STEM FREQUENCY DISTRIBUTION AND SAMPLING OF THE HARDWOOD COMPONENT OF PINE-HARDWOOD STANDS IN SOUTHEASTERN OKLAHOMA

By

WILLIAM WYCOFF THOMAS

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1965

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
MASTER OF SCIENCE
December, 1973

STEM FREQUENCY DISTRIBUTION AND SAMPLING OF THE

HARDWOOD COMPONENT OF PINE-HARDWOOD STANDS

IN SOUTHEASTERN OKLAHOMA

Thesis Approved:

Nat Walker
Thesis Adviser

Ronald W. McNew

J Edward Fmzeniey

N N Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to thank Dierks Division of the Weyerhaeuser Company for the research data, and the Oklahoma State University Forestry Department for providing the funds and facilities necessary to complete this research project.

I wish to express my deep appreciation to my major advisor, Dr. Nathanial Walker, and the other members of my committee, Drs. R. W. McNew and J. E. Langwig for their assistance, encouragement and patience in the pursuit of this research project and the preparation of this thesis.

Special thanks to my typist, Verna Harrison, for her diligent efforts and expertise in my behalf.

Last, but certainly not least, special gratitude to my wife, Jo Ann, for her sacrifices, love and support in this endeavor, and my son, Brian, for being the best little inspiration I've ever known.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Modern forestry recognizes that there exist many important resources besides trees in the forest. The public demands that water, wildlife, range and recreation be considered in any forest management plan. Consequently, management costs have risen considerably. These costs must be recovered for the forest to be economically self-sustaining, and therefore, of greatest benefit to society. The industrial forest owner must obtain a profit. Timber is the resource with the greatest potential for generating revenues equal to or exceeding costs. Such net revenues can be realized only if proper management is applied. The opportunities for management of forest stands exist continuously through three general periods in time, namely, through the establishment, growth and harvest periods. This thesis is concerned with research in the field of standing timber management, that is, during the growth stage.

Most forest stands contain stems of many different diameters. For convenience, discrete diameter classes are established. Generally, diameter classes have a range of either one or two inches and are designated by the mid-point of the class. For example, the 10-inch diameter class will contain all stems from 9.6 inches to 10.5 inches, or it will contain all stems from 9.1 inches to 11.0 inches, depending on whether one-inch or two-inch classes are used. The two-inch diameter interval was used throughout this research study. The point on the stem at which the diameter is measured is the standard 4.5 feet above the ground, that is, diameter at breast height (dbh).

While the manager's ultimate goal is to maximize the wood production of the forest, he must consider two elements: first, the average volume of wood (cubic feet or board feet) obtainable from trees of various diameters; second, the frequency distribution of the various diameter trees over a given area (structure). By combining these two elements of forest structure, more information helpful to the manager is available. Much research has been done relating volumes of trees to their respective diameters. It is the purpose of this research to investigate the second element, that is, frequency distributions (structure) of hardwood stands in the pine-hardwood forests of Southeastern Oklahoma.

Differences between the observed structures and the desired structures form a basis for management plans. The desired distributions of stems of various diameters are constructed from the empirical observations of the spatial arrangements of trees similar in forest types, on similar sites, and under similar conditions. Having determined what structure is desirable for a given forest, the manager must then have an accurate picture of the actual structure that exists.

Conventional inventory methods now employed use standard statistical procedures to obtain estimates of the average number of trees per acre of the individual diameter classes. When these mean frequencies are plotted by diameter class, a geometric series representing the distribution of diameters on the average acre often results. It is this average acre on which regulatory practice is based. Averages often conceal much of the total picture. The ability to predict the frequency distributions of stems from which the mean is derived aids in bringing the actual forest structure into sharper focus. The research described by this paper is considered to add an additional dimension to current methodology rather than to supplant any facet of that methodology.

The work of several investigators in the field (Ker, Walker, et al) suggests that tree stem distributions, in common with many other biological forms, tend to exhibit "contagion", characterized statistically by variances exceeding means, and by clumpy spatial arrangements of the plants. If satisfactory mathematical equations can be found to describe such distributions, then it becomes possible to use such equations as predictors, providing that the required parameters of the predicting equations can be supplied. Such a predictor has been found in the negative binomial distribution, first formulated by Montmort, a French mathematician, in 1714, and later developed by Anscombe, Bliss and Fisher. Since in forestry, the costs of sampling normally are directly related to the intensity of sampling, the success of any predictive function must be measured by the degree of sampling necessary to achieve reasonable estimates of the required parameters.

With a goal of satisfactory parameter estimation, the results of numerous computer runs of different sampling intensities were obtained for comparison. Additionally, the frequency data between combinations of two diameter classes reflect the tolerance, or intolerance, to coexistence in the same area. Included in this paper are the results of the examination of joint frequencies for all possible combinations of two diameter classes.

Data for this project were obtained from 1253 one-tenth acre permanent inventory plots located in the Oklahoma Block of the Dierks Division of the Weyerhaeuser Company. The total area represented by these plots is 890,000 acres. Only the hardwood component of the pine-hardwood type was analyzed in this study. In view of the findings of Walker (1970), it was expected that less contagion would be found in the hardwood component than in the pine component which was studied by Dahlem (1972).

The negative binomial series requires the estimation of two parameters - the mean, μ, and a positive exponent, $\underline{k}$. The IBM 360 computer was utilized to obtain multiple samples at both 10 and 20 percent sampling intensities. The means of the distributions of sample means were used as a basis for computing confidence limits. Both the moment estimate of $\underline{k}$ ($\tilde{k}$) and the attendant maximum likelihood estimate of $\underline{k}$ ($\hat{k}$) were calculated for each sample. The relationships that existed between these two estimates of $\underline{k}$ were used in improving the programming efficiency of deriving the maximum likelihood estimate from the moment estimate by the iterative method of Fisher (1953). Particular attention was given to those instances in which the $\tilde{k}$ values were negative.

This research will (1) compare the actual frequency distributions of a hardwood population to the negative binomial distributions calculable from the μ and $\underline{k}$ values of the population, (2) examine the significance of associations between occurrences in combination of two diameter classes, (3) look into the effects of sampling intensity on confidence limits of the mean and the effects on the resultant frequency distributions, (4) provide an efficient computer method for obtaining $\hat{k}$ from $\tilde{k}$, regardless of whether $\tilde{k}$ is a positive or a negative value.

CHAPTER II

LITERATURE REVIEW

Although the statistical models of numerous frequency distributions of individual populations have existed for centuries, only since about 1900 has much work been done to compare the calculated distributions to the distributions of populations actually found in nature. It was felt that most natural populations were distributed at random over a given area and therefore could be fitted with the Poisson series. Observers soon discovered that with both plant and animal populations, a tendency existed for individuals to occur together. Chronologically, the movement toward use of the negative binomial to describe forest populations has been as follows:

Clapham (1936) noted the aggregative tendency of individuals in grassland communities and used the term "over-dispersion" to describe it. Much preferred has been the term "contagion" as coined by Polya in 1931 and used by Neyman (1939) in his work on insect and bacterial colonies. Neyman showed that contagious distributions can arise from combining the random distribution of groups per unit area with the logarithmic distribution of the individuals within the groups. Based upon his observations, he formulated his own frequency distribution series in which the frequency values are assumed to be proportional to the mean frequency value. He found many bimodal populations were satisfactorily described by his series.

Haldane (1941) proposed that binomial distributions could be fitted more efficiently by using the maximum likelihood estimates of the parameters rather than the moment estimate normally used. Fisher (1941), in the same publication, dealt specifically with the negative binomial series. He noted that the negative binomial differed from the positive binomial in that the exponent must be estimated as well as the mean of the distribution. He also demonstrated that with large samples, an efficiency of ninety percent is obtainable with moment estimates of the exponent $\underline{k}$.

Cole (1946) viewed most contagious distributions as being composed of intermingled random distributions. Each distribution is based on the mean frequencies of groups containing 0, 1, 2, ..., X individuals. Individual probabilities of 0, 1, 2, ..., X counts are combined and multiplied by the $\underline{N}$ of the population to give the expected frequencies. Empirical results obtained through analysis of rodent ectoparasite populations gave supportive results. Cole admitted that his concept of contagiously distributed populations is simpler than many of the cases found in nature.

Neyman's series was given support by Archibald (1948) working with bimodal populations of small plants. He concluded that Neyman's series was a generalized form of the Poisson. But where the Poisson failed because of too many empty quadrats and too few quadrats with a few individuals, Neyman's series excelled. This spotlighted the fundamental point that heterogenous biological populations are likely to show contagion.

A proof that a logarithmic distribution of individuals within groups combines with the Poisson distribution of the groups to form the negative binomial was provided by Quenouille (1949). He suggested that the variance/mean ratio provides a relative quantification of the degree of contagion.

