156
68

Survival Equations for Loblolly Pine Trees in Cutover, Site-prepared Plantations

by

Olga B. Avila

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE
in
Forestry

APPROVED:

Harold E Burkhart
H.E. Burkhart, Chairman

Timothy G Gregoire
T.G. Gregoire

Jesse A. Logan
J.A. Logan

August, 1990
Blacksburg, Virginia

Survival Equations for Loblolly Pine Trees in Cutover, Site-prepared Plantations

by

Olga B. Avila

H.E. Burkhart, Chairman

Forestry

(ABSTRACT)

The probability of mortality for an individual tree with certain characteristics growing under certain conditions was modeled. A particular algorithm SCREEN was used to find the best set of predictors variables. This algorithm was specially created to be used when the dependent variable can take only two values like in this binary case (dead or alive tree). The logistic model with different independent variables, which were found to be significant through the SCREEN algorithm, was fitted to the data.

For the unthinned plots the logistic model with the following variables, CR (crown ratio), HH (total height/height of dominant and codominant trees) and CI (competition index) was compared with the survival model applied in a published distance-dependent model PTAEDA. The logistic model with CR, HH and DD (quadratic mean diameter/dbh) was compared with the survival model already used in a distance-independent model TRULOB. In both cases the behavior of the logistic model was quite similar to the published models.

For the thinned plots the predictor variables DD, HH, CI and CR raised to 1.5 were used in the logistic model to predict mortality for individual trees.

Mortality is difficult to predict. In this particular study the logistic model was used. The final distance-dependent model for unthinned plots includes as predictor variables CR, HH and CI. For thinned plots the final logistic model employs HH, CI and CR raised to 1.5 as independent variables. The final distance-independent model for unthinned plots includes as predictor variables HH,

DD and CR. For thinned plots the final logistic model uses HH, DD and CR raised to 1.5 as independent variables.

Differences between deterministic and the stochastic treatments of mortality were also studied. No practical differences in several stand characteristics such as average height, total volume, basal area were found when using these two approaches. Further, no significant differences were found in the diameter distribution for dead or alive trees.

Acknowledgements

My deepest gratitude goes to my advisor, Dr. Harold E. Burkhart, for his support, guidance and encouragement throughout this study. Special thanks are extended as well to Dr. Timothy G. Gregoire and Dr. Jesse A. Logan for their valuable suggestions and comments. I would like to express thanks to Ralph L. Amateis for his assistance.

I also wish to thank supporters of the Loblolly Pine Growth and Yield Research Cooperative at VPI and SU for supplying the data and financial support for this study.

Very special thanks to my husband, Ruben, whose encouragement and love were always present during both good and bad times.

To my parents, brother and sister my appreciation for their continuous support.

Thanks are also given to my officemates and all the Biometrics Section. Their friendship, advice and technical assistance were very important during my stay at Virginia Tech.

Table of Contents

List of Illustrations

List of Tables

Chapter I

Introduction

Knowing an individual tree's likelihood of dying within the next growing season is essential to a tree-level forest growth simulator (Buchman, 1979). Mortality is one of the important components in a growth and yield model (Buchman et al., 1983). Although anyone who wants to develop a growth and yield model must deal with the problem of mortality, it remains one of the least understood components of growth and yield simulation models (Hamilton, 1986). As Glover and Hool (1979) mentioned, mortality is one of the most difficult and least reliable aspects of evaluating tree growth. They also said that mortality will never be well predicted because of the complexity and uncertainty of forest ecosystems. Dudek and Ek (1980) provide a thorough review of past studies in mortality of trees.

