversity which describes the diversity of resources that an organism utilises, beta (β) diversity which is defined as the degree of change of species diversity between habitats, and habitat diversity which describes the structural complexity of the environment.

Structural complexity is of special importance to wooded lands where the vertical distribution of vegetation may determine the variety and numbers of ground vegetation as well as other life forms. It has been observed [4] that the numbers of vertical layers of vegetation in the temperate woodlands of North America provide a better prediction of bird species when used in conjunction with plant species diversity. Other studies in British woodlands [5] incorporating tree taxonomic diversity and architectural complexity have also shown that insect diversity measured as species richness is more closely related to the combination of structural complexity and plant species diversity than to plant species diversity alone. A possible explanation for these observations is that trees and other woody vegetation are examples of plants which vary greatly in their growth forms and ecological roles during different phases of their lives [6].

Because the species abundance distribution of the non-woody life forms in forested and other wooded lands reflects the ecosystem's well-being, a reliable measure of structural diversity is of as great value to ecologists who may be interested in bird species diversity, as it is to silviculturalists who are interested in measuring the structural variety that results from a selection system. Clearly, a good measure of structural diversity, which is defined here as the number of vertical layers present and the abundance of vegetation within them [6], is required in order to explain how the species numbers vary between communities such as different forest types.

This study therefore examines the measurement of structural diversity in forested areas which are classified under the terrestrial habitat system and woodland formation [7]. The aim of the study is to introduce a measure of diversity that takes into account the spatial arrangements of the trees of different sizes within forested areas.

1.1 Measuring structural diversity

Structural diversity has been measured in many ways. Foliage height diversity, for example, has been measured by visually estimating the proportion of total foliage in subjectively chosen horizontal layers [4]. Variants of this method have also been used in a number of other studies [8-10], and even a stratiscope exists to facilitate the measurement of structural diversity in this way [11]. Another way of measuring structural diversity is to divide it into two components [5]. The first component then involves the determination of plant spatial diversity by recording the number of touches by the vegetation to a vertical pole. This allows for the construction of spatial diversity profiles. The second part is to measure architectural complexity by defining a number of categories into which the plant structure at any one site can be divided. One disadvantage in the use of these methods is that the quantification of diversity is done with the help of indices that were developed in connection with species diversity. While species are easily countable, the amount of foliage at a given height, for example, is a continuous variable. The indices comprise parametric abundance models and a number of non-parametric indices such as Margalef's index [12], the Q-statistic [13], the Shannon index [14], Simpson's index [15], McIntosh's index of diversity [16], the Berger-Parker index *d* [17] and the Brillion index *HB* [14] among others. A brief examination of the parametric models follows.

1.2 Abundance models

Species abundance models describe the distribution of species abundances. These models range from those which characterise high eveness to those which characterise cases where abundance is very unequal [6]. The diversity of a community is therefore described by a model which provides the closest fit to the observed distribution of species abundances. A single index of diversity is then found by estimating an appropriate parameter using data from random samples.

A good diversity index has to take into account two factors: Species richness i.e., the number of species or distinct classes of any other characteristic of interest, and species equitability - a measure of eveness. Usually high eveness coupled with high richness indicates high diversity, but this is rarely the case in reality.

Generally, inference is made about the population of elements (usually species) where each element belongs to only one of the classes $C_1, C_2, \ldots, C_S$ and s may be large. If p_i is considered to be the proportion of elements belonging to class C_i, $i = 1, 2, \ldots, s$, then the proportion p_i is called the relative abundance of class C_i. Species abundance models strive to describe the vector $\boldsymbol{p} = (p_1, p_2, \ldots, p_s)$ or some function of $\boldsymbol{p}$ by making inferences based on the model parameters after statistical treatment of sample data.

Species abundance can be described by a number of distributions [14]. However, only four models have provided the most frequently used measures of diversity [6]. These models are now described.

1.2.1 The geometric series

The geometric series, which is also called the niche pre-emption hypothesis, represents a situation where a proportion of some limiting resource is partitioned in such a way that the most dominant species utilises a given percentage, the second most dominant species utilises the same percentage of what is left of the resource, and so on until all the species have been accounted for. Clearly, data which fit the geometric series type of species abundance suggest that the areas under study are environments which suffer from poor species representation, or areas that are in very early stages of a succession [18-20].

The geometric series gives the abundance of species as ranked from the least abundant as, [21]

$$n_i = NC_k k(1-k)^{i-1}$$

where n_i is the number of individuals of species i, N is the total number of individuals, $C_k = [1-(1-k)^S]^{-1}$ is a constant which ensures that $\sum n_i = N$, S is the total number of species and k is the proportion of the available resource that each species utilises.

1.2.2 The log series

The log series is related to the geometric series [21,22]. Both the geometric and log series have been found to adequately describe the same abundance distributions [23] although different biological explanations exist in relation to the use of either series [6]. The log series is associated with situations where one or a few factors is responsible for the distribution of species. A good example is a coniferous forest plantation where light is largely responsible for the distribution of the ground vegetation and dependent species [6]. Here the trees are the most dominant species. The log series is of the form [24]

$$\alpha x,\ \alpha x^2/2,\ \alpha x^3/3, \ldots,\ \alpha x^n/n.$$

The series has two parameters, α, the log series index which is a measure of diversity and N, the total number of individuals. These parameters are related as $N = \alpha \ln(1+N/\alpha)$.

The factor x is related to α and N by

$$\alpha = N(1-x)/x$$

and x is related to the number of species as

$$S = \alpha[-\ln(1-x)].$$

1.2.3 The log normal model

The log normal distribution of species abundance results from a common niche space that is occupied by several species and each one of the species utilises a portion of the niche space that is proportional to its relative abundance. Unlike the geometric distribution, a condition here is that the size of any one of the niche sub-space is independent of the sizes of the sub-spaces that are occupied by the other species [6].

Also, the arrival of new individuals, if considered random, will result in the subdivision of any one of the niche sub-spaces with equal probability. If the smallest sub-space is repeatedly subdivided for some reason, a log series type of species abundance distribution will result. When the opposite is true i.e., the largest sub-space is repeatedly subdivided, very high diversity will result.

When the number of individuals is plotted against the species abundance (number of species) using the logarithmic scale for the latter, a symmetric bell-shaped curve is achieved for data that fit the log normal model. The resulting classes represent doubling in species numbers if $\log_2$ is used, for example , so that the modal class can easily be read off the diagram. The modal class corresponds to the class which has most species.

If the classes to either side of the modal class are numbered as C_j, $j = 1,2, \ldots s$ and the modal class which is the reference point is denoted by C_0, the number of species in class number j can be predicted form the following equation

$$S(C_j) = S(C_0)\exp(-2\sigma^2 j^2)$$

where $S(C_j)$ is the number of species in the j^{th} class to the right or left of the modal class, $S(C_0)$ is the number of species in the modal class and σ^2 is the standard deviation of the distribution.

Because the log normal distribution for species abundance is usually truncated, (as a result of the difficulty in sampling all the rare species) S^*, the total number of species has to be estimated. The diversity index for the log normal distribution, λ, then becomes,

$$\lambda = S^*/\ \sigma.$$

1.2.4 The broken stick model

The broken stick distribution is found in situations where only one important resource is being shared almost uniformly between species [25]. A good example of this is a community consisting of taxonomically related organisms. The model is characterised by only one parameter, S, the number of species which is also used as the index of diversity. Although S is sufficient as an index of diversity since the broken stick distribution depicts a highly even representation of species, it is also dependent on the sample size [26]. Close attention should therefore be paid to the sample size when data seems to fit the broken stick model of species abundance distribution.

The model has the following form

$$S(n) = [S(S-1)/N](1-n/N)^{S-2}$$

where $S(n)$ is the number of species in the abundance class with n individuals.
It may be preferable to replace the species with other variables such as an indicator of plant size. A good example is tree crown length which allows for the estimation of foliage height diversity. Such a measure of diversity would however lead to some loss of information because most of the objectively measurable indicators of plant size are continuously scaled.

Further, all the above models make important assumptions regarding the underlying probability distributions. A quick examination of the log series reveals that the absolute abundances (the number of elements of C_i) are considered to be s independent observations from a gamma distribution with given parameters [22].
If for any reason the elements do not belong to one and only one class, different authors will use different abundances and consequently obtain different diversity measures for the same data.

Another disadvantage which is also shared by the non-parametric indices is that diversity is measured only in terms of the total number of classes present. Some of the indices include a measure of eveness as well. If sampling is done in a large area, diversity can be very high in small pockets thus giving an unbalanced representation of diversity on a unit area basis. Besides, one who is interested in estimating foliage height diversity or tree size diversity, for example, can easily determine the number of trees or plants in their respectively designated classes as well as the number of classes. Unless sampling is carried out in very small areas, information on the inter-mixing of trees belonging to the various classes will be lost in the process of estimating diversity.