Anscombe (1950) gave a concise statement of four ways in which the negative binomial distribution can arise: (1) inverse binomial sampling, (2) heterogenous Poisson sampling, (3) randomly distributed colonies and (4) the immigration-birth-death process. He also provided a very good comparative analysis between several of the distributions most frequently used in the evaluation of biological populations. These included Neyman Type A, Thomas, Polya and the negative binomial. The negative binomial is felt to be superior since it is always unimodal and is easiest to work with. In the case of a negative moment estimate of $\underline{k}$ ($\tilde{k}$) which is caused by the distribution mean exceeding the distribution variance, Anscombe felt that the related $\underline{k}$ would be infinite and therefore the distribution would be Poisson. The availability of computers permitted the problem of maximum likelihood estimates of $\underline{k}$ ($\hat{k}$) that are derived from negative $\tilde{k}$ values to be studied in this research project. This thesis will show that negative moment estimates of $\underline{k}$ have attendant maximum likelihood values of $\underline{k}$ that are very large positive integers.

Bliss (1953) clarified and simplified the procedure utilized to obtain the negative binomial frequency distribution. Working with various plant and animal populations having variances significantly larger than the means, he found the negative binomial to be the best predictor of the frequencies of groups of individuals of the populations. His work also provided the first example of the method of maximum likelihood estimation of $\underline{k}$ as developed by Fisher (1953).

Fisher's method of maximum likelihood estimation of $\underline{k}$ is given, beginning on page 15. Robinson (1954) worked on a quadrat basis with high and low prairie plant populations and found the negative binomial fit to be best.

The fit of pine seedling distributions to the Poisson and the negative binomial frequency distributions was examined by Ker (1954). He found that seedlings in quadrats conformed to the negative binomial except in isolated instances where the stocking density per unit area was greater than ninety-five percent.

Waters and Henson (1959) suggested the negative binomial be the standard for work with forest insects. They examined the frequency data gathered by many studies of various forest insects, and found good fit to the negative binomial in almost every instance. Pielou (1960) found that a positive correlation between the nearest neighbor and the sum of the trunk circumferences existed for regular, random and aggregated populations of trees.

The strongest evidence of the frequency distribution of forest trees and size class associations was presented by Walker (1970). His data and analysis of various types of forests indicate the wide applicability of the negative binomial. Dahlem (1972), working with pine, showed that when a population conforms to the negative binomial, relatively low intensity samples may provide an estimate of the mean that can be combined with a known $\hat{k}$ and the resultant negative binomial frequency distribution will closely resemble the actual forest structure. Gerrard and Cook (1972) have shown that an unbiased, minimum variance estimate of the mean is obtainable from the presence and absence of data of inverse binomial sampling.

This review of literature shows, in a historical perspective, how the negative binomial has come to be recognized as generally the best reflector of the structure of actual biological populations.

CHAPTER III

PROCEDURE

The Oklahoma Block of the Dierks Division, Weyerhaeuser Company, provided the raw data for this study. The block is sub-divided into six districts of unequal geographical size and contains approximately 890,000 acres of mixed pine and hardwood species. Data collected on 1253 Continuous Forest Inventory (CFI) plots became the population on which the research was performed. The CFI plots are circular with a fixed radius and area of one-tenth acre.

The number of stems in each diameter class was recorded on a per acre basis, by plot and by district. In later recording on IBM computer cards, the data were converted to a stems per plot basis. One card was punched for each plot. Standard computer programs provided frequency tables for each district and the block. The form of the frequency tables may be seen by referring to the tables beginning on page . The tables show the frequency of plots having 0, 1, 2, ..., X four-inch, six-inch, eight-inch, et cetera, stems. Accumulative frequencies were derived and entered on IBM cards. These cards became the data deck for a computer program that produced means, variances, maximum likelihood values of <u>k</u> and expected univariate negative binomial frequencies. The mean frequency of each individual distribution was calculated as:

$$\bar{x} = \frac{\Sigma fx}{N},$$

where N is the population size (number of plots) and Σfx is the total count. The variance of each distribution was calculated as:

$$s^2 = \frac{\Sigma(fx^2) - \frac{(\Sigma fx)^2}{N}}{N - 1}.$$

The $\hat{k}$ value for each distribution was arrived at through the iterative balancing of the equation (Fisher, 1953):

$$Z_i = \Sigma\left(\frac{Ax}{k_i + x}\right) - N \cdot \ln\left(1 + \frac{\bar{x}}{k_i}\right),$$

so that Z_i will equal zero plus or minus 0.0001. The initial k_i value chosen in each case is the moment estimate, $\tilde{k}$, which is obtained as:

$$\tilde{k} = \frac{\bar{x}^2}{s^2 - \bar{x}}.$$

An example of the above and more detailed discussion are presented in the next chapter.

The fit of generated negative binomial distributions to their respective actual distributions was tested by chi-square. The goodness-of-fit was examined for all distributions by diameter class, for each district, and for all the districts combined (block).

The computer was used to construct joint frequency tables for all possible combinations of diameter classes. Two-way contingency tables showing presence-absence relationships were then built for each joint frequency table. The null hypothesis that different diameter classes were distributed independently was tested by chi-square. According to standard statistical procedures, significance was established at the five percent level, with one degree of freedom. All significant associations were then examined to determine if they were positive (one diameter class related to the presence of another) or negative (one diameter class related to the absence of another). This was done by the method of Cole (1946) in which * = $\underline{N} \cdot a/[(a+b)(a+c)]$, where a, b, c, d, represent the values of the upper left, upper right, lower left and low lower right cells of a contingency table, respectively. If * is greater than one, the association is positive, and if * is less than one, the association is negative. Logic points to perhaps a simpler test for the determination of the direction of association. Comparing the product of cross-multiplication of cells a and d with the product of cells b and c shows that when (a·d) is greater than (b·c), the association is positive and when the reverse is true, the association is negative.

Sampling of negative binomial distributions was performed by diameter class, for each district and the block. Negative binomial frequency distributions were calculated using the $\hat{k}$ of the block, N of the district and the $\bar{x}$ of the distribution of sample means. These expected distributions were compared with the actual distributions of the districts by diameter class. Chi-square could not be used as a test for goodness-of-fit since the $\hat{k}$, N and $\bar{x}$ values were from different levels. Therefore, $\hat{k}$ and N were held constant and $\bar{x}$ was varied. The resultant distributions were tested with chi-square. The number of computer samples considered sufficient to provide a good $\bar{x}$ (mean of the distribution of means) was 250 and the intensities examined were 10 and 20 per cent.

The possibility of establishing confidence limits in the estimation of the mean was studied. By diameter class and intensity, the number of sample means falling within plus or minus 10 per cent of the mean value of the distribution of means was calculated. Division by 250 then yielded the probability of any single sample mean falling within the desired range, at the intensity under investigation. It was felt that any mean falling within plus or minus ten per cent of the true mean will prove to be practicable in the generation of a representative negative binomial distribution. By holding the probability and desired range values constant, the variation in sampling intensity by diameter class was calculated.

Trial and error programming led to improvements in the efficiency of calculating the maximum likelihood estimate of k from either a positive or negative moment estimate of k. The results are discussed in more detail beginning on page 43.

CHAPTER IV

RESULTS AND DISCUSSION

Stem Frequency Distribution

The negative binomial was the only distribution used in this research work. Sufficient evidence of its applicability and suitability in describing the stem frequencies of forest stands was provided by Walker (1970). The reader may find it helpful to be able to see how it is related to the Poisson distribution, therefore a limited discussion of the two distributions follows.

The negative binomial was derived from empirical observations of variable biological populations in the early 18th century. The populations did not fit the Poisson distribution as had been anticipated. The spatial frequencies of individuals did not reflect the randomness of the Poisson, but tended to be aggregative. This gregariousness yielded distributions in which the variances exceeded the means. Poisson distributions are characterized by variances equal to the means. Dependent upon the closeness of the mean and variance to equality, distributions range from nearly random (Poisson) to highly aggregative (logarithmic).

Neyman, Thomas and Polya were but three of many who formulated numerous distributions based on the specific relationship between the mean and variance of particular populations. These tended to fail when applied on a broad basis because they could not handle the changes in the degree of contagion. The negative binomial, however, appears to have a much greater flexibility, tending to fit distributions with varying degrees of aggregation, providing that the distribution curves are unimodal.

With populations distributed according to the Poisson, the probability of any individual being present is equal to the probability of any other individual being present. In aggregated (contagious) populations, the probability of an individual being present is affected by the presence or absence of other individuals of the same type. Quenouille (1949) found in counting individual bacterium and colonies that "whereas the colony counts followed Poisson's distribution, the number of bacteria per colony were logarithmically distributed, and that consequently the bacterial counts were distributed in the negative binomial form". Since the mean and variance are equal in the Poisson, the mean is the only parameter required for the generation of a model. With the negative binomial, a second parameter is required to reflect the inequality of mean and variance. The parameter is called $\underline{k}$. It is the exponent in the basic probability equation $(q - p)^{-\underline{k}}$, where $q = 1 + p$ and $p = \mu/\underline{k}$. The relative frequencies (θ's) of individual units having 0, 1, 2, ..., et cetera, sequential counts are obtained through expansion of the probability equation.