There are two types of mortality, regular and irregular (Lee, 1971, Yang, 1988). As Lee pointed out "Regular mortality is found when trees grow so close together that they must compete for survival". This type of competition for limited resources such as light, moisture and nutrients is most intensive in overstocked stands, and non-existent in fully open-grown trees. When trees grow under highly competitive conditions, the suppressed trees eventually die. Those trees that are best adapted to the environment conditions are likely to survive. This probability of survival increases with size and age until the tree is mature. The other type of mortality caused by insect, disease, windfall, fire or other damage is known as "irregular mortality". The last type of mortality is very difficult to predict. Typically only the non-catastrophic (regular) mortality is considered (Monserud,1975,1976) or a specific type of irregular mortality (e.g. Stage's (1973) model for mountain pine beetle (*Dendroctonus ponderosae* Hopk.) induced mortality in lodgepole pine (*Pinus contorta* Dougl.)).

There are different types of growth projection systems. They may be classified into several classes. Among these are individual tree models (either distance-dependent or distance-independent models according to wether the inter-tree distance information is available or not), stand growth models and diameter distribution models. According to the type of growth projection systems used, mortality is treated in a different way. Diameter distribution models may use either an explicit or implicit mortality function to control the number of trees per diameter class over time. Stand growth models may use stand mortality models that predict overall reduction in trees per acre or by diameter class. The individual-tree modeling approach requires mortality functions that predict the probability of survival for each tree in the stand (Hamilton, 1980). No matter which type of growth projection system is used in trying to find the equation that explains the probability of survival, the first step is to screen potential independent variables for those that best explain variations in mortality rate (Hamilton and Edwards, 1976). In this particular case the dependent variable (probability of survival) is a discrete variable that only can take two different values (0,1). Thus, this is the simplest of the qualitative response models (dichotomous dependent variable).

Qualitative response models are regression models in which the dependent variable is categorical, i.e., it can assume only a limited number of discrete values. These models calculate the probabilities that an individual will take a particular value among the set of alternatives. As Nerlove and Press (1973) point out, in these models the important aspect is the probability that the response takes on a particular value and not the value itself that this response can take. In this study the qualitative response model is binary, because the dependent variable can take only two values (dead or alive tree).There are several examples in forestry research that deal with these kinds of models. Some of them are developed by Reed et al. (1981), Strub et al. (1986), and Daniels et al. (1979). But, mortality prediction is the most common application of qualitative response models in forestry research.

In the mortality (or survival) model, as the true probability of an individual tree within a certain interval is bounded by 0 and 1, it would be desirable to find a function to model mortality that provides estimates of probability in the interval (0,1) (Hamilton, 1974). The logistic function has been suggested as the preferred model to express the relationship between a 0,1 dependent variable and a set of independent variables (Neter et al., 1970).

An important aspect in the mortality prediction problem is the dichotomy between stochastic and deterministic models. Deterministic models predict the probability that a certain tree with certain characteristics will live. The tree frequency can then be adjusted by multiplying the original frequency times the predicted probability of survival. In stochastic models this probability is compared with a random number generated from a uniform random variate between 0 and 1; if the predicted probability that the tree will live is less than the random number, the tree is considered to have died. On the other hand, if this probability is greater than or equal to the random number the tree will be assumed to stay alive.

When considering various types of thinning, it is advisable to recognize their impact on mortality. There are different questions that still remain under study concerning this aspect, such as: What is the immediate impact of thinning on mortality rate?, What is the duration of this impact?, What is the relationship between intensity of thinning and mortality rate?.

Hamilton (1986) suggested many variables that may account for the impact of thinning in the stand. Variables such as time since thinning, basal area prior to thinning, residual basal area, residual basal area as a percentage of basal area prior to thinning and thinning method can be considered as descriptors of thinning.

Objectives

The main purpose of this study was to model the probability of survival of trees in unthinned and thinned cutover, site-prepared plantations of loblolly pine (*Pinus taeda* L.). Specific objectives were to:

1. Model the probability of survival of unthinned trees in cutover, site-prepared plantations.

2. Validate the model obtained above.

3. Determine if stochastic and deterministic approaches to mortality prediction give similar results.

4. Model survival for trees in thinned cutover, site-prepared plantations.

Chapter II

Literature review

There are, in general, three basic steps when dealing with the development of a mortality model as it is described in Hamilton and Edwards (1976).