2 Diversity in Forests

A great number of endangered animal, plant and insect species inhabit forested lands [27]. The diversity of these organisms depends to a certain extent on the proportional distribution of the tree species and the age-class distribution of the forest stand. Diversity in forests can therefore be examined at two levels. Firstly there is the species diversity which in turn depends on the age-class or tree size distribution, and secondly there is the tree size distribution itself [27].

These levels of diversity are best exemplified by natural forests whose development leads to a combination of species mixture and tree size diversity on a given unit area of forest land [28,29,30] . Forests of this type are characterised by an abundance of small trees whose numbers tend to gradually diminish as they get older [30,31,32]. Natural and other forests which are managed to resemble them are described as all-aged forests [33] as opposed to even-aged forests which comprise of trees of approximately uniform age.

Even-aged stands are groups of trees which originated in a relatively short period of time [34]. These trees belong to the same age class. The width of the age class may vary from one or two years as in plantations to as many as 20 years [34]. Uneven-aged stands on the other hand have trees which do not display a single reproductive period. The forest originates more or less continuously and if undisturbed the stands usually have trees varying from germinating seedlings to overmature individuals. More shade tolerant species tend to form uneven-aged stands in the absence of disturbance. Categorising stand structure by age is however not very useful in many practical instances mainly due to the reason that age may be impossible to determine in many natural stands. In cases where it can be determined, age data is usually imprecise and time consuming to collect. A better description of forest stands would therefore be all-sized and even-sized stands. This description results from tree size being very well correlated to its age except for suppressed trees and on rare occasions where stands display slow growth on poor sites and trees of widely diverse ages may show little variation in size [34].

An easily measurable variable representing tree size is diameter at breast height (DBH). Diameter at breast height is very well correlated to both crown diameter and tree volume [35]. DBH will, for this reason, be used throughout the rest of this paper as a measure of tree size.

2.1 Diameter distributions

2.1.1 All-sized stands

A possible model for the diameter distribution of an all-sized stand is the exponential distribution which is characterised by a reduction in the number of trees as the DBH increases. In many managed all-sized stands, a balanced all-sized forest results. This means that the current growth can be removed periodically while maintaining the exponential diameter distribution which ensures the all-sized stand structure.

The shape of the exponential distribution is shown in the Figure 1.

The density function for the exponential distribution is,

$$f(x) = (1/\lambda)e^{-x/\lambda} \qquad \lambda > 0,\ \ 0 \le x < \infty.$$

where x = DBH.

The parameter λ is the mean diameter as well as the standard deviation of the population. Low values for λ reflect a high stocking of seedlings or trees of the smallest diameter class. The value of λ is determined by the species present, site and the silvicultural treatment of the stand [36].

If the trees in an all-sized stand are put into diameter classes with a subjectively chosen width denoted by a, the number of trees in successive diameter classes follow the geometric series of the form α, αy, αy^2, αy^3, . . . , where y is the ratio of the series [34].

α, the proportion of trees in the smallest diameter class, is given by $\int_0^a f(x)dx$. The next largest diameter class has the proportion $\int_a^{2a} f(x)dx$ and in general, the proportion of trees in the n^{th} largest diameter class is given by $\int_{na}^{a(n+1)} f(x)dx = e^{-na/\lambda}(1-e^{-a/\lambda})$.

The ratio of the series is therefore given by $e^{-(n+1)a/\lambda}(1-e^{-a/\lambda})/\, e^{-na/\lambda}(1-e^{-a/\lambda})$ which simplifies to $e^{-a/\lambda}$.

This ratio is a measure of the rate at which the number of stems diminishes as one moves from one diameter class to the next larger diameter class and it is usually referred to as the ‘diminution quotient’.

The ratio is also positively related to $1/\lambda$, the inverse of the mean diameter, indicating that mortality is higher in stands with a high density of seedlings or small trees [37]. The exponential distribution therefore provides a model for the distribution of diameters which can be compared to the geometric species abundance distribution model.

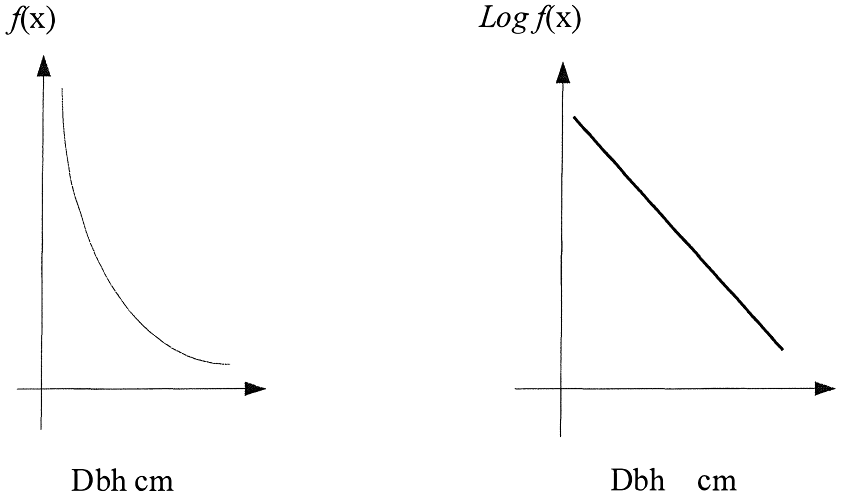


Figure 1. The density function of the exponential distribution and its logarithmic transformation.

2.1.2 Even-sized stands

Even-sized stands (which can also be considered as all-sized depending on the objective of the study [32]) have diameter distributions which depend to a certain degree on the stand age. Very young plantations have their diameter distributions (diameter at ground level, DGL) taking the shape of a symmetrical bell shaped curve with most trees clustering near the average diameter. As the stand grows older, the

diameter class distribution changes. The total number of trees in the stand decreases with an increase in the variation of the diameters as well as the number of different diameter classes. Some even-sized stands may however have their diameters resembling the symmetrical bell-shaped curve for a considerably long time [38].
The diameter distribution of an even-sized stand is also influenced by site, the species composition of the stand, and disturbances resulting from heterogeneous spatial patterns of survival following regeneration and early growth, or silvicultural treatment.

2.2 Choosing a probability distribution to model diameters.

Because a number of factors influence the diameter distribution of even-sized stands, several shapes are taken by their frequency distributions. Most of the distributions are unimodal and positively skewed. Although it is not uncommon to come across diameter distributions which follow a bimodal form as a result of physical disturbances to an all-sized stand.

As a result of this, a number of statistical distributions lend themselves for fitting diameter data, depending on the skewness and kurtosis of the diameter frequency diagram. Some of the distributions include the Gram-Charlier series [39], the Pearl-Reed curve [40,41], Pearsonian curves [42], the gamma distribution [40], and the three parameter log-normal distribution [43]. The beta distribution has also been applied to diameter distributions [44].

The Weibull distribution has been successfully used to fit diameter data [45,46,47]. This distribution, unlike the gamma and log normal distributions which are limited to positively skewed shapes, has been used to model diameter distributions because it can take on shapes depicting the full range of unimodal continuos distributions with both positive and negative skewness as well as the shapes described by all-sized stands [44].

Since diameter distributions show some variation in skewness and peakedness or kurtosis, it has been suggested that a probability distribution can be selected according to the observed skewness and kurtosis from random samples [48].

Skewness is measured by $\sqrt{\beta_1}$, the skewness coefficient, and it is given by $\sqrt{\beta_1}$ = $\mu_3/\mu_2^{3/2}$. Kurtosis is measured by $\beta_2 = \mu_4/\mu_2^2$ where μ_k is the k^{th} central moment.

$$\mu_k = \int_{-\infty}^{\infty} (x - E[X])^k f(x)dx$$

and $f(x)$ is the probability density function of the random variable X.

$\sqrt{\beta_1}$ and β_2 may be estimated from sample data by the method of moments. Their moment estimators are given by, [48].

$$\sqrt{b_1} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{[\sum_{i=1}^{n}(X_i - \overline{X})^2]^{3/2}} \quad \text{and} \quad b_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{[\sum_{i=1}^{n}(X_i - \overline{X})^2]^2}$$

The asymptotic variances are $\sqrt{6/n}$ and 24/n for $\sqrt{\beta_1}$ and b_2 respectively [49], when the parent population is normal.

The β_1 and β_2 space can be used to demonstrate the range of skewness and peakedness covered by the various statistical distributions [50]. Figure 2 shows the β_1-β_2 graph and the statistical distributions that have been used for describing diameter distributions.