There are two commonly used methods for obtaining an estimate of $\underline{k}$. The moment estimate, $\tilde{k} = \bar{x}/(s^2 - \bar{x})$, is most efficient when $\underline{k}$ is less than 1.0, indicating a highly contagious distribution. However, most hardwoods are distributed with only a moderate degree of contagion, and have $\underline{k}$ values greater than 2.0 (Walker, 1970). In this situation, the maximum likelihood estimate, $\hat{k}$, proves to be the best. Calculating $\hat{k}$ requires iterative solutions in balancing the following equation, so that Z_i is within plus or minus 0.0001 of zero. The equation is:

$$Z_i = \Sigma \left(\frac{Ax}{k_i + x}\right) - N \cdot \ln\left(1 + \frac{\bar{x}}{k_i}\right),$$

where Ax is the accumulated frequency for x count and k_i is the value at the ith iterate. In the second term of the equation, 2.3025851(N) Log_{10} may be substituted for N·ln. The moment estimate is used as a first estimate of k ($k_i = \tilde{k}$). In the example that follows, 0.001 is used as a limit rather than 0.0001. Table I shows the data necessary for begining the calculation of a Z_i value.

TABLE I

THE DISTRIBUTION OF 12-INCH HARDWOOD STEMS ON 1253 SAMPLE PLOTS

Stem Count (x)	Number of Samples (f)	Total Count (fx)	Accumulated Frequency (Ax)
0	826	0	427
1	300	300	127
2	96	192	31
3	22	66	9
4	6	24	3
5	3	15	0
	$\Sigma f = N = 1253$	$\Sigma fx = 597$	

From the above table, estimates of the mean and variance are obtained. Using the formulas presented earlier, we find $\bar{x} = 0.47645$ and $s^2 = 0.61386$. The moment estimate of k then becomes:

$$\tilde{k} = \frac{(0.47645)^2}{0.61386 - 0.47645} = \frac{0.22700}{0.13741} = 1.65196.$$

As stated previously, this value becomes k_1 in the search for the maximum likelihood value of $\underset{\sim}{k}$. Table II demonstrates the calculation of the first term of the equation and the first iteration proceeds as follows:

TABLE II

VALUES OF THE FIRST TERM OF THE Z_1 ITERATION EQUATION IN THE SEARCH FOR $\hat{k}$

Stem Count (x)	Accumulated Frequency (Ax)	$\frac{1}{k_1 + x}$	$Ax\left(\frac{1}{k_1 + x}\right)$
0	427	0.60534	258.48020
1	127	0.37707	47.88903
2	31	0.27382	8.48858
3	9	0.21496	1.93466
4	3	0.17692	0.53078
			Σ = 317.32328

By substitution, we get;

$$Z_1 = 317.32328 - 2.3025851\ (1253)\ \text{Log}_{10}\left(1 \quad \frac{0.47645}{1.65196}\right)$$

$$= 317.32328 - 317.52134 = -\ 0.13806$$

Note that 2.3025851(1253) will remain a constant throughout the search for this $\hat{k}$. Since Z_1 is negative, the value of k_1 must be decreased. A value of 0.1 is selected for the magnitude of change and $k_2 = 1.55196$.

The first term of the equation for the second iteration is 336.14592 and the second term of the equation is 335.47018, yielding a Z_2 = 0.67573. As this value is positive, it follows that k_3 is greater than k_2 and less than k_1. By interpolation, k_3 is determined to be 1.63499.

Successive calculations yield;

$$k_3 = 1.63499 \text{ produces } Z_3 = -0.07050,$$

$$k_4 = 1.62714 \text{ produces } Z_4 = -0.00735,$$

$$k_5 = 1.62633 \text{ produces } Z_5 = -0.00075.$$

The maximum likelihood estimate of $\underline{k}$ is therefore 1.62633, since the resultant Z_i meets the requirement of falling in the range of 0.000 ± 0.001.

Once the $\hat{k}$ is obtained for the distribution of a diameter class, the next step is the calculation of the expected negative binomial frequencies (θ's). Following the procedures developed primarily by Bliss (1953). To avoid the difficulties inherent in working with exponents, it is best to use the logarithmic equivalent for the θ_0 equation. So, instead of:

$$\theta_0 = \frac{N}{q}k, \text{ where } q = \left(1 + \frac{\bar{x}}{\hat{k}}\right),$$

we obtain, Log θ_0 = Log $\underline{N}$ - $\hat{k}$ Log q. The main statistics required are q and a satisfactory estimate of $\underline{k}$ (either $\tilde{k}$ or $\hat{k}$). Using the 12-inch diameter class of the block, we find the frequency of plots having zero stems is;

$$\text{Log } \theta_0 = \text{Log } (1253) - 1.62633 \text{ Log } (1.29296),$$

$$= 3.09795 - 0.18146 = 2.91646,$$

and the antilog gives a θ_0 value of 825.04948. Successive frequencies are calculated from:

$$\theta_x = \frac{\hat{k} + x - 1}{x} \cdot \underline{R} \cdot \theta_{x-1}$$

where x = stem count of 1, 2, 3, et cetera, sequentially, and $\underline{R}$ = $\bar{x}/(\hat{k} + \bar{x})$. Since the values for $\bar{x}$ and $\hat{k}$ do not change within the diameter class, $\underline{R}$ has a constant value of 0.22658. The expected negative binomial frequencies for the 12-inch class of the block are shown in Table III.

TABLE III

EXPECTED NEGATIVE BINOMIAL FREQUENCIES (θ's) AND CHI-SQUARE FOR THE 12-INCH DIAMETER CLASS

Count (x)	$\frac{\hat{k} + x - 1}{x}$	Calculated Expectation (θ_x)	Rounded Expectation (θ_x)	Observed Frequencies	Chi-square $\frac{(f - \theta_x)^2}{\theta_x}$
0		825.04948	825	826	0.00121
1	1.62633	304.02566	304	300	0.05263
2	1.31316	90.45885	90	96	0.40000
3	1.20878	24.77528	25	22	0.36000
4	1.15658	6.49257	7 }	6 }	0.00000
5	1.12527	1.65536	2 }	3 }	
		$\Sigma = N =$	1253	= 1253	$\Sigma =$ 0.81384

The chi-square value included in the above table is used to determine the goodness-of-fit between the expected and the observed distributions. The number of degrees of freedom (df) is calculated as Y - 3, where Y equals the number of individual frequency counts of the expected distribution. In Table III, six expected frequencies were calculated (0 through 5). The fifth and sixth entries were combined in order to meet the requirement of using no expected frequency value less than three. The number of entries is therefore reduced to five. One d.f. is lost on the total and two more on the estimated statistics, $\bar{x}$ and $\hat{k}$. The number of degrees of freedom available is subsequently

5 - 3 = 2. A chi-square table gives an approximate probability value of 0.67 for $X^2 = 0.81384$ with two degrees of freedom. For goodness-of-fit testing, the limit of significance is considered to occur where the probability value is equal to 0.05. With $\underline{P} = 0.67$ for the 12-inch diameter class, the satisfactory fit of the observed distribution to the calculated negative binomial distribution, is indicated. Due to the need for a minimum one degree of freedom, distributions with fewer than four count categories could not be tested. This generally occurred in the larger diameter classes (14-inch, 16-inch, etc.). Table IV shows the means, variances and maximum likelihood estimates of $\underline{k}$ on the 1253 plots, by diameter class for each of the six districts and the block. In the larger diameter classes, the means are smaller and reflect a reduction in the length of the distributions. The decrease in the values of the variances with increasing diameter class reflects the inefficiency of fixed radius plots, in providing good representation of the true variability that exists in a population. The $\hat{k}$ values are therefore the most critical values in distribution generation, as they reflect the combined effects of length and variation of the observed distributions. The large $\hat{k}$ value for the 14-inch class of district four is caused by the mean being greater than the variance, thus giving a negative moment estimate. It is caused, in part, by the heavy proportion of plots having zero stems and the few number of count categories in the distribution. Some additional discussion of this situation can be found on page 43. Values in Table IV were used to generate the expected negative binomial frequencies shown in Tables V, VI, VII, VIII, IX, X and XI on the following pages.