- The first step is to screen for the best set of independent variables that explains the probability of survival.
- The second step establishes that once the independent variables have been selected it is necessary to estimate the parameters of the mortality model.
- The third step requires the validation or verification of the model.

How to find the best set of predictors variables

The screening algorithm (Hamilton and Wendt, 1975) SCREEN is a procedure that allows selection of the best predictors independent variables given a dichotomous dependent variable. The algorithm is discussed by Gleser and Collen (1972) and the theory behind the algorithm was presented by Sterling et al. (1969). An updated FORTRAN version of this algorithm was presented by Hamilton (1988).

The algorithm consists of two procedures: SEARCH and GRAPH. The SEARCH procedure screens the data for relations between the proportion of the two possible outcomes of the dependent variable and the explanatory independent variables. The GRAPH procedure prints the results of this screening in a format similar to a decision "tree".

Figure 1 is a diagram of a decision "tree". The squares are referred to as nodes. At each node, SEARCH determines the most significant independent variable at a certain significance level given by the user. SCREEN selects the set of independent variables that best explain variations in mortality rate by first identifying the single best predictor of mortality. Given this variable to be used as a predictor, the variable that explains the most significant amount of the remaining variation in mortality is identified. This procedure is repeated until no further significant predictors are found.

As an example, Figure 2 shows output from Hamilton and Wend (1975) that resulted from analyzing a population of western white pine that had been measured at 5 year intervals for 20 years (1941-61). This example is carried out by these authors in order to explain all the steps of the SCREEN algorithm. The new version of the algorithm in FORTRAN (Hamilton, 1988) allows easy access to the program.