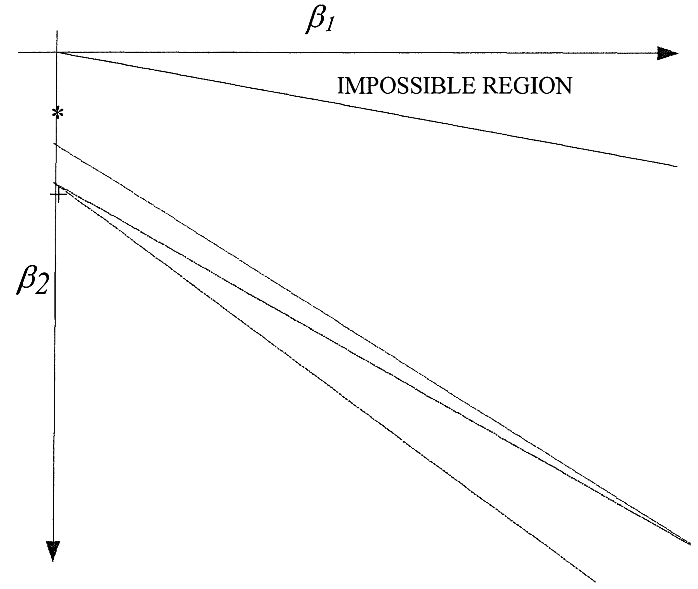


Figure 2. The β_1-β_2 space and selected distributions which are used to fit diameter data.

1. _______________ WEIBULL LINE.
2. _______________ GAMMA LINE.
3. _______________ LOG-NORMAL LINE.
4. (The meeting piont of 1 and 2) EXPONENTIAL DISTRIBUTION.
5. * UNIFORM.
6. + NORMAL.

The 'impossible region' refers to the combinations of β_1 and β_2 which are mathematically impossible. The normal, exponential and uniform distributions have only one shape and are therefore represented on the plane by single points. The Weibull, gamma and log-normal show more flexibility in terms of their ability to take on a variety of shapes. Their three lines lie rather close to one another which explains why the three distributions may perform equally well when fitting diameter data. It can also be seen that the exponential distribution is a special case of both the gamma and the Weibull distributions.

The beta distribution covers the entire region between the gamma, the β_2 axis and the 'impossible region' which shows that the beta distribution is much more flexible than the others in terms of skewness and kurtosis. This distribution, however, has its range of positive density from 0 to 1, making it necessary to identify upper and lower limits of any data set to which the distribution is to be fitted so that the appropriate transformation of scale can be made.

Another distribution which is even more flexible than the beta is a four parameter distribution called the Johnson's S_B [48]. This distribution covers the β_1-β_2 space between the log-normal and the gamma lines in addition to the area spanned by the beta distribution. Unlike the beta distribution, its maximum likelihood estimators have closed form solutions.

Johnson's S_B is a transformation on the normal and it is relatively easy to fit for both diameter and height data [48].

The probability distribution which is ultimately chosen to describe a diameter distribution will depend on the likelihood or the goodness-of-fit results. The distribution must also be relatively simple to fit in terms of parameter estimation, sufficiently flexible and be easily manipulated for estimation of proportions in the various size classes.

It should also be noted that skewness and kurtosis by themselves do not uniquely determine a distribution. These parameters only help in identifying candidate distributions.

The models should be examined in terms of their biological sense. For example, does the model make any specific assumptions about the biological or ecological processes needed to generate a specific pattern of diameter distribution? Or is the model solely statistical in the sense that it represents nothing more than a mathematical fit to the empirical data?

3 Describing tree size diversity

Tree size diversity can be described by examining the distribution of any of the variables that are indicators of tree size e.g., tree height, diameter at breast height (DBH), crown diameter, or tree volume. These, on their own, give only a general idea of the tree size diversity because irregularities shown in small areas of a stand may tend to even out when larger areas are considered for measurement. For example, combining about equal areas of all the size classes for a fully regulated even-sized forest may produce a distribution typical of that of an uneven-sized forest. The actual size distribution would reveal that trees occur in clusters whereby most of the trees are surrounded by trees of their own size.

A better description of tree size diversity can be arrived at by considering the actual dissimilarity in the tree sizes of the individual trees in relation to their proximity to one another.

An index which can be used to describe tree size diversity, and which is based on the information obtained from the dissimilarity between two neighbouring trees is introduced here.

A randomly located tree and its nearest neighbour provide the smallest possible area over which the variation in tree sizes is measurable as well as a good idea of how this variation is spread over the stand.

3.1 The size differentiation index

One index which makes use of information based on the dissimilarity in sizes between neighbouring trees is the ‘size differentiation’ index [51]. The index is based on the ratio of the diameters of neighbouring trees. The differentiation at breast height diameter, denoted by TD_i, is defined as the ratio of the smaller diameter to the bigger diameter (r_{ij}), subtracted from unity,

$$TD_i = \frac{1}{n}\sum_{j=1}^{n}(1 - r_{ij})$$

where i = tree ($i = 1,2,...,I$)

j = nearest neighbour ($j = 1,2,...,n$)

r_{ij} = (smaller DBH)/ (bigger DBH) of the two neighbouring trees.

TD_i, which lies between zero and one may also be used with other indicators of tree size, such as height or volume.

The principle of size differentiation is explained by Figure 3. The breast height diameters of the nearest tree to the sampling point and those of its three nearest neighbours are shown.

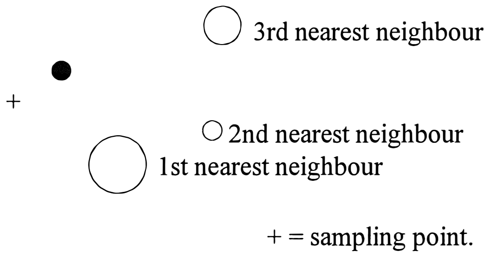


Figure 3. An illustration of how to sample for the size differentiation index.

The value of TD_i increases with increasing average diameter difference between trees. A zero value indicates that neighbouring trees are equal in size.

This index is most useful in studying the effects of thinning. It can be calculated for the stand as a whole irrespective of the tree species, or for a given sub-population. One advantage of using this type of index is that one can study both the short range differentiation ($n = 1$) and the long-range differentiation ($n = 3,4,...$) because the index allows for more than one nearest neighbour. Long-range differentiation is an average over a larger number of trees. Consequently this results in the reduction of the TD_i variation within a stand and leads to a more general description of tree size differentiation.

Unfortunately, there are no assumptions made about the diameter distributions thus ignoring the frequencies with which trees of a particular diameter appear in the population. For this reason, the size differentiation index is of greater help when studying relative changes in stand structure, such as modifications brought about by thinnings, rather than describing a particular forest state.

3.2 Indices based on the Coefficient of Variation

Another index of variation is the coefficient of variation which is a relative measure of spread. The coefficient of variation gives the spread or standard deviation of a random variable as a ratio to its mean. It is given by,

$$V = \frac{S}{\overline{X}}$$

where S is the standard deviation and $\overline{X}$ is the mean.

The coefficient of variation depends on the underlying probability distribution. Samples taken from an exponential distribution, for example, will have their coefficients of variation spread around an expected value of one. More dispersed populations have a value greater than one and values of less than one are obtained for less dispersed populations [52]. A similar index where the variance (S^2) is used in place of the standard deviation (S) can be used to test hypothesis concerning the spatial distribution of objects [53]. Here, the objects (trees etc.,) are assumed to be randomly distributed according to a Poisson process. Any deviation of the ratio $S^2/\overline{X}$ from one is therefore an indication of a non-random pattern of spatial distribution e.g., clustering.

The coefficient of variation may be used to describe tree size diversity. However, if the samples are collected from large areas, the values obtained for the coefficient of variation will be average values over the large areas. This will eventually lead to some loss of information because the coefficient of variation on its own does not make use of the information on the spatial arrangement of trees of varying sizes.

3.2.1 The dissimilarity coefficient

In order to estimate the coefficient of variation small plots are randomly distributed in a forest stand. If J plots are randomly distributed within a stand, the mean and standard deviation for the j^{th} plot are given by $\overline{x}_j$ and s_j respectively, $j = 1,2, \ldots, J$. The coefficient of variation V_j can then be calculated for every plot

$$V_j = \frac{s_j}{\overline{x}_j}$$

The measure of dissimilarity is the average of the V_j over the plots, i.e.,

$$\overline{V} = \frac{1}{J}\sum_{j=1}^{J} V_j$$

A desirable attribute of the diversity measure is that it should provide information on the dissimilarities between the individual trees. This can be achieved by comparing only two neighbouring trees at a time, thereby reducing the plot to the smallest possible size.

The two trees are selected by locating the tree which lies closest to a randomly selected sampling point and its nearest neighbour.