These seven tables demonstrate the fit obtained on all districts

TABLE IV

MEANS, VARIANCES AND MAXIMUM LIKELIHOOD VALUES OF k

	Diameter Classes						
	4	6	8	10	12	14	16
District 1							
Mean	6.4787	3.6436	1.7500	.8085	.5319	.3191	.1596
Variance	21.9514	7.0862	2.4880	.8401	.7102	.4110	.1883
$\hat{k}$	2.6004	3.3357	4.0105	22.3663	1.5375	.8719	.9993
District 2							
Mean	6.8009	3.5972	1.5355	.7251	.4739	.2417	
Variance	29.0840	9.6798	2.4690	.9431	.6600	.2794	
$\hat{k}$	1.8094	1.9383	2.6608	2.3800	1.1230	1.7178	
District 3							
Mean	6.0699	3.6021	1.6237	.7849	.4624	.2903	.1452
Variance	21.0708	8.3598	3.7279	.9805	.5526	.3369	.1572
$\hat{k}$	2.3373	2.6072	1.2780	3.0885	1.9684	1.5707	2.7618
District 4							
Mean	5.0721	2.5625	1.3317	.7212	.5385	.3125	
Variance	18.7919	5.9188	1.8266	.9267	.7811	.3028	
$\hat{k}$	1.8622	2.2694	3.3692	2.2260	1.2608	112.7985	
District 5							
Mean	4.2896	2.3122	1.1086	.7149	.4570	.2760	
Variance	14.5976	5.7793	1.7882	1.0593	.5129	.3189	
$\hat{k}$	1.8434	1.6573	1.6573	1.7087	3.0990	1.7743	
District 6							
Mean	4.9540	2.7448	1.3347	.7071	.4100	.2008	
Variance	21.1870	6.3169	1.9379	1.0231	.4950	.2284	
$\hat{k}$	1.4261	2.0575	2.5581	2.1540	2.6127	1.4759	
Block							
Mean	5.5619	3.0439	1.4334	.7406	.4765	.2706	.1604
Variance	21.7895	7.4142	2.3656	.9638	.6139	.3093	.1699
$\hat{k}$	1.8074	2.0015	2.1573	2.6105	1.6257	1.8159	2.7578

TABLE V

COMPARISON OF OBSERVED DISTRIBUTIONS AND NEGATIVE BINOMIAL EXPECTED DISTRIBUTIONS BY DIAMETER CLASS FOR DISTRICT 1

Stem Count	Diameter Classes: 4		6		8		10		12		14		16	
	Observed (O) and Expected (E) Frequencies													
	O	E	O	E	O	E	O	E	O	E	O	E	O	E
0	5	7	22	16	45	44	86	85	119	119	144	143	162	162
1	19	14	22	28	51	54	64	66	48	47	30	34	23	22
2	17	17	26	32	44	41	28	27	13	15	12	8	2	3
3	17	19	31	29	22	25	8	8	6	5	2	2	1	1
4	18	19	21	24	14	13	2	2	2	1		1		
5	14	18	24	19	6	6				1				
6	17	16	17	14	4	3								
7	20	14	9	9	2	1								
8	12	12	6	6		1								
9	9	10	5	4										
10	8	8	2	3										
11	3	7	2	2										
12	2	6	0	1										
13	7	5	0	1										
14	6	4	1											
15	4	3												
16	3	2												
17	1	2												
18	3	1												
19	1	1												
20	1	1												
21	0	1												
22	0	1												
23	1													
d.f. =	15		10		5		2		2		2		1	
χ^2 =	13.52883		8.04896		1.17916		.10941		.48794		3.47758		.37879	
P ~	.60		.60		.95		.95		.80		.20		.60	

TABLE VI

COMPARISON OF OBSERVED DISTRIBUTIONS AND NEGATIVE BINOMIAL EXPECTED DISTRIBUTIONS BY DIAMETER CLASS FOR DISTRICT 2

Stem Count	Diameter Classes											
	4		6		8		10		12		14	
	Observed (O) and Expected (E) Frequencies											
	O	E	O	E	O	E	O	E	O	E	O	E
0	16	13	31	28	61	63	112	112	143	142	168	168
1	16	18	30	35	67	61	63	62	43	47	37	36
2	17	20	35	33	34	41	23	25	21	15	4	6
3	19	20	24	28	27	23	9	8	2	5	2	1
4	21	19	19	23	12	12	3	3	1	1		
5	19	17	27	18	6	6	1	1	1	1		
6	10	16	9	13	0	3						
7	12	14	10	8	3	1						
8	11	12	9	7	0	1						
9	12	10	10	5	1							
10	12	9	2	4								
11	8	8	1	3								
12	7	6	0	2								
13	7	5	1	1								
14	2	4	1	1								
15	3	4	0	1								
16	5	3	2									
17	5	3										
18	3	2										
19	0	2										
20	3	1										
21	0	1										
22	1	1										
23	1	1										
24	0	1										
25	0	1										
26	0											
27	0											
28	0											
29	1											
d.f. =	18		13		5		3		3		1	
χ^2 =	11.67940		16.75954		7.54443		.30113		4.54747		1.69444	
P ~	.85		.20		.20		.96		.23		.20	

TABLE VII

COMPARISON OF OBSERVED DISTRIBUTIONS AND NEGATIVE BINOMIAL EXPECTED DISTRIBUTIONS BY DIAMETER CLASS FOR DISTRICT 3

Stem Count	Diameter Classes 4		6		8		10		12		14		16	
	Observed (O) and Expected (E) Frequencies													
	O	E	O	E	O	E	O	E	O	E	O	E	O	E
0	10	9	21	19	64	65	93	92	124	123	143	143	161	162
1	12	16	26	29	50	47	56	58	42	46	33	35	24	22
2	29	19	31	31	29	30	26	24	16	13	9	7	0	2
3	17	20	29	27	16	18	7	8	4	3	1	1	1	
4	11	19	24	22	13	11	3	3		1				
5	19	18	15	17	4	7	1	1						
6	14	15	10	13	2	4								
7	13	13	11	9	5	2								
8	13	11	5	6	1	1								
9	9	9	4	4	1	1								
10	8	8	4	3	1									
11	9	6	3	2										
12	8	5	2	1										
13	1	4	0	1										
14	4	3	1	1										
15	2	2		1										
16	1	2												
17	2	2												
18	1	1												
19	1	1												
20	0	1												
21	0	1												
22	0	1												
23	1													
24	0													
25	1													
d.f. =	15		10		5		3		2		1		1	
X^2 =	17.06188		5.38955		7.11178		.37150		2.38160		.68571		.68799	
P ~	.30		.85		.20		.95		.30		.40		.40	

TABLE VIII

COMPARISON OF OBSERVED DISTRIBUTIONS AND NEGATIVE BINOMIAL EXPECTED DISTRIBUTIONS BY DIAMETER CLASS FOR DISTRICT 4

Stem Count	Diameter Classes 4		6		8		10		12		14	
	Observed (0) and Expected (E) Frequencies											
	O	E	O	E	O	E	O	E	O	E	O	E
0	19	18	34	37	70	68	113	111	132	133	150	152
1	23	24	47	45	59	65	56	61	54	50	53	47
2	19	26	48	39	43	40	27	24	12	17	3	8
3	36	24	24	30	20	20	8	8	7	6	2	1
4	21	21	21	21	11	9	4	3	2	2		
5	14	18	13	14	3	4		1	1			
6	16	15	5	9	1	2						
7	14	13	7	6	1							
8	7	10	3	3								
9	7	8	2	2								
10	10	7	1	1								
11	3	5	2	1								
12	4	4	0									
13	3	3	0									
14	4	3	0									
15	2	2	0									
16	2	2	1									
17	1	1										
18	0	1										
19	1	1										
20	0	1										
21	0	1										
22	1											
23	1											
d.f. =	14		11		4		2		2		1	
χ^2 =	12.70836		7.54801		1.53211		.82087		2.46477		4.91727	
P ~	.60		.75		.80		.70		.30		.025	

TABLE IX

COMPARISON OF OBSERVED DISTRIBUTIONS AND NEGATIVE BINOMIAL EXPECTED DISTRIBUTIONS BY DIAMETER CLASS FOR DISTRICT 5

Stem Count	Diameter Classes: 4		6		8		10		12		14	
	Observed (O) and Expected (E) Frequencies											
	O	E	O	E	O	E	O	E	O	E	O	E
0	21	24	57	56	95	95	122	122	146	144	171	171
1	34	31	49	49	63	63	58	61	52	58	41	41
2	33	31	34	37	33	33	30	25	20	15	7	8
3	31	28	23	26	12	16	8	9	3	3	2	1
4	25	24	23	18	11	8	2	3		1		
5	14	19	13	12	6	3	0	1				
6	11	15	8	8	1	2	0					
7	14	12	6	5		1	0					
8	10	9	3	3			1					
9	4	7	1	2								
10	5	5	2	1								
11	7	4	0	1								
12	3	3	1	1								
13	0	2	1	1								
14	3	2		1								
15	0	1										
16	3	1										
17	2	1										
18	1	1										
19		1										
d.f. =	12		9		3		3		2		1	
χ^2 =	7.97006		3.61281		2.29167		1.59199		3.31513		1.12500	
P ~	.80		.93		.55		.60		.20		.30	