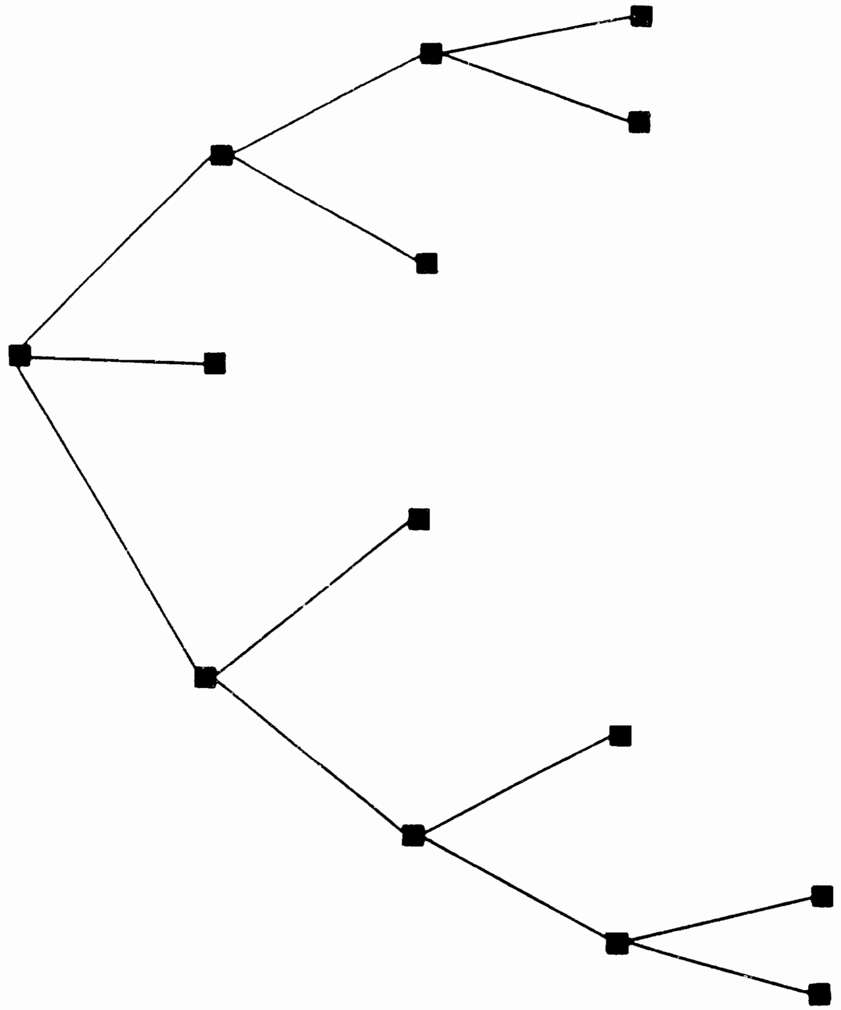
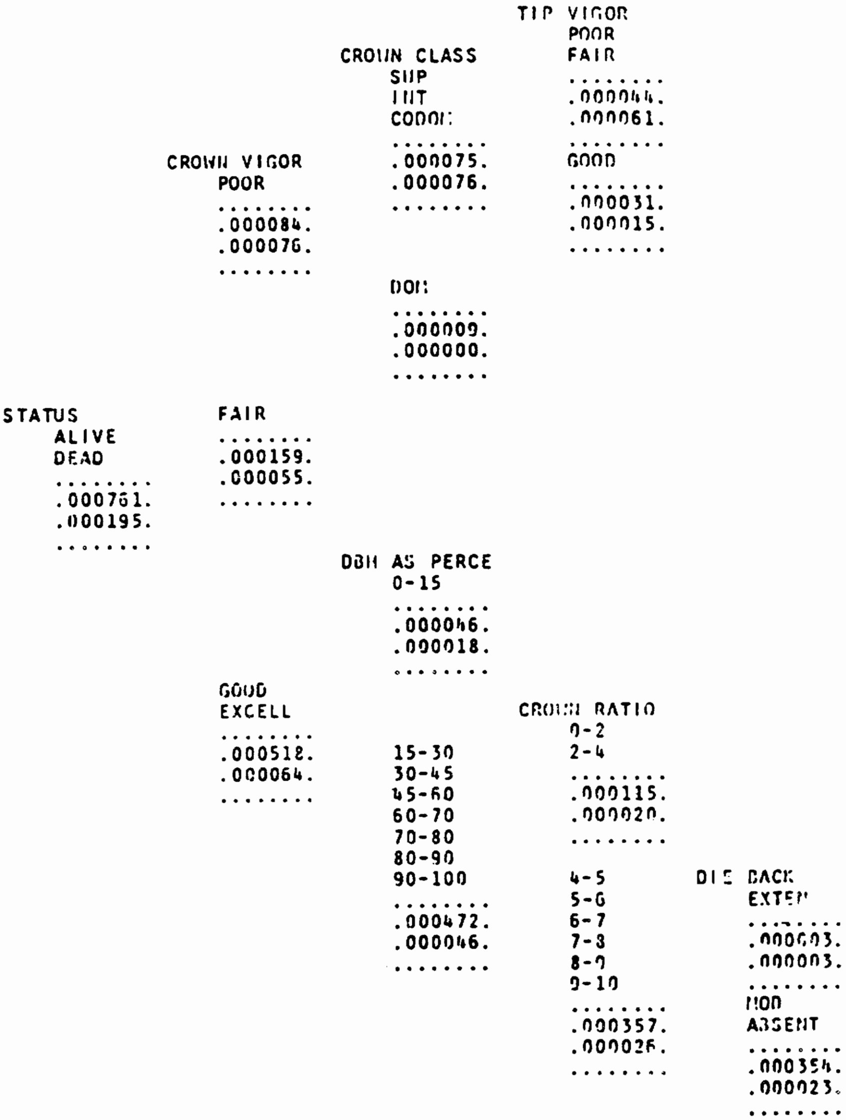