Suppose that the tree diameters in the j^{th} plot are denoted by x_{1j} and x_{2j} and their mean by $\overline{x_j}$, their standard deviation is given by,

$$s_j{}^2 = \frac{\sum_{i=1}^{2}\left(x_{1j} - \overline{x}_j\right)^2}{(2-1)} = \left(x_{1j} - \overline{x}_j\right)^2 + \left(x_{2j} - \overline{x}_j\right)^2$$

$$= \left\{x_{1j} - \frac{\left(x_{1j} + x_{2j}\right)}{2}\right\}^2 + \left\{x_{2j} - \frac{\left(x_{1j} + x_{2j}\right)}{2}\right\}^2$$

$$= \frac{1}{4}\left\{\left(x_{1j} - x_{2j}\right)^2 + \left(x_{1j} - x_{2j}\right)^2\right\} = \frac{1}{2}\left(x_{1j} - x_{2j}\right)^2$$

and hence, $$s_j = \frac{1}{2}\left\{\left(x_{1j} - x_{2j}\right)^2\right\}^{1/2} = \frac{\left|x_{1j} - x_{2j}\right|}{\sqrt{2}}$$

The coefficient of variation is then given as,

$$V_j = \frac{\left|x_{1j} - x_{2j}\right|\sqrt{2}}{\left(x_{1j} + x_{2j}\right)}$$

And an average for the coefficient of variation over all the sample plots is,

$$\overline{V} = \frac{\sqrt{2}}{J} \sum_{j=1}^{J} \frac{|x_{1j} - x_{2j}|}{(x_{1j} + x_{2j})}$$

The dissimilarity coefficient [52] is obtained from the coefficient of variation but, for simplicity reasons $\sqrt{2}$ is replaced by 1. If Hj denotes this measure, then,

$$Hj = \frac{|x_{1j} - x_{2j}|}{(x_{1j} + x_{2j})}$$

and the dissimilarity coefficent is accordingly defined as,

$$H = \frac{1}{J} \sum_{j=1}^{J} H_j = \frac{1}{J} \sum_{j=1}^{J} \frac{|x_{1j} - x_{2j}|}{(x_{1j} + x_{2j})}$$

The measure Hj lies within the interval 0 and 1. It is close to zero when both trees in the pair are of similar size and 1/3 when one tree is double the size of the other. Hj tends to one as the difference between the tree diameters gets larger

.

Another property of Hj, as shown in the figure below, is that when the difference between the diameters of the two trees i.e., $|x_{1j} - x_{2j}|$ is held at a constant value, Hj becomes inversely proportional to the sum of the diameters i.e., $Hj \propto 1/(x_{1j} + x_{2j})$ when $|x_{1j} - x_{2j}| = \text{constant}$.

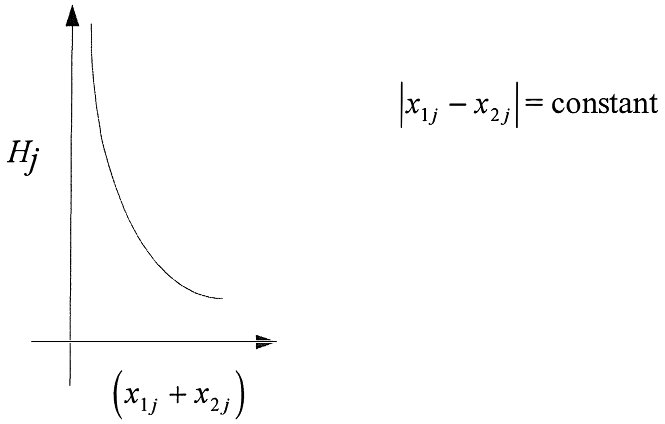


Figure 4. The relationship between Hj and the sum of the diameters

This shows that smaller values for H_j will be obtained for pairs of large trees whose diameters differ by the same margin as pairs of smaller trees. Consequently, a young even-sized stand will be considered to be more diverse than an older stand with a similar size distribution. This difference is however only marginal as illustrated by the following numerical example.

Consider any pair of neighbouring trees from a stand with $x_1 = 5$cm and $x_2 = 6$cm.
Then $H_j = |5 - 6|/(5 + 6) = 0.090$
Consider also a similar pair from an older stand with $x_1 = 24$cm and $x_2 = 25$cm.
Then $H_j = |24 - 25|/(24+25) = 0.020$

The H_j values for both the pairs in their respective stands are very low (< 0.1), showing that the difference is only marginal. This follows from the fact that the index is a relative measure.

3.3 The dissimilarity coefficient in a theoretical all-sized stand.

In natural forests or all-sized stands which are regulated to ensure an exponential diameter distribution, X_1 and X_2 are assumed to be exponentially distributed random variables. Since the exponential is a special case of the gamma distribution, the probability density function for the random variable H_j can be derived assuming that the random variables X_1 and X_2 are gamma distributed. The reason for using the gamma density function is that the distributional assumption allows for the derivation of the density function for H_j which will serve for all-sized as well as even-sized stands whose diameter distributions can be modelled using the gamma distribution. X_1 and X_2 are also assumed, for simplicity, to be independent of each other.
Given the above assumptions, the probability distribution function for H_j can be derived according to the following transformation.

$$H_j = \left|\frac{X_1 - X_2}{X_1 + X_2}\right| \text{ which can be re-written as, } \left|\frac{X_1}{X_1 + X_2} - \frac{X_2}{X_1 + X_2}\right|$$

$$\text{Let } \frac{X_1}{X_1 + X_2} = Y_1 \text{ and } \frac{X_2}{X_1 + X_2} = Y_2$$

then $H_j = | Y_1 - Y_2 |$ and the simultaneous distribution function for X_1 and X_2 is given as,

$$f_{X_1,X_2}(x_1,x_2) = \frac{x_1^{\alpha-1}x_2^{\alpha-1}e^{\frac{-x_1}{\beta}-\frac{-x_2}{\beta}}}{\Gamma(\alpha)\Gamma(\alpha)\beta^{\alpha}\beta^{\alpha}} \qquad 0<x_1<\infty,\ 0<x_2<\infty.$$

$$\alpha, \beta > 0.$$

If $Y_3 = (X_1+X_2)$, then $X_1 = Y_1Y_3$, and $X_2 = Y_3(1-Y_3)$ and the simultaneous distribution function for the random variables Y_1 and Y_3 are given below.

$$f_{Y_1,Y_3}(y_1,y_3) = \frac{y_1^{\alpha-1}(1-y_1)^{\alpha-1}y_3^{2\alpha-1}e^{\frac{-y_3}{\beta}}}{\Gamma(\alpha)\Gamma(\alpha)\beta^{\alpha}\beta^{\alpha}} \qquad 0<y_1<\infty,\ 0<y_3<\infty \quad \alpha,\beta > 0,$$

where the Jacobian of the transformation, which is the determinant of the matrix of partial derivatives, is given by $J = |y_3(1-y_1) + y_1y_3| = y_3$.

The marginal distribution for Y_1 is then obtained by integrating the simultaneous distribution function with respect to Y_3 as follows,

$$f_{Y_1}(y_1) = \frac{\Gamma(2\alpha)\beta^{2\alpha}}{\Gamma(\alpha)\Gamma(\alpha)\beta^{\alpha}\beta^{\alpha}}y_1^{\alpha-1}(1-y_1)^{\alpha-1}\int_0^{\infty}\frac{1}{\Gamma(2\alpha)\beta^{2\alpha}}y_3^{2\alpha-1}e^{\frac{-y_3}{\beta}}dy_3,$$ where the

The kernel of the above distribution is recognised as the beta density function. The integral of this function simplifies to 1. As a result of this, the marginal distribution for the random variable Y_1 is given by,

$$f_{Y_1}(y_1) - \frac{\Gamma(2\alpha)y_1^{\alpha-1}(1-y_1)^{\alpha-1}}{\Gamma(\alpha)\Gamma(\alpha)},\quad 0 \le y_1 \le 1,\ \alpha > 0,$$ which in turn is

recognised as the beta density function with parameters (α, α).

Also, given that $H_j = |Y_1 - Y_2|$ and that $Y_1+Y_2 = 1$, then $H_j = |2Y_1 - 1|$.

Let U = $2Y_1$-1, then $H_j = |U|$ and the probability distribution function for U can be obtained from the transformation $F_U(u) =$ P($U \le u$) = P ($2Y_1$ -1 $\le u$) and

$$f_U(u) = \frac{dF_U(u)}{du} = f_{Y_1}(y_1)\left|\frac{dy_1}{du}\right|.$$

Since $u = 2y_1+1$, then $y_1 = \frac{u+1}{2}$, and therefore $\left|\frac{dy_1}{du}\right| = \frac{1}{2}$.

Hence $f_U(u) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\left(\frac{u+1}{2}\right)^{\alpha-1}\left(1-\frac{u+1}{2}\right)^{\alpha-1}\frac{1}{2}$, which simplifies to,

$f_U(u) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\frac{(1-u^2)^{\alpha-1}}{2^{2\alpha-1}}$, $-1 \le u \le 1$. Since $Hj = |U|$, the last transformation can be done in two steps.

Step 1. $U > 0$ and therefore, $Hj = U$. In this case, $\left|\frac{du}{dh_j}\right| = 1$ and

$f_{H_j}(h_j) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\frac{(1-h_j^{\,2})^{\alpha-1}}{2^{2\alpha-1}}$.