TABLE X

COMPARISON OF OBSERVED DISTRIBUTIONS AND NEGATIVE BINOMIAL EXPECTED DISTRIBUTIONS BY DIAMETER CLASS FOR DISTRICT 6

Stem Count	Diameter Classes											
	4		6		8		10		12		14	
	Observed (O) and Expected (E) Frequencies											
	O	E	O	E	O	E	O	E	O	E	O	E
0	31	28	40	42	83	82	127	130	162	163	198	198
1	26	31	57	49	72	72	75	69	61	58	35	35
2	30	29	34	43	40	44	26	27	14	14	5	5
3	31	26	35	33	16	23	7	9	0	3	1	1
4	17	22	23	24	21	11	2	3	1	1		
5	22	19	21	17	7	5	1	1	1			
6	12	16	7	11		2	0					
7	10	13	9	7			0					
8	14	11	4	5			1					
9	9	9	4	3								
10	10	7	1	2								
11	5	6	2	1								
12	7	5	2	1								
13	3	4		1								
14	2	3										
15	0	2										
16	5	2										
17	1	2										
18	1	1										
19	0	1										
20	0	1										
21	1	1										
22	1											
23	0											
24	0											
25	0											
26	0											
27	0											
28	1											
d.f. =	15		8		3		3		1		1	
X^2 =	9.66349		6.94844		11.59718		2.40578		2.16131		-0-	
P ~	.85		.55		.01		.50		.15		> .90	

TABLE XI

COMPARISON OF OBSERVED DISTRIBUTIONS AND NEGATIVE BINOMIAL EXPECTED DISTRIBUTIONS BY DIAMETER CLASS FOR THE BLOCK

Stem Count	Diameter Classes													
	4		6		8		10		12		14		16	
	Observed (O) and Expected (E) Frequencies													
	O	E	O	E	O	E	O	E	O	E	O	E	O	E
0	102	99	205	197	418	417	653	653	826	825	974	974	1072	1072
1	130	135	231	238	362	360	372	377	300	304	229	229	163	163
2	145	143	208	215	223	227	160	150	96	90	40	42	16	17
3	151	137	166	173	113	125	47	51	22	25	10	7	2	1
4	113	124	131	131	83	65	16	16	6	6		1		
5	102	109	113	95	32	32	3	5	3	2				
6	80	93	56	67	8	15	0	1		1				
7	83	78	52	46	11	7	0							
8	67	65	30	31	1	3	2							
9	50	54	26	21	2	1								
10	53	44	12	14	1	1								
11	35	35	10	9										
12	31	29	5	6										
13	21	23	2	4										
14	21	18	3	3										
15	11	15	0	2										
16	19	12	3	1										
17	12	9												
18	9	7												
19	3	6												
20	4	4												
21	1	3												
22	3	3												
23	4	2												
24	0	2												
25	1	1												
26	0	1												
27	0	1												
28	1	1												
29	1													
d.f. =	22		16		9		4		3		1		1	
χ^2 =	18.11314		9.81455		11.43453		2.84671		.81384		.59524		1.05882	
P ~	.70		.85		.23		.60		.85		.45		.30	

and the block, by diameter class. There were 44 distributions calculated and tested with chi-square. Only the 14-inch class of district four and the eight-inch class of district six failed to show a good fit. The 42 good fits represent 95.45 percent of the total.

The variation in the strengths of the fits, as reflected by the P values, is more related to the $\hat{k}$ value than to the value of $\bar{x}$. Walker (1970) and others have shown that a k greater than 1.0 indicates only mild contagion and is not as critical in the building of the frequency distributions as a k less than 1.0. Hardwoods generally have k values greater than one, so there is a relatively wide range of $\hat{k}$ values that will meet the requirement of Z_i being within $\pm$ 0.0001 of zero. Depending upon the k_i arrived at in the computational process, the expected distribution will vary in its conformity to the observed. Lest the reader obtain a false impression of the range of acceptable k values from which a satisfactorily fitting expected distribution can be generated, it should be noted that as k approaches zero from above, the range narrows quickly. The deviation is always greater above k than below it. For example, when k is approximately one, a k value 10 percent less or 20 percent greater may yield acceptable results. When k is approximately five, a value 20 percent smaller or 400 percent larger may yield acceptable results (Walker, 1970). Knowledge of the limits of deviation in acceptable k values for a particular area and species should be determined before application is attempted.

Joint Frequency Study

The extent of association between diameter classes is another dimension of forest structure requiring consideration. Some association between size classes may be due to factors such as shade intolerance, seed dispersal or other ecological or statistical factors. The existence of significant associations can be determined by testing two-way contingency tables for the null hypothesis of independence with chi-square. Significant associations may be either positive or negative. Positive association is shown by correlation in the joint presence or absence of two diameter classes on a plot and negative association exists when the presence of one diameter class is related to the absence of another diameter class.

Joint frequency tables were constructed with a computer program for all combinations of two diameter classes from 4-inch through 30-inch diameter on the block level only. The counts (x's) for one diameter class run horizontally across the top of the output page, with counts ranging from 0 to 30. Counts for the second diameter class are listed vertically and range from 0 to 16. The frequency of plots having the joint counts indicated is shown at the intersections of the coordinates

The next step was the construction of two-way contingency tables for each of the joint frequency tables. Headings of 'present', 'absent' and 'total' are placed horizontally for one diameter class, and vertically for the other. Four cells are thus formed. The upper left is designated cell a and the upper right is designated cell b. The lower left and lower right are cell c and cell d, respectively. Cell a contains the number of plots in which both diameter classes are present. Cells b and c contain the number of plots where one class is present and the other class absent. Cell d contains the number of plots where both diameter classes are absent. The entering of data into the table begins with the total plot count, N, which is placed at the coordinate intersection of column 'total' and row 'total'. From the univariate frequency tables the number of empty plots for each diameter class is readily determined and entered to its respective 'absent-total' position. Subtraction gives the respective 'present-total' entries. The value entered in cell d is obtained directly from the joint frequency table. Subtraction then provides the data entries for cells a, b and c.

Expected frequencies are obtained from the general formula:

$$E_y = \frac{(T_r)(T_c)}{T_g}$$

where y = cell designation; T_r = row total; T_c = column total; amd T_g = grand total. The expected frequencies all deviate from the observed frequencies by the same amount. The sign for two of the cell deviations is positive and for the other two, negative. The sum of the deviations of all cells is therefore zero. To test the fit of the observed to the expected, a chi-square is calculated. Since the number of degrees of freedom in the χ^2 test is d.f. = (r - 1)(c - 1), one degree of freedom is available for the 2 x 2 contingency table. Chi-square equals the square of the deviation value times the sum of the reciprocals of the four expected values, as shown below:

$$\chi^2 = (D) \quad \frac{1}{E_a} + \frac{1}{E_b} + \frac{1}{E_c} + \frac{1}{E_d}.$$

The null hypothesis is that the diameter classes exist independently of one another (not associated) and is tested by the chi-square value at the five percent level and one degree of freedom. Since hardwoods in general are able to regenerate under low light conditions (shade tolerance), it is expected that the hypothesis would not be rejected. A probability (P) value greater than 0.05 results in acceptance of the hypothesis and P values less than 0.05 warrant rejecting the hypothesis.

Out of 91 tests made, only 15 had P values less than 0.05 and therefore showed significant association between two diameter class distributions (rejection of the null hypothesis). Association occurred most generally in adjacent size classes and is not surprising. As the results did not greatly differ from the anticipated findings, additional analysis was felt to be unnecessary. For information only, all tables were tested for positive or negative tendency in two ways. First, by the method of Cole (1946) in which

$$\lambda = \frac{N \cdot a}{(a + b)(c + d)}.$$

If λ is greater than one, the association is positive, and if less than one, the association is negative.

The second test is derived from simple analysis of the relationship that must exist between the four cells for a positive or negative association to occur (Walker, 1970). When the product of (a•d) is less than the product of (b•c), the association is negative. If (a•d) is greater than (b•c), the association is positive. This formulates as:

$$\text{Negative Association} = (a \cdot d) < (b \cdot c),$$
$$\text{Positive Association} = (a \cdot d) > (b \cdot c).$$

All 15 of the significant associations were positive. Of the remaining 76 tables, 38 tested positive and 38 tested negative.

Sampling of Negative Binomial Distributions

In subsection one, sufficient evidence that the stem frequency distributions found on the sample plots conform to the negative binomial was presented. For purposes of sampling, the 1253 plots are redefined as a population. The $\bar{x}$, s^2 and $\hat{k}$ values previously determined are now considered the μ, σ^2 and $\underline{k}$ values of the theoretical population. The diameter distributions on all districts and the block were sampled with a computer program. Minimum input data for the program included specification of the number of samples to be drawn from each diameter class (sampling intensity (n/N) expressed as a percent), the number of times that the specified intensity sample is to be drawn for each diameter class and the accumulative frequencies for each diameter class. There is one IBM card (two where the count exceeds 16) for each diameter class, with a district (or block) identifier, the total N and the accumulative frequencies spaced horizontally from left to right. The first frequency given is for the zero count, second frequency is the one count, et cetera.

The computer derives and stores the relative frequencies for each count. Beginning from a random start, the frequencies are sampled on a probability basis until the desired number of samples have been drawn. The $\bar{x}$, s^2 and $\hat{k}$ values are calculated and stored. Sampling continues in this manner until the specified number of samples has been drawn. Once the final sample has been drawn, the output is prepared. The output consists of (1) a list of the sample $\bar{x}$, s^2 and $\hat{k}$ values, (2) a distribution showing the number of sample means falling within plus and minus three standard deviations of the mean of the means. The distribution range is given in 19 value points. The mid-point is the value of the mean of the means and each of the value points extending outward are plus or minus one-third of one standard deviation, (3) a distribution, as described in (2) above, for the sample variances, (4) a distribution, as described in (2) above, for the sample maximum likelihood estimates of k and (5) under each of the above distributions, the mean, variance, standard deviation and the maximum and minimum sample values encountered.