Figure 1. Decision "tree" from GRAPH procedure.



```
                                          TIP VIGOR
                                              POOR
                           CROWN CLASS        FAIR
                               SUP            ........
                               INT            .000044.
                               CODOM          .000061.
                               ........       ........
             CROWN VIGOR       .000075.       GOOD
                 POOR          .000076.       ........
                 ........      ........       .000031.
                 .000084.                     .000015.
                 .000076.                     ........
                 ........
                               DOM
                               ........
                               .000009.
                               .000000.
                               ........
STATUS           FAIR
    ALIVE        ........
    DEAD         .000159.
    ........     .000055.
    .000761.     ........
    .000195.
    ........
                           DBH AS PERCE
                               0-15
                               ........
                               .000046.
                               .000018.
                               ........
                 GOOD
                 EXCELL                   CROWN RATIO
                                              0-2
                 ........                     2-4
                 .000518.      15-30          ........
                 .000064.      30-45          .000115.
                 ........      45-60          .000020.
                               60-70          ........
                               70-80
                               80-90
                               90-100         4-5        DIE BACK
                               ........       5-6            EXTEN
                               .000472.       6-7            ........
                               .000046.       7-8            .000003.
                               ........       8-9            .000003.
                                              9-10           ........
                                              ........       MOD
                                              .000357.       ABSENT
                                              .000026.       ........
                                              ........       .000354.
                                                             .000023.
                                                             ........
```

Figure 2. Output "tree" produced by GRAPH.

If an output as shown in this figure is obtained, this will mean that the first best predictor of mortality rate is crown vigor; for the trees with poor crown vigor the next best predictor variable is crown class. From this output, there are no significant differences among suppressed, intermediate and codominant crown classes; but they are different from the dominant class. For the dominant class there are no further explanatory variables. For those trees in the suppressed, intermediate and codominant crown classes, the mortality is further explained by the tip vigor.

SEARCH also provides ideas related with interactions between independent variables, as it is shown in the output in Figure 2. For trees with poor crown vigor, crown class is the next most significant predictor of mortality. For those trees with good or excellent crown vigor, dbh is more significant than crown class. This can be explained by a significant interaction between crown class and crown vigor and between dbh and crown vigor.

The results of a screen of relationships between dependent and independent variables in this particular algorithm do not assume any specific functional relationship between the dependent and the independent variables.

When using SCREEN, the SEARCH part of the algorithm provides the user with an optimal set of independent variables for describing the behavior of the dependent variable, but the model relating those variables must be constructed by the user.

In order to use SCREEN the independent variables must be either discrete with no more than eight classes, or they must be transformable into discrete classes. When continuous variables are transformed, the classes need not be of equal width. Considerable care has to be taken when defining discrete classes for continuous independent variables. If classes are too broad, differences in the relationship between dependent and independent variables may be masked. The algorithm combines classes in which the relationship is similar. It is then usually preferable to define too many discrete classes than to define a few broad classes, which may mask the true relationship between dependent and independent variables.

The algorithm described in the SEARCH procedure does not make any assumptions about the numerical structure of data. No distributional assumptions are required for the independent variables other than they have to be discrete or readily transformed to discrete variables. This aspect constitutes the most important difference with the stepwise multiple regression procedure. Because in many cases the data available are distributed non-normally and non-homoskedastically (conditions neccesary for the stepwise to be applied) the SCREEN algorithm is a preferred method. The chisquare distribution is involved in the SEARCH procedure to establish the best set of independent variables.

Logistic function

In developing a model for estimating the probability of mortality (or survival) for an individual tree as a function of independent variables, an important characteristic arises when dealing with a dichotomous dependent variable. For estimating the probability of individual tree mortality, the dichotomous dependent variable assumes the value 1 when a tree dies and the value 0 otherwise. As Hamilton (1980) discussed, a major difficulty in using a linear regression model for this particular case is that probabilities greater than one or less than zero may occur. This is most likely to occur when an independent variable assumes a value near or outside the limits of the initial dataset. A nonlinear model appears then, to be preferable. Walker and Duncan (1967) and Neter and Maynes (1970) have suggested that the logistic function, which limits estimates of probability to the interval (0,1), is the preferred model to express the relationship between a 0,1 dependent variable and a set of independent variables.