Step 2. $U < 0$, hence $Hj = -U$ and $\left|\frac{du}{dh_j}\right| = 1$ and $f_{H_j}(h_j) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\frac{(1-(-h_j)^2)^{\alpha-1}}{2^{2\alpha-1}}$,

$0 < hj < 1$.

The probability density function for Hj is then given as follows,

$f_{H_j}(h_j) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\frac{(1-h_j^{\,2})^{\alpha-1}}{2^{2(\alpha-1)}}$, $0 < hj < 1$, $\alpha > 0$.

The expected value of the random variable Hj is then obtained as,

$E\left[H_j\right] = \int_0^1 \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\frac{1}{2^{2(\alpha-1)}} h_j(1-h_j^{\,2})^{\alpha-1} dh_j$. If $x = h_j^{\,2}$, then $h_j = \sqrt{x}$ and

$dh_j = \frac{1}{2\sqrt{x}}dx$, which implies that $E\left[H_j\right] = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\frac{1}{2^{2\alpha-1}}\int_0^1 x^{1-1}(1-x)^{\alpha-1}dx$, and upon rearranging,

$$E\left[H_j\right] = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}\frac{1}{2^{2\alpha-1}}\frac{\Gamma(1)\Gamma(\alpha)}{\Gamma(\alpha+1)}\int_0^1 \frac{\Gamma(\alpha+1)}{\Gamma(1)\Gamma(\alpha)}x^{1-1}(1-x)^{\alpha-1}dx,\ 0 < x < 1.$$

The kernel of the integrand is the beta probability density function with parameters $(1,\alpha)$ and the integration simplifies to 1. The expected value of Hj, as a result, becomes $E\left[H_j\right] = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)[\Gamma(\alpha+1)]}\frac{1}{2^{2\alpha-1}}$.

For the special case of the gamma distribution with $\alpha = 1$, the exponential distribution results and as a consequence the probability density function for Hj becomes,

$f_{H_j}(h_j) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \frac{1}{2^{2(1-1)}} (1 - h_j^2)^{1-1} = 1$, $0 \le h_j \le 1$. The random variable Hj is therefore uniformly distributed on the interval (0,1), and its density function is given as,

$$f_{H_j}(h_j) = 1 \qquad 0 \le h_j \le 1.$$

and its expected value is $E[H_j] = \frac{\Gamma(2)}{\Gamma(1)\Gamma(2)} \frac{1}{2^{2-1}} = \frac{1}{2}$.

It is also seen that Hj is not dependent on λ, the mean diameter as illustrated by the following Table.

Table 1. An illustration of the results from a numerical examples.

Distribution	Mean of Hj	Std. dev of Hj
exp (1)	0.499	0.00458
exp (3)	0.510	0.00457
exp (100)	0.501	0.00452

These values were calculated from simulations of samples (4000 pairs) independently drawn from the three exponential distributions.

The expected value and variance for Hj are 1/2 and 1/12 respectively. And because H is the average of J independent $U(0,1)$ random variables, the central limit theorem can be used to show that H is asymptotically normally distributed with an expectation of 1/2 and variance $1/(12J)$ respectively.

For distributions other than the gamma, the derivation of the probability distribution function for Hj is less tractable.

When deriving the distribution of Hj for a stand whose diameter sizes follow the exponentially distribution, the size of one tree is considered to be independent of the size of its neighbour, resulting in a uniform pattern of spatial distribution of trees with respect to size. This type of forest serves as an idealised standard so that the deviations from this standard can serve as an index of dissimilarity.

Clearly, values for H that are close to 0.5 imply that the size dissimilarity coefficient of the stand under consideration is near that of the desirable standard. Values below 0.5 suggest that the stand has trees of similar sizes and values above 0.5 indicate a high dissimilarity in tree sizes but raises questions about the proportional representation of the different diameter classes. As the diameter distribution tends to uniform, a high value of H will be obtained, assuming that the trees are uniformly distributed on the ground with respect to size.

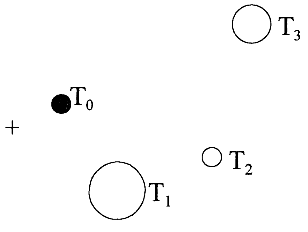
3.4 The tree size diversity index d_i

One advantage which the size differentiation index [51] has over the coefficient of dissimilarity [52] is that the former index allows one to study both the short range differentiation ($n = 1$) and the long range differentiation ($n = 2,3,4, ...$). The dissimilarity coefficient is limited to the dissimilarity between the sizes of two trees at any one time, thus declaring redundant the information contained in the 2nd, 3rd, or nth nearest neighbours. This information, although not as useful as that contained in that size of the nearest neighbour, is of great value mainly because the stand can not be reduced to particular tree pairs which are clearly defined and distinguishable from other similar pairs.

The tree size diversity index, d_i, makes use of the information contained in the sizes of any number of neighbours to a reference tree. The index is a weighted linear combination of the dissimilarity coefficients H_j. In this way one can combine the advantages of the dissimilarity coefficient with those of the size differentiation index TD_i into one new index.

A weighted linear combination makes it possible to distribute the weights according to the importance given to the information contained in the sizes of any number of neighbours to a reference tree.

The principle of the tree size diversity index is shown by the figure that follows.



+ = sampling point.

T_0 = reference tree

T_1 = 1st nearest neighbour

T_2 = 2nd nearest neighbour

T_3 = 3rd nearest neighbour

Figure 5. An illustration of the principle of tree size diversity.

H_1 is the dissimilarity coefficient based on the sizes of the tree nearest to the sampling point (the reference tree) and its nearest neighbour, i.e., $H_1 = \frac{|T_0 - T_1|}{T_0 + T_1}$ and in general,

$$H_j = \frac{|T_0 - T_j|}{T_0 + T_j}.$$

The tree size diversity index is defined as $d_i = \sum_{j=1}^{J} w_j H_j \qquad i = 1,2, \ldots, n$

$$j = 1,2, \ldots J$$

where d_i = the tree size diversity index (obtained from plot i).

H_j = the dissimilarity coefficient between the reference tree and its j^{th} neighbour.

w_j = the weight attached to each H_j.

Because H_j is a random variable, a weighted linear combination of H_j's is also a random variable and the expected value of a weighted linear combination is a weighted linear combination of expected values. As a result, the expected value of the diversity index is given by, $E[d_i] = E\left[\sum_{j=1}^{J} w_j H_j\right] = \sum_{j=1}^{J} w_j E[H_j]$.

The weights should be distributed in such a way so as to reflect the relative importance of each of the J neighbours. The distribution of the weights should also make it possible for one to obtain an unbiased estimate of the tree size diversity index in a stand. If the mean tree size diversity index on the plot is given by μ, the error made when estimating the diversity index using a weighted linear combination of sample values is given by, $e_i = \hat{\mu}_i - \mu$ where $\hat{\mu}_i = \sum_{j=1}^{J} w_j H_j$. Then, in order to arrive at an unbiased estimate, the expected value of the random variable e_i should be set to zero. One way of choosing the weights is therefore to choose only those sets of weights which give e_i an expected value of zero as shown.

Since $E[e_i] = E\left[\hat{\mu}_i - \mu\right]$, then, $E[e_i] = E\left[\hat{\mu}_i\right] - \mu$, and $E[e_i] = 0$. Hence, $E\left[\hat{\mu}_i\right] - \mu = 0$ and $E\left[\hat{\mu}_i\right] = \mu$.

therefore, $E\left[\sum_{j=1}^{J} w_j H_j\right] = E\left[H_j\right]$ and upon simplification, $\sum_{j=1}^{J} w_j E\left[H_j\right] = E\left[H_j\right]$ and $\sum_{j=1}^{J} w_j = 1$.

A condition is therefore that the sum of the weights should equal to 1. As a consequence of this, the tree size diversity index for a theoretical all-sized stand is given by, $\sum_{j=1}^{J} w_j E\left[H_j\right] = 1(0.5) = 0.5$ since $E\left[H_j\right] = 0.5$ and $\sum_{j=1}^{J} w_j = 1$.

For stands other than the theoretical all-sized stand, a procedure for selecting the weights must be established, given that the weights sum to one. One such procedure is to distribute the weights in such a way that they decrease with increasing distance from the reference tree. The weights can be chosen to be inversely proportional to the distance between the reference tree and its neighbours. However, to keep the calculations simple, the tree size diversity index will be calculated using a ranking procedure based on the distance between the reference tree and its neighbours. Trees are ranked according to their distance form the reference tree. The nearest neighbour to the reference tree is ranked 1, the 2nd nearest neighbour is ranked 2, and so on. The weights are then chosen to be proportional to the inverse of the ranks. This procedure ensures that the neighbouring trees contribute to the tree size diversity index depending on their locations. A neighbouring tree ranked j contributes relatively more to the overall diversity index than a tree ranked (j+1), for example.