When a sample had a mean equal to the variance, indicative of a

Poisson distribution, the sample was rejected and another sample was drawn. The number of rejects was printed out for information purposes. As expected, this situation arose most frequently in the larger diameter classes where the counts were generally 0, 1, 2 and 3. In such cases, the majority of frequencies were in the zero count, and the distributions had means of less than one.

After testing, it was decided that 250 samples of a diameter class at any given intensity would sufficiently cover the variability of the class. The sampling intensities selected for evaluation were ten and twenty percent. The output data permitted examination of several results including (1) how the sample values ($\bar{x}$, s^2, $\hat{k}$) are distributed and how they correspond to the actual values (μ, σ^2, $\underline{k}$), (2) how the distribution means relate to the actual values, (3) how the district sample values relate to the actual block values and (4) how the three statistics are related to each other.

The distribution of means was approximately normal regardless of diameter class or size of the population or intensity of sampling. The distribution of variances also resembled the normal, but with a slight positive skew. The distributions of variances apparently were not affected by the diameter class, population size or sampling intensity. The sample $\hat{k}$ values showed no uniform predictive distribution tendencies. Their distributions ranged from multimodal dispersal to unimodal clustering about the mean. There was no apparent correlation between the distribution of the $\hat{k}$ values and the three factors described above.

As expected, the district means of the distributions of means and variances satisfactorily approximated the actual district and block means and variances. The block means of the distributions of means and variances were excellent approximations of the actual block means and variances. All estimates of k were sufficiently poor to exclude the possibility of low intensity sampling for that parameter, considering the relatively large effect on calculated distributions of relatively small changes in k values (Walker, 1970). Consequently, emphasis was directed to the other distribution parameter, the mean.

The basic premise under which this portion of the study was done, is that a good estimate of k for a large geographical area will combine with a low intensity estimate of the mean of a portion of the total area, and provide a satisfactory estimate of the actual frequency distribution of stems on the small area. Sampling for the mean of a population requires consideration of two separate facets. First is the effect of variation of $\bar{x}$ values on the adequacy-of-fit of the resultant distributions. Second, the degree of confidence that any single $\bar{x}$ value will fall within acceptable limits must be known, since multiple sampling is not practical in field work. For determining the acceptable variation in the $\bar{x}$ value, the original expected distributions generated from the μ's and k's of the districts and the block are accepted as the observed distributions. The value of the mean is then changed and new expected distributions are generated. The adequacy-of-fit of the expected distribution to the observed is determined by calculating the chi-square and finding the attendant probability. The tails of the distributions were collapsed to a minimum expected frequency of three, except in the larger diameter classes where the total distribution length consisted of only four frequency values. In such cases, collapsing the tails would reduce the number of degrees of freedom to zero and thereby invalidate the chi-square test. Therefore, the minimum expectation rule was suspended in these instances, since the allowable deviation of the mean is relative to a specific diameter class. In addition, the contribution of the larger diameter classes to the total volume of the area is particularly significant and worthy of consideration.

The initial variation in the value of the mean selected for evaluation was plus and minus 10 percent of the mean. From that point, arriving at the percents of positive and negative variation that resulted in an inadequately fitting distribution, was a trial and error procedure. Table XII shows the effect of variations in the mean for the six-inch diameter class of the block and Table XIII shows the effect for the 14-inch diameter class of the block.

For the six-inch diameter class, an adequate fit between the observed observed distribution and expected distribution is maintained when the $\bar{x}$ value does not exceed the value of the true mean by more than 13 percent, or does not fall below the mean by more than 10 percent. The plus and minus percentage deviations that are allowable for the 14-inch diameter class are approximately 12 and 10 percent, respectively. It should be kept in mind that values have been rounded, so that the exact extent of the effects are not measured in this work. The effects of increasing the value of the mean are (1) the frequencies of the counts less than the mean are decreased, (2) the frequencies of the counts larger than the mean are increased and (3) the tail of the distribution is extended. A decrease in the value of the same magnitude will reverse the above effects to the same degree. Decreases of the same magnitude of an increase result, however, in significantly poorer fits with the observed distribution. This is primarily due to the shortening of the distribution tail and the accompanying reduction in the number of degrees

TABLE XII

DISTRIBUTION COMPARISON SHOWING THE EFFECTS OF VARIATIONS IN THE MEAN

(Six-inch DBH class of the Block)
μ = 3.0439
$\underline{k}$ = 2.0015
N = 1253

	Percent Variation												
Count	+15%	+14%	+13%	+12%	+11%	+10%	μ	-10%	-11%	-12%	-13%	-14%	-15%
0	166	168	169	171	173	175	197	223	226	229	232	235	238
1	211	213	214	216	218	219	238	258	260	262	264	267	269
2	201	202	203	204	205	206	215	224	224	225	226	227	227
3	171	171	171	172	172	172	173	172	172	172	172	171	171
4	136	136	136	135	135	135	131	125	124	123	122	121	121
5	104	103	103	102	102	101	95	86	86	85	84	83	82
6	77	76	76	75	75	74	67	58	57	56	56	55	54
7	56	55	55	54	53	53	46	39	38	37	36	35	35
8	40	40	39	38	38	37	31	25	24	24	23	23	22
9	28	28	28	27	26	26	21	16	16	15	15	14	14
10	20	20	19	19	18	18	14	10	10	10	9	9	9
11	14	14	13	13	12	12	9	6	6	6	6	6	5
12	9	9	9	9	8	8	6	4	4	4	4	3	3
13	7	6	6	6	6	6	4	3	2	2	2	2	2
14	4	4	4	4	4	4	2	2	2	1	1	1	1
15	3	3	3	3	3	3	2	1	1	1	1	1	
16	2	2	2	2	2	2	1	1	1	1			
17	1	1	1	1	1	1	1						
18	1	1	1	1	1								
19	1	1	1	1	1								
20	1												
d.f.=	15	14	14	14	14	14		12	11	11	11	11	11
χ^2 =	28.6	24.3	22.2	19.4	16.2	14.7		17.3	21.1	26.7	34.9	41.2	53.3
P ~	.02	.04	.075	.175	.30	.43		.13	.035	.005	.000	.000	.000

of freedom. Therefore, errors in the estimation of the mean in a positive direction are less serious than the same degree of error in a negative direction. This fact may be concealed in the larger diameter classes since the relative change in the zero count is less significant and the distribution length may not be changed. Consequently, the number of degrees of freedom do not change and the higher chi-square values accompanying the decreases are generally not sufficient to change the probability in the chi-square table.

TABLE XIII

DISTRIBUTION COMPARISON SHOWING THE EFFECTS OF VARIATIONS IN THE MEAN

(14-inch Diameter Class of the Block)
μ = 0.2706
$\underline{k}$ = 1.8159
$\underline{N}$ = 1253

	Percent Variation												
Count	+15%	+14%	+13%	+12%	+11%	+10%	μ	-10%	-11%	-12%	-13%	-14%	-15%
0	940	942	945	947	949	951	974	997	999	1001	1004	1006	1009
1	250	249	247	246	244	243	229	214	213	211	209	208	206
2	51	51	50	49	49	48	42	36	35	35	34	33	32
3	10	9	9	9	9	9	7	5	5	5	5	5	5
4	2	2	2	2	2	2	1	1	1	1	1	1	1
d.f.=	1	1	1	1	1	1		1	1	1	1	1	1
χ^2 =	5.92	5.10	4.30	3.76	3.40	2.93		3.25	3.89	4.33	5.36	6.26	7.57
P ~	.017	.024	.040	.054	.069	.090		.076	.050	.040	.021	.013	.008

The maximum positive and negative deviations from the value of the mean that still permits generation of distributions that adequately fit the observed distributions are given in Table XIV. The fact that the block permits less tolerable error than the districts is expected since the larger the N, the less tendency for the distribution-of-means curve to flatten out.