They also discussed several reasons for expecting a curvilinear relationship rather than a linear one when the dependent variable is dichotomous. Thus, a nonlinear model, such as the logistic, appears to be preferable (Hamilton, 1974).

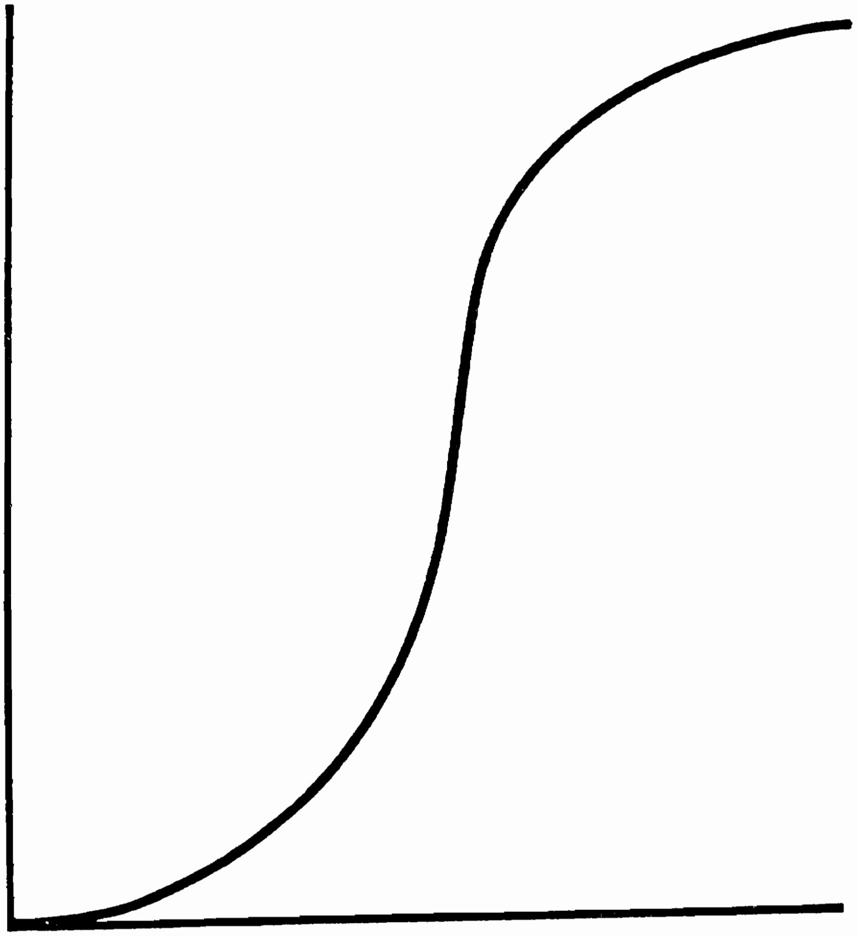


Figure 3. Logistic Function.

Mathematical models for predicting tree survival have assumed a variety of forms. Lee (1971), Keister (1972) and Lenhart (1972) used linear models to predict stand mortality. Hamilton (1974) and Hamilton and Edwards (1976) used a logistic function. Monserud (1976) proceeded also with a logistic function raised to a power. Glover and Hool (1979) and Somers et al. (1980) used forms of the Weibull function for modeling mortality. Buchman (1979) used a variant of the logistic function that included an upper asymptote parameter to ensure that survival predictions would be strictly less than one. Buchman (1983) pointed out three important reasons for using the logistic function to compute probability of mortality;

1. The function is bounded by 0 and 1.

2. The logistic function, or slight modifications of it, can be used to describe the patterns of mortality.

3. Nonlinear estimations (either nonlinear least squares regression or maximum likelihood) can be used to estimate the parameters.