If rj denotes the rank given to the j^{th} nearest neighbour, the weight assigned to the j^{th} nearest neighbour then becomes,

$$w_j = \frac{\frac{1}{r_j}}{\sum_{j=1}^{J} \frac{1}{r_j}},$$

where the term $\frac{1}{\sum_{j=1}^{J} \frac{1}{r_j}}$ ensures that $\sum_{j=1}^{J} w_j = 1$, thereby satisfying the unbiasedness condition.

The tree size diversity index is consequently defined as,

$$d_i = \frac{\sum_{j=1}^{J} \frac{1}{r_j} H_j}{\sum_{j=1}^{J} \frac{1}{r_j}},$$

where i = the sampling plot.

The tree size diversity index which has so far been described is the index value for only one sample location. The index value for the whole stand may be obtained as the average of the sample values, i.e., $d = \frac{1}{n} \sum_{i=1}^{n} d_i$.

which in turn can be considered to be a linear combination of the sample values, with a weight of $\frac{1}{n}$ attached to each sample value.

A broad range of indices can be developed by using weights which are inversely proportional to any power of the ranks, i.e.,

$$d_i = \frac{\sum_{j=1}^{J} \frac{1}{r_j^{\,p}} H_j}{\sum_{j=1}^{J} \frac{1}{r_j^{\,p}}}$$

Different choices of the exponent p will result in different indices. As p decreases, the weights tend to be more similar. When p approaches 0, the index becomes a simple average of the J dissimilarity coefficient values. For progressively larger values of p, the closest trees contribute progressively more to the index. As p approaches ∞, all the weight is given to the closest tree. The dissimilarity coefficient [52] then becomes a special case of the tree size diversity index.

The tree size diversity index shares many properties with the dissimilarity coefficient. The index ranges from 0 to 1. Values for d_i which are close to 0.5 imply that the size diversity of the stand under examination is near that of the desirable all-sized standard stand. Values below 0.5 suggest a low size diversity and values above 0.5 indicate a high size diversity.

There are however subtle differences between the two indices. The dissimilarity coefficient is an average measure of the dissimilarity between trees and their immediate neighbours. The tree size diversity index , on the other hand, is a measure of the dissimilarity between one tree per sampling plot, the reference tree, and its J neighbours. The index is also an average measure, but the contribution made by any of the neighbours to the overall index value is dependent on its location in relation to the reference tree.

4 Simulation study

The simulation study that follows was conducted for the following three reasons.

1. In order to confirm that the average value of the tree size diversity index is 0.5, and that it does not depend on the mean value of the distribution when the two diameter values are generated from an exponential distribution.

2. In order to find out the average values for d_i when the diameter values are generated from distributions other than the exponential, i.e., distributions which may be used to model diameter distributions from even-sized stands.

3. In order to study the effects of removing parts of the distributions, which correspond to certain diameter classes, on the tree size diversity index.

The Weibull distribution has been used here because it has proved useful in modelling the diameter distributions of both all-sized and even-sized stands [44]. This distribution can take on the full range of unimodal continuos distributions as well as reversed *J* shapes which can not be fitted by the exponential distribution. The reversed *J* shape of diameter distributions is a general name given to diameter distributions which resemble the exponential distribution in shape.

Random numbers were generated from 19 different distributions and used to simulate different types of forest stands. Four thousand (4000) randomly generated diameter values were used to calculate the index of tree size diversity for each of the forest types viz.: I: A forest with the reverse *J* diameter distribution. II: An even-sized forest with a diameter distribution that takes the shape of a symmetric bell-shaped curve, and III: an even-sized stand with a diameter distribution which is positively skewed.
The mean, median and standard deviation of the tree size diversity index were calculated and are presented in Table 3.

Several of the distributions are truncated at some point and -trun(x) indicates that truncation is at point x. When two diameter values are generated from the same distribution with one value less than some truncation point x and the other value greater than another point y, the notation v1< x, v2>y is used. Exp(λ) represents an exponential distribution with parameter λ and $Wei(\alpha,\beta)$ represents a Weibull distribution with parameters α and β.

Table 2. The mean, median and standard deviation of d_i for the different distributions.

Distribution	Mean of d_i	Median of d_i	Std dev. of d_i
exp(1)	0.496	0.493	0.288
exp (25)	0.492	0.483	0.289
exp(100)	0.501	0.504	0.288
exp(25)-trunc5	0.385	0.367	0.237
exp(25)-trunc10	0.323	0.300	0.207
exp(25)v1, v2<40	0.453	0.430	0.289
exp(25)v1, v2<50	0.465	0.444	0.289
exp(25)trunc10-15	0.512	0.519	0.301
exp(25)v1<15, v2>25	0.736	0.760	0.177
exp(25)v1 10-40, v2 30-50	0.322	0.327	0.186
Wei(0.5,25)	0.685	0.788	0.303
Wei(0.7,25)	0.605	0.622	0.302
Wei(2.5,25)	0.253	0.214	0.187
Wei(3.6,25)	0.185	0.152	0.145
Wei(2,5,25)-trunc5	0.234	0.202	0.169
Wei(2.5,25)-trunc10	0.003	0.002	0.004
Wei(2.5,25)v1<20, v2>35	0.269	0.239	0.188
Wei(2.5,25)v1, v2<35	0.235	0.196	0.183
Wei(2.5,25)v1,v2<30	0.228	0.182	0.183

4.1 Results of the simulation study

The results from the exponential distribution confirm that the mean of d_i is independent of the mean of the diameter distribution. When small values are removed, the mean of d_i is appreciably lowered. A truncation at the 5 cm point for a distribution with a mean of 25 gives the mean value for d_i as 0.38. When the larger values of the distribution are excluded, the mean d_i value remains close to 0.5. If the middle section of the distribution is deleted, the mean value of d_i remains close to 0.5 unless a section large enough to remove many small values is deleted. When the diameter values are selected from the two ends of the distribution, the mean value of d_i becomes close to one as expected.

Reverse *J* distributions can also be obtained via the two parameter Weibull distribution with the first shape parameter set to less than one. The mean *d* values obtained from a Weibull distribution with $0.5 \le \alpha \le 1$ are close to those obtained from an exponential distribution, thereby giving mean d_i values close to 0.5.

For a Weibull distribution with $\alpha = 3.6$, which gives a shape for the distribution that is similar to that of the a normal distribution, the mean for d_i is very low. Positively skewed Weibull distributions give low average d_i values too. Removal of the small diameter values from such a distribution results in even lower size diversity while the removal of the larger diameter values do not greatly affect the tree size diversity index, leaving it at 0.2 for a *Wei*(2.5,25), for example.

The assumption of independence that was used to carry out the simulations may not be realistic. In this case it may be desirable to somewhat relax this assumption and simulate the diameter values form multivariate distributions which allow for dependence. One such distribution is a multivariate distribution which has exponential marginals. Because the tree size diversity index is a function of the dissimilarity coefficient which is calculated using two diameter values at a time, the bivariate exponential distribution (BVE) provides a good source of random numbers which may be used to represent diameter values.

If (X_1,X_2) is BVE, then there exist independent random variables Z_1, Z_2 and Z_{12} such that $X_1 = \min(Z_1,Z_{12})$ and $X_2 = \min(Z_2,Z_{12})$, [56]. This method is used to generate random numbers from twelve distributions. Again four thousand (4000) pairs of random numbers representing diameter values were generated using MINITAB and used to calculate the tree size diversity index. The results are shown on table 4. $Exp(\lambda_1,\lambda_2,\lambda_{12})$ denotes a bivariate exponential distribution with λ_1,λ_2 and λ_{12} being the expected values of the independent and exponentially distributed random variables Z_1, Z_2, and Z_{12} respectively.

Table 4. Results of the simulation study

Distribution	Mean(*d*)	Stdv.(*d*)
Exp(25,25,10)	0,190	0,289
Exp(25,25,15)	0,242	0,310
Exp(25,25,20)	0,280	0,318
Exp(25,25,25)	0,30	0,320
Exp(25,25,30)	0,331	0,321
Exp(25,25,35)	0,346	0,324
Exp(25,25,40)	0,347	0,325
Exp(25,25,50)	0,385	0,323
Exp(25,25,70)	0,410	0,315
Exp(25,25,100)	0,434	0,308
Exp(25,25,500)	0,488	0,294
Exp(25,25,1000)	0,495	0,291

The results from the second simulation study in which dependent exponentially distributed random variables were used to calculate the diversity measure showed that stronger dependence yields low values as expected. Weak dependence gives values that are close to 0,5 which compares well with the values obtained when independence was assumed. Only values from the bivariate exponential distribution were used in this study.

4.2 Testing for size segregation.

Even when all the diameter classes are equally represented, the value for d, the size diversity index, could still be well below 0.5. One reason for this could be that the trees are segregated according to their sizes which results in large trees occurring in the neighbourhood of other large trees and small trees in the neighbourhood of small trees. If such a spatial pattern is suspected to occur, it may be helpful to test whether the pattern is purely due to chance. This can be done by a randomisation test.

A randomisation test can be used to test the hypothesis that a pattern present in the data is purely due to a chance effect of the observations in a random order [54].