TABLE XIV

ALLOWABLE ERROR IN ESTIMATING THE VALUE OF THE MEAN FOR ALL DIAMETER CLASSES ON ALL DISTRICTS AND THE BLOCK

Area	Diameter Class 4		6		8		10		12		14	
	+	-	+	-	+	-	+	-	+	-	+	-
District 1	37	18	28	16	25	17	22	13	34	18	38	17*
District 2	43	22	33	18	27	17	28	17	37	18	35*	--
District 3	40	19	34	17	39	22	29	16	28	17	33	24*
District 4	41	18	31	18	23	17	30	15	31	19	27*	--
District 5	38	19	39	20	31	18	30	19	24	15	29	24*
District 6	48	22	32	18	24	18	27	16	25	15	34*	--
Block	16	12	13	10	12	9	11	9	14	10	12	10

-- = insufficient entries to provide a degree of freedom for testing
* = minimum expectation rule of at least three entries suspended

Now that there is an indication of the allowable error in the estimation of the mean that will still permit calculation of an adequately fitting expected distribution, attention is centered on the confidence that any single estimate of the mean will fall within the acceptable deviation. The statistical probability is expressed in the following equation: $\zeta = P(\mu - \alpha\mu < \bar{x} < \mu + \beta\mu)$, where α equals the percent negative deviation (expressed as a decimal) and β equals the percent positive deviation (expressed as a decimal). The α and β values were taken from Table XIV, for this portion of the study. The $\alpha\mu$ and $\beta\mu$ values are individually divided by the standard error of the mean ($\sigma_{\bar{x}}$). The resultant values are termed z and are located in the normal distribution. The attendant probability values are found from a table of the normal distribution and are summed to obtain the probability of the mean being within the desired deviation range. The confidence values are shown in Table XV for a sampling intensity of 10 percent and in Table XVI for a sampling intensity of 20 percent. The confidence values reflect the confidence of obtaining an adequately fitting negative binomial distribution 95 percent of the time, as well as the confidence of obtaining sample means that fall within the acceptable limits. The values in the two tables have been rounded to the nearest one-tenth of one percent for convenience.

The final aspect of sampling to be examined is the determination of the sample size required, by diameter class, when the confidence level is held constant at 90 percent. The sample size (n) is found by solving

$$n = \frac{(1.645)^2 \ \sigma^2}{\alpha^2 \ \mu^2}$$

where α, μ and σ are values previously defined and 1.645 is the 90 percent limit for the normal distribution (Remington and Schork, 1970). Using the information in Table XV the sample sizes were calculated for all districts and the block. Since the population sizes vary, expressing the results in terms of percents was done to simplify comparison. Table XVII contains the sampling intensities for a 90 percent confidence level.

TABLE XV

THE PROBABILITY OF AN ESTIMATE OF THE MEAN OBTAINED FROM ANY SINGLE TEN PERCENT SAMPLE FALLING WITHIN THE ALLOWABLE DEVIATION FROM THE TRUE MEAN

	Diameter Class					
Area	4	6	8	10	12	14
D-1	0.8448	0.7805	0.6788	0.4902	0.5128	0.4371
D-2	0.8923	0.7910	0.6646	0.5520	0.5242	
D-3	0.8495	0.7856	0.7088	0.5469	0.4495	0.4595
D-4	0.8174	0.7383	0.6270	0.5434	0.5066	
D-5	0.8197	0.7772	0.6441	0.5688	0.4378	0.4565
D-6	0.8708	0.7880	0.6703	0.5299	0.4270	
Block	0.9288	0.8427	0.7209	0.5999	0.5817	0.4506

TABLE XVI

THE PROBABILITY OF AN ESTIMATE OF THE MEAN OBTAINED FROM ANY TWENTY PERCENT SAMPLE FALLING WITHIN THE ALLOWABLE DEVIATION FROM THE TRUE MEAN

	Diameter Class					
Area	4	6	8	10	12	14
D-1	0.9357	0.9010	0.8319	0.6419	0.6630	0.5751
D-2	0.9641	0.9052	0.8165	0.7081	0.6718	
D-3	0.9366	0.8969	0.8476	0.7001	0.5965	0.6110
D-4	0.9118	0.8718	0.7882	0.6916	0.6609	
D-5	0.9196	0.8931	0.7957	0.7268	0.5832	0.6091
D-6	0.9490	0.9050	0.8278	0.6844	0.5701	
Block	0.9871	0.9508	0.8695	0.7627	0.7435	0.6016

TABLE XVII

PERCENTAGE SAMPLING INTENSITIES BY DIAMETER CLASS AND AREA FOR NINETY PERCENT CONFIDENCE THAT A SAMPLE MEAN WILL FALL WITHIN PLUS AND MINUS THE MINIMUM ACCEPTABLE DEVIATION

Area	Diameter Class					
	4	6	8	10	12	14
District 1	23.23	30.01	40.46	109.46	111.51	201.03
District 2	16.66	29.61	46.47	79.60	116.32	
District 3	23.04	32.43	42.50	90.44	130.10	100.97
District 4	29.33	36.19	46.36	103.01	97.07	
District 5	26.90	33.09	54.98	70.30	133.64	88.99
District 6	20.19	29.30	38.01	90.50	148.17	
Block	10.56	17.28	30.69	46.85	58.39	91.22

It should be noted that the sampling intensities shown are the maximum that would be required. This is true because the minimum confidence deviations allowable were used in the calculations.

The larger the sample size (intensity), the smaller the deviation and vice versa. The optimum sampling intensity falls between the values obtained through the use of α or β alone. It is the result of using some factor of $(\alpha - \beta)$ in the equation that determined $\underline{n}$. It was felt that the use of α alone provided a satisfactory indication of the sampling intensity required and the complexity of solving for the optimum intensity was not justified. It is sufficient to note that the optimum intensities are less than the values given in Table XVII.

The unrealistically high sampling intensities indicated for the 10-inch and larger diameter classes can be reduced by increasing the plot size to obtain a better estimate of the variances. Although this aspect was beyond the scope of this research effort, it is suggested that the use of different basal-area-factor prisms to obtain variable radius plots might be an improvement in sampling technique. The differences between the block and any of the districts sampling intensities point up the significant effect of increases in the sample size.

The preceding material has attempted to show that an estimate of the mean frequency of stems on a small subdivision of a large area, obtained through relatively low intensity sampling, will combine with the established $\hat{k}$ of the large area and the subdivision n, to generate a stem frequency distribution according to the negative binomial which reflects the actual frequency distribution of stems existing in the subdivision.

Comments on the Maximum Likelihood Value of k Obtained from the Moment Estimate

The calculation of the maximum likelihood estimate of k of the negative binomial frequency distribution from the moment estimate of k involves an iterative procedure (previously described) that is very laborious when done by hand and can be very costly when using a computer. The major problem is that the distance between the $\tilde{k}$ and the $\hat{k}$ values is not predictable. Only general trends are observable regarding the relationship.

When $\hat{k}$ is a small positive value, it will most closely approximate the $\hat{k}$. As $\tilde{k}$ increases, the distance between it and $\hat{k}$ also increases. Unfortunately, the relationship does not appear to be linear. The size of the $\bar{x}$ and s^2 values affect the relationship.

When $\tilde{k}$ is negative, regardless of its magnitude, $\hat{k}$ is always a very large positive value. The relationship here is even less predictable than for positive $\tilde{k}$ values. It appears that while increases in the value of a negative $\tilde{k}$ result in an increase in the value of $\hat{k}$, increases in $\bar{x}$ and s^2 values result in a decrease in the value of $\hat{k}$. Due to time limitations, a detailed study was not possible.

Figure 1 shows in a very general fashion, the procedural flow for the computer calculation of $\hat{k}$. It is purposely given in broad terms so that it might be more adaptable to the specific computer facilities available to the reader. Two facets deserve specific comment as they were derived directly from this research, and are not believed to have previously been used. The first is the variable incrementation (VINC) and the second is the negative moment estimate.

Increment, in this thesis, refers to either a positive or negative change in the iterative value of $\tilde{k}$. As originally designed, the program had a constant increment of 0.1. This is satisfactory as long as $\tilde{k}$ is positive and the $\hat{k}$ is relatively small (less than 2.0). Walker (1970) has shown that these conditions are most often met by pine forests. Pine-hardwood and hardwood forests generally have $\underline{k}$ values greater than 2.0. This resulted in an excessive number of iterations in obtaining the $\hat{k}$ value. Variable incrementation significantly reduced the number of iterations and hence the cost. It is most profitable when $\tilde{k}$ is negative, since the distance separating $\tilde{k}$ and $\hat{k}$ is extremely large.