In order to perform a randomisation test, a test statistic, T, is chosen to measure the extent to which the data show the pattern in question. The value t_O of T for the observed data is then compared with the distribution of T that is obtained by randomly reordering the data. If all the possible orders of the data are equally likely to occur, then the value t_O will be a typical value from the randomisation distribution of T. An extreme value of t_O suggests that there is weak evidence to support the hypothesis that all the possible orderings of the data are equally likely to occur.

The test statistic which is used to test for size segregation is d, the tree size diversity index. The randomisation distribution for d will reveal that extremely small values of d indicate segregation of trees according to their sizes.

The hypothesis is formally stated as follows:

H_O: The trees do not segregate according to their sizes.

H_A: The trees segregate according to their sizes.

The significance level or p-value is then the proportion of values that are as extreme or more extreme than that of the observed size diversity index among those values in the randomisation distribution.

The randomisation test can be performed as follows,

1. List the diameter values for n pairs of sample trees where each pair consists of the reference tree and its j^{th} nearest neighbour. Save the results in n different columns.

2. Generate random permutations of the diameter values in the columns corresponding to the neighbouring trees using a computer and a quick algorithm [55].

3. Calculate the index d_i and its average d for each of a large sample of permutations. This is a computer intensive method. The resulting distribution will be called the randomisation distribution.

4. If the observed index looks like a typical value from the randomisation distribution, do not reject the null hypothesis and conclude that there is no evidence of size segregation. On the other hand, if the observed value is unusually small, then the data contradicts the null hypothesis and it can be concluded that the alternative hypothesis is more plausible.

If the observed value for the index is in the bottom 5% of the randomisation distribution, one can conclude that the result is significant at the 5% level.
A similar randomisation test can be performed to test whether small trees always occur in the neighbourhood of large trees and vice versa, if there is reason to believe so.

5 Discussion.

A criterion that can be used to describe forest structure as well as to study the effects of management on the forest structure is described here. This criterion, named the tree size diversity, can be used in combination with other variables to decide the weight given to thinnings in even-sized forests. In order to study the various effects of removing particular diameter sizes from a forest stand, a simulation study was carried out. The diameter values used in the simulation study were generated from the Weibull and exponential distributions. The Weibull distribution is used to model diameter sizes because of its flexibility. This distribution has been used in other studies [44] to fit both diameter and height data.

Parts of the distribution are deleted in an effort to see the effects of what can be considered as 'thinnings'. The points at which the distributions are removed are denoted by the suffix -trunc.

The results form the simulation study show that the tree size diversity index stays close to 0.5 for minor deviations from a diameter distribution that resembles the exponential distribution. This is confirmed by the results from the simulations in which parts the exponential as well as the other reversed *J* shaped distributions from the Weibull distribution are deleted. A large truncation point will however result in a low value for the tree size diversity index. Care should therefore be taken when deciding the minimum acceptable diameter.

Young plantations with a diameter distribution resembling that of the normal distribution have low *d* values as expected. The same is true for older even-sized stands unless large portions of these stands are removed.

The removal of parts of the distribution for those distributions which are used to model the diameter distribution of an even-sized stand do not result in a major improvement in the tree size diversity index. This is mainly because the smallest diameter classes are markedly underrepresented. One should however expect that regeneration and ingrowth will occur after thinnings are done. These should bring the d values close to 0.5, but only after some time.

The tree size diversity index d_i depends on the diameter distribution or the relative abundance of the various diameter classes and the spatial arrangement of the trees. If there is reason to suspect an unusual pattern of spatial distribution, hypotheses concerning these patterns can be tested. The hypotheses tests can be used to test whether the trees are clustered according to their sizes, randomly inter-mixed, or whether small trees always occur in the neighbourhood of large trees and vice versa. Randomisation tests are used for such hypotheses, which implies that the resulting p-values are only approximations since they are obtained from samples because obtaining results from large sets of possible permutations is very time consuming. Part of the simulation study was also carried out with the simplifying assumption that the size of one tree does not affect the sizes of its neighbours. The implication is that a tree could not only be infinitely close to another, which is not realistic, but it could as well be much larger or smaller than its neighbours. For these reasons, real life data may give somewhat different results.

On the other hand, the same assumptions were used to derive the distribution for the 'ideal natural' forest whose diameter distribution resembles the exponential distribution. This forest is a standard against which other forests are compared and it should be possible to measure the interference to a forest using the standard.
When quantifying the impact of silvicultural treatment, no assumptions are required concerning the diameter distribution of the stand. An estimate of the tree size diversity before and after a silvicultural treatment such as a thinning will indicate the direction to which the treatment has shifted the spatial arrangement of trees of different diameters. The index d_i is therefore more useful when it is used to study the modifications of forest structure resulting form management as compared to when describing a particular forest state.

The d_i values are rather variable within the same stand and better results should be obtained when smaller areas are studied at a time instead of sampling from very large and variable stands.

The term diversity as used in this study does not fit the traditional meaning of biological diversity. Biodiversity indices use the relative abundances as well as the distribution of species or other characteristics of interest in a slightly different way. An area with a large number of species whose abundance is uniform across the different classes is considered to be very diverse in the traditional sense. For this reason, the broken stick model is chosen as a ‘yardstick’ because it seems to represent the maximum equitability for biological communities. In this study, a forest stand with a diameter distribution that resembles the exponential distribution, and whose tree sizes are not dependent on the sizes of their neighbours, is considered to be the standard.

While the geometric model of species abundance distribution indicates a high level of dominance and generally poor diversity, its continuous equivalent, the exponential distribution, when used to model diameters, provides the standard for evaluating tree size diversity. The tree size diversity of any forest under study can then be obtained by comparing it to this standard.

References.

1. Soulé, M.E and Kohm, K.A., eds 1989. *Research Priorities for conservation Biology*, Island Press.

2. Sandlund, O.T. et al., 1992. *Conservation of Biodiversity for Sustainable Development,* Scandinavian University Press.

3. Groombridge, B., 1992. *Global Biodiversity - Status of the Earth's living resources,* Chapman and Hall.

4. MacArthur, R.H. and MacArthur, J. W., 1961. On bird species diversity. *Ecology,* **42**, 594-8.

5. Southwood, T.R.E. et al., 1979. The relationship of plant and insect diversities in succession. *Biol. J. Linn. Soc.,* **12**, 327-48.

6. Magurran, A.E., 1987. *Ecological Diversity and its measurement*, Chapman and Hall.

7. Elton, C.S., and Miller, R.S., 1954. The ecological survey of animal communities with a practical system of classifying habitats by structural characteristics. *J. Ecol.,* **42**, 460-96.

8. MacArthur, R.H. and Horn, H.S., 1969. Foliage profile by vertical measurements. *Ecology,* **50**, 802-4.

9. Terborgh, J., 1977. Bird species diversity on an Andean elevation gradient. *Ecology,* **58**, 1007-19.

10. Moss, D., 1978. Diversity of woodland song-bird populations. *J. Anim. Ecol,* **47**, 521-47.

11. Blondel, J. and Cuvillier, R., 1977. Une methode simple et rapide pour decrire les habitats d'oiseaux: le stratiscope. *Oikos,* **29**, 326-31.

12. Margalef, R., 1972. Homage to Evelyn Hutchson, or why is there an upper limit to diversity. *Trans Connect. Acad. Arts Sci.* **44**, 211-35.

13. Kempton, R.A. and Tylor, L.R., 1976. Models and statistics for species diversity. *Nature,* **262**, 818-20.

14. Pielou, E.C., 1975. *Ecological Diversity,* Wiley, New York.

15. Simpson, E.H., 1949. Measurements of diversity. *Nature,* **163**, 688.

16. McIntosh, R.P., 1967. An index of diversity and the relation of certain concepts to diversity. *Ecology,* **48**, 392-404.

17. Berger, W.H. and Parker, F.L., 1970. Diversity of planktonic Forminifera in deep sea. *Science,* **168**, 1345-7.

18. Whittaker, R.H., 1965. Dominance and diversity in land plant communities, *Science,* **147**, 250-60.

19. Whittaker, R.H., 1970. *Communities and Ecosystems*, Macmillan, New York.

20. Whittaker, R.H., 1972. Evolution and measurement of species diversity, *Taxon*, **21**, 213-51.

21. May R.M., 1975. Patterns of species abundance and diversity. In: *Ecology and Evolution of Communities* (eds M.L. Cody and J.M Diamond), Havard University Press, Princeton.

22. Engen. S., 1978. *Stochastic abundance models with Emphasis on Biological Communities and Species Diversity,* Chapman and Hall, London.

23. Thomas, M.R. and Shattock, R.C., 1986. Filamentous fungal associations in the phylloplane of *Lolium perenne. Trans. Brit. Mycol. Soc.* **87**, 255-68.

24. Poole, R.W., 1974. *An introduction to Quantitative Ecology,* McGraw- Hill Kogakusha, Tokyo.

25. May, R.M., 1974. *General Introduction. In: Ecological Stability* (eds M.B Usher and M.H. Willamson), Chapman and hall, London, pp1-14.

26. Cohen, J. E., 1968. Alternative derivations of a species abundance relation. *Amer. Nat.,* **102**, 165-72.

27. Saastamoinen, O. and Tyrvänen, L., 1993. Biodiversity and Finnish Forestry: Discussion on Concepts and Issues. In Linddal, L. and Naskali, A., (eds). *Valuing Biodiversity. Proceedings from a workshop on the social costs and benefits from preserving endangered species and biodiversity of the Boreal forests.* Scandinavian Forest Economics **34**:110-123.

28. Whitemore, T.C., 1978. Gaps in the forest canopy: In Tomlinson, P. B. and Zimmerman, M.H., (eds). *Tropical trees as living systems. Proc. Cabot. symp. Havard Forest (1976).* Cambridge, England. pp 639-655.

29. Daniel, Th. W., et al., 1979. *Principles of Silviculture.* 2nd ed. McGraw Hill Book Company. 500p.

30. Lähde, E., et al., 1991. Stand structure of advanced virgin forests in Finland. Scandinavian *Journal of Forest Research* **6(4):** 527-537.

31. Hånell, B., 1984. Skogsdikningsboniteten hos Sveriges torvmarker. Sveriges lantbruksuniversitet. *Rapporter i skogsekology och skoglig marklära* 50. 125p.

32. Lähde, E., et al 1992. Stand structure of thinning mature conifer-dominated forests in Boreal zone. In Hagner, M., (ed). *Silvicultural alternatives.* Proceedings from an internodic workshop June 22-25 1992. Swedish University of Agricultural Sciences. Department of Silviculture. Reports, No **35**:58-65.

33. Smith, A. D.,1962. *The practice of Silviculture.* 7th ed. New York. John Wiley and Sons.

34. Husch, B., et al., 1982. Forest mensuration. 3rd ed. New York. John Wiley and Sons.

35. Schrender, H.T., et al., 1993. *Sampling methods for Multiresource Forest Inventory.* New York. John Wiley and Sons.

36. Phillip, M. S., 1994. *Measuring Trees and Forests.* 2nd ed. Cambridge University Press 310pp.

37. Meyer, H. A., 1953. *Forest Mensuration.* Penns Valley Publishers Inc., State College, DA, USA.

38. Gringrich, S. F.,1967 Measuring and evaluating stocking and stand density in upland hardwood forests in the Central States. *Forest science* **13**:38-53.

39. Meyer, H. A., 1930. Diameter distribution series in even-aged forest stands. Yale Univ. Sch. for. Bull. 28, 105p.

40. Nelson, T. C., 1964. Diameter distribution and growth of loblolly pine. *Forest Science.* **10**:105-115.

41. Osborne , J. G. and Schumacher, F.X., 1935. The construction of a normal-yield and stand tables for even-aged timber stands. *J. Agric. Res.* **51**:547-564.

42. Shnur, G.l., 1934. Diameter distributions for old-field loblolly pine stands in Maryland. *J. Agric. Res.* **49**:731-743.

43. Bliss, C. L. and Reinker, K. A., 1964. A log-normal approach to diameter distributions in even-aged stands. *Forest Science*. **10**:350-360.

44. Clutter, J.L. and Bennett, F. A., *Diameter distributions in old-field slash pine plantations.* Ga. For. Res.Counc. Rep. 13 p.

45. Bailey, R. L. and Dell, T.R.,1973. Quantifying Diameter distributions with the Weibull function. *Forest Science* **19**:97-104.

46. Little, S.N., 1984. Weibull diameter distributions for mixed stands of different conifers. *Canadian Journal of Forest Research* **13(1)**, 100-107.

47. Renolls, K., et al., 1985. Characterising diameter distributions by the use of the Weibull distribution. *Forest* **58(1)**, 57-66.

48. Hafley, W, L and Schreuder, H. T., 1977. Statistical distributions for fitting diameter and height data in even-aged stands. *Canadian Journal of Forest Research* **7**:481-487.

49. Cramér, H., 1963. *Mathematical Methods of Statistics*. Princeton University Press. Princeton. pp 356-57.

50. Johnson, N.L. and Kotz, S., 1970. *Distributions in Statistics*. Houghton Muffin Co. Boston MA.

51. Fülder, K, and Gadow von Klaus., 1994. How to define a Thinning in a Mixed Deciduous Beech Forest. *Paper presented at the IUFRO Conference in Lousa, Portugal, 25-29 April 1994.*

52. Hagner, M., and Nyquist, H., 1995. *Description of dissimilarities between neighbouring trees*. Unpublished manuscript.

53. Pielou, E.C., 1969. *An introduction to Mathematical Ecology.* Wiley, New York.

54. Manly, B. F. J., 1991. *Randomization and Monte Carlo methods in Biology.* Chapman and Hall.

55. Knuth, D.E., 1981. *The Art of Computer programming,* Vol 2, Semi-numerical Algorithms, Addison-Wesley, Reading, Massachusetts.

56. Marshall, W.A., 1967. A multivariate exponential distribution. *American Statistical Journal pp 30-44.*

Serien Arbetsrapporter utges i första hand för institutionens eget behov av viss dokumentation.

Författarna svarar själva för rapporternas vetenskapliga innehåll.

1995 1 Kempe, G. Hjälpmedel för bestämning av slutenhet i plant- och ungskog. ISRN SLU-SRG-AR--1--SE

2 Riksskogstaxeringen och Ståndortskarteringen vid regional miljöövervakning. - metoder för att förbättra upplösningen vid inventering i skogliga avrinningsområden. ISRN SLU-SRG-AR--2--SE.

3 Holmgren, P. & Thuresson, T. Skoglig planering på amerikanska västkusten - intryck från en studieresa till Oregon, Washington och British Columbia 1-14 augusti 1995. ISRN SLU-SRG-AR--3--SE.

4 Ståhl, G. The Transect Relascope - An Instrument for the Quantification of Coarse Woody Debris. ISRN SLU-SRG-AR--4--SE.

5 Törnquist, K. Ekologisk landskapsplanering i svenskt skogsbruk - hur började det?. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--5--SE.

1996 6 Persson, S. & Segner, U. Aspekter kring datakvalitėns betydelse för den kortsiktiga planeringen. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--6--SE.

7 Henriksson, L. The thinning quotient - a relevant description of a thinning? Gallringskvot - en tillförlitlig beskrivning av en gallring? Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--7--SE.

8 Ranvald, C. Sortimentsinriktad avverkning. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--8--SE.

9 Olofsson, C. Mångbruk i ett landskapsperspektiv - En fallstudie på MoDo Skog AB, Örnsköldsviks förvaltning. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--9--SE.

10 Andersson, H. Taper curve functions and quality estimation for Common Oak (Quercus Robur L.) in Sweden. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--10--SE.

11 Djurberg, H. Den skogliga informationens roll i ett kundanpassat virkesflöde. - En bakgrundsstudie samt simulering av inventeringsmetoders inverkan på noggrannhet i leveransprognoser till sågverk. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--11--SE.

12 Bredberg, J. Skattning av ålder och andra beståndsvariabler - en fallstudie baserad på MoDo:s indelningsrutiner. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--12--SE.

13 Gunnarsson, F. On the potential of Kriging for forest management planning. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--13--SE.

14 Holm, S. & Thuresson, T. samt jägm.studenter kurs 92/96. En analys av skogstillståndet samt några alternativa avverkningsberäkningar för en del av Östads säteri. ISRN SLU-SRG-AR--14--SE.

15 van Kerkvoorde, M. A sequential approach in mathematical programming to include spatial aspects of biodiversity in long range forest management planning. ISRN SLU-SRG-AR--15--SE.

16 Tormalm, K. Implementering av FSC-certifiering av mindre enskilda markägares skogsbruk. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--16--SE.

1997 17 Engberg, M. Naturvärden i skog lämnad vid slutavverkning. - En inventering av upp till 35 år gamla föryngringsytor på Sundsvalls arbetsomsåde, SCA. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN-SRG-AR--17--SE.

18 Christoffersson, P & Jonsson, P. Avdelningsfri inventering - tillvägagångssätt och tidsåtgång. ISRN-SRG-AR--18--SE.

19 Ståhl, G., Ringvall, A. & Lämås, T. Guided transect sampling - An outline of the principle. ISRN-SRG-AR--19--SE.

20 Cedervind, J. GPS under krontak i skog. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN-SRG-AR--20--SE.

21 Holm, S. & Thuresson, T. samt jägm.studenter kurs 93/97. En analys av skogstillståndet samt några alternativa avverkningsberäkningar för en stor del av Östads säteri. ISRN SLU-SRG-AR--21--SE.

22 Ali, Abdul Aziz. Describing Tree Size Diversity. ISRN SLU-SRG-AR--22--SE.