The original program was designed to immediately convert a negative $\tilde{k}$ to its absolute value. This was improved upon by establishing a large positive value as equal to any negative $\tilde{k}$. Trial and error showed that for the hardwood population under consideration, a value of 70.0 was best. The actual optimum values for incrementation and negative $\tilde{k}$ conditions will have to be determined for the different types of forest

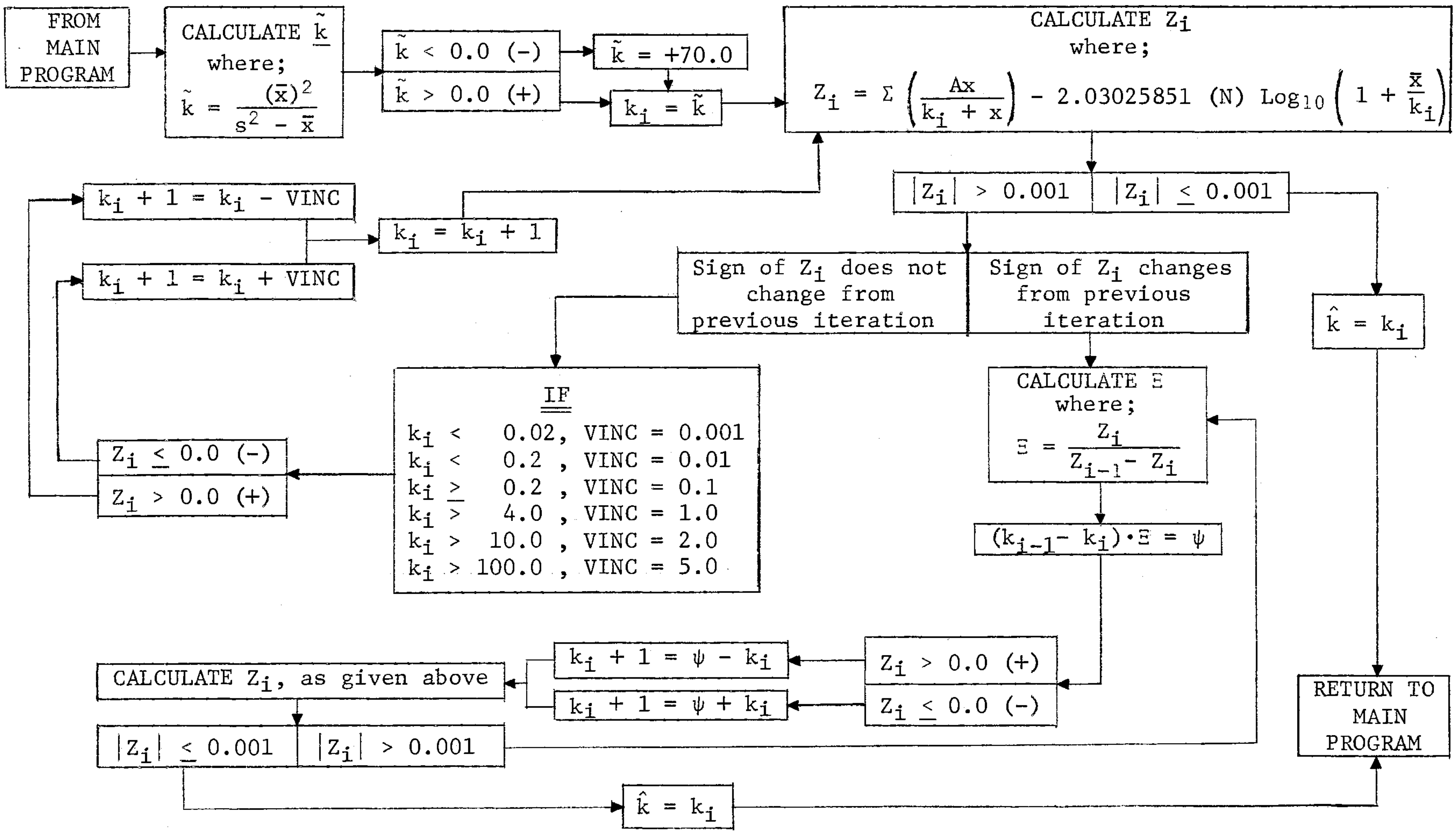

Figure 1. Generalized Flow Chart for Calculation of the Maximum Likelihood Estimate of k.

structures individually. This section is intended only to show how some efficiencies were achieved in this study and as a directional guide, not as hard and fast rules.

CHAPTER V

SUMMARY

The objectives of this thesis were to examine (1) the fit of a hardwood population drawn from a pine-hardwood forest to a mathematical model generated from the negative binomial series, (2) the association that exists between diameter classes, (3) the sampling intensity and confidence limit estimation possibilities and (4) the possibility for programming efficiently the calculation of the maximum likelihood estimate of the parameter $\underline{k}$ of the negative binomial series.

Comparison of the frequency distribution of stems by diameter class with expected negative binomial frequency distributions showed sufficiently strong correlation to suggest that the frequency of the hardwood stems over an area of pine-hardwood forest is predictable.

Obviously, for any predictive frequency distribution to be valid, the parameters must be accurately estimated. Samples of various sizes were drawn from the population segments called districts. Frequency distributions that were built according to the negative binomial series using the sample level $\bar{x}$, district level $\underline{N}$ and block $\hat{k}$, compared favorably with the actual district population although a definitive measurement of the goodness-of-fit was not obtainable since the statistics were drawn from different population levels. The confidence that could be placed in a relatively low intensity sample $\bar{x}$ of providing an acceptable distribution was examined at both 10 and 20 percent intensity, for the 4-inch through 14-inch diameter classes. The confidence level was then held constant and the various sampling intensities required to meet the objective were studied. The results indicate that for the smallest diameter classes, samples from small plots will be adequate. For the larger classes (10-inch and up), an increase in plot size will probably be necessary.

Joint frequency tables were constructed in order to evaluate the association between all combinations of two diameter classes. Significant associations were noted in 15 of 91 tests. These were all positive associations and generally occurred in adjacent diameter classes. As the nature of hardwoods lends itself to the expectation of a minor degree of association, the matter was not pursued further.

Some comments of the calculation of the maximum likelihood estimate of $\underline{k}$ have been given for informational purposes only. The maximum likelihood estimate of $\underline{k}$ is critical to the calculated negative binomial distribution and it is hoped that the discussion will provide a smoother path for others working with the negative binomial series.

The negative binomial has shown good signs of becoming a valuable forest management tool. Properly applied, it may be able to provide a more accurate analysis of forest structure with a reduction of sampling time in the field and thereby effect a savings in management costs.

SELECTED BIBLIOGRAPHY

Anscombe, F. J.

1950 "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions." Biometrika, 37, 358-382.

Archibald, E. E. A.

1948 "Plant Populations, I. A New Application of Neyman's Contagious Distribution." Ann. Bot. (NS), 12, 221-235.

Bliss, C. I.

1953 "Fitting the Negative Binomial Distribution to Biological Data." Biometrics, 9, 176-196.

Clapham, A. R.

1936 "Over-Dispersion in Grassland Communities and the Use of Statistical Methods in Plant Ecology." J. Ecology, 24, 232-251.

Cole, L. C.

1946 "A Theory for Analyzing Contagiously Distributed Populations." Ecology, 27, 329-341.

Dahlem, M. J.

1972 "Low Intensity Forest Sampling Through Use of Stem Frequency Distribution and Population Parameters." (Unpub. M. S. thesis, Oklahoma State University.)

Fisher, R. A.

1941 "The Negative Binomial Distribution." Ann. Eugenics, 11, 182-187.

1953 "Note on the Efficient Fitting of the Negative Binomial." Biometrics, 9, 197-200.

Gerrard, D. J., and R. D. Cook.

1972 "Inverse Binomial Sampling as a Basis for Estimating Negative Binomial Population Densities." Biometrics, 28, 971-980.

Haldane, J. B. S.

1945 "On a Method of Estimating Frequencies." Biometrika, 33, 222-225.

Ker, J. W.

1954 "Distribution Series Arising in Quadrat Sampling of Reproduction." J. For., 52, 838-841.

Neyman, J.

1939 "On a New Class of 'Contagious' Distributions, Applicable in Entomology and Bacteriology." Ann. Math. Stat., 10, 35-57.

Pielou, E. C.

1960 "A Single Mechanism to Account for Regular, Random, and Aggregated Populations." J. Ecology, 48, 575-584.

Quenouille, M. H.

1949 "A Relation Between the Logarithmic, Poisson, and Negative Binomial Series." Biometrics, 5, 162-164.

Remington, R. D., and M. A. Schork.

1970 Statistics with Applications to the Biological and Health Sciences. New Jersey: Prentice-Hall, Inc.

Robinson, P.

1954 "The Distribution of Plant Populations." Ann. Bot. (NS), 18, 35-45.

Walker, N.

1970 "Tree Stem Frequency Distribution and Size Class Association in Natural Forest Types." (Unpub. Ph.D. thesis, N. Carolina State University.)

Waters, W. E., and W. R. Henson.

1959 "Some Sampling Attributes of the Negative Binomial Distribution with Special Reference to Forest Insects." For. Sci., 3, 321-328.

VITA

William Wycoff Thomas

Candidate for the Degree of

Master of Science

Thesis: STEM FREQUENCY DISTRIBUTION AND SAMPLING OF THE HARDWOOD COMPONENT OF PINE-HARDWOOD STANDS IN SOUTHEASTERN OKLAHOMA

Major Field: Forest Resources

Biographical:

Personal Data: Born in Oklahoma City, Oklahoma, May 4, 1942, the son of Mr. and Mrs. John W. Thomas.

Education: Graduated from Harding High School, Oklahoma City, Oklahoma, in May, 1960. Attended Westminster College, Fulton, Missouri, Oklahoma City University, Oklahoma City, Oklahoma before receiving Bachelor of Science Degree in General Business from Oklahoma State University in May, 1965. Completed the requirements for the Master of Science Degree at Oklahoma State University in December, 1973.

Professional Experience: Graduate Research Assistant, Oklahoma State University, 1971-73.