So, the logistic equation is well suited to predict the probability that an individual tree will stay alive (Lowell and Mitchell, 1987). However, as Myers (1990) points out, special difficulties naturally occur when the response variable is not continuous. The following are two difficulties with the use of ordinary least squares:

1. Distribution of the error term is discrete and thus not normal.

2. Error variance is not homogeneous.

Thus, for the special case of a binary response in the dependent variable naturally brought about by the binary response situation. One solution is to use weighted least squares but a more popular approach is to model the mean response against the regressors through a logistic function.

The general form of the logistic mortality model is;

$$E(p) = 1/[1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k)}]$$

where;

$E(p)$ = expected probability of survival

β_i = i^{th} nonlinear regression coefficient (model parameter) $i = 0,1,...,k$

x_j = j^{th} independent variable $j = 1,2,...,k$

There are two ways of estimating the parameters of the logistic function; the linearization of the model and then the use of the least squares regression or a much preferred approach, the use of nonlinear regression to estimate the parameters (Hamilton, 1980). The first procedure has as a major shortcoming that is very demanding of data. When this approach is applied, each of the independent variables is broken into discrete classes. The cells formed by all possible combinations of the classes serve as the data points. If there are several independent variables, each with several classes, the number of possible cells may be very large. Each cell that is used in the estimation process must have at least one tree dead for the dependent variable to be defined (but not all trees dead). This can lead to the necessity of a great number of observations. The use of the nonlinear regression procedure to estimate the parameters is then a much preferred one. When this is done individual trees serve as data points.

Logistic regression is used in many scientific areas. Part of this interest comes from the fact that many software packages contain options for doing logistic regression. For example SAS (1983) offers PROC LOGIST. While the maximum likelihood estimation of coefficients in the model does not allow a "closed form" solution, it is relatively easy to use iterative procedures for estimating the coefficients.

Several commercial packages produce estimates of the parameters for the logistic model as well as information that leads to statistical inference regarding the roles of the variables in the model. In addition, estimates of the asymptotic variance-covariance matrix of the coefficients are often displayed. This allows us to make use of standard errors of coefficients for testing individual parameters. The latter constitutes an analog to the t-test that are used in ordinary regression.

Measures of lack of fit of the logistic model can be determined by making use of known features of maximum likelihood estimation. In order to see if the logistic model is appropriate, we can use the concept of deviance, which involves the likelihood ratio between a perfect model and the logistic model. These concepts are developed in Myers (1990).

In this context we find that the logistic model is inappropriate if the deviance is sufficiently large. The concept of deviance can be used to test hypotheses concerning the significance of set of variables among a larger number of independent variables. In this case, the partial chi square tests on individual coefficients can be done. Also, the use of asymptotic theory allows us to determine approximate variances and covariances of regression coefficients. This can be used in order to do tests of individual parameters of the logistic model, as an analog to the t-tests that are used in ordinary regression.

Thus, there are two ways for testing individual coefficients in the logistic model, one method uses the concept of deviance and the second one uses the approximate variance-covariance matrix. In both methods, the chi square distribution is involved.

Apart of the hypothesis testing mentioned above, the analyst has access to some other criteria to evaluate the performance of the logistic model when the method of maximum likelihood is used in order to estimate the parameters. There is an analog of the R^2 and other measures such as outlier and influence diagnostics. These are well explained by Pregibon (1981). Arabatzis (1990) also discussed a number of criteria for the selection of the model and for the detection of outliers and influence diagnostics. As he points out, these aspects are not easy to solve when the dependent variable is discrete, and many difficulties arise when trying to express a model's goodness of fit by a single criterion.

Halperin et al. (1971) also assert that the maximum likelihood procedure is preferred because it yields asymptotically unbiased coefficient estimates independently of any distributional assumptions about the data.

Several of the most recent nonlinear packages have as options both least squares and maximum likelihood parameter estimation procedures (Payandeh, 1983). Also, Nerlove and Press (1973) pointed out that estimation of the parameters of the logistic model using maximum likelihood procedure is well established. Usually the statistical packages that allow for nonlinear estimation procedures also contains procedures to evaluate the computed equation and the significance of the independent variables used as predictors .

Different models have been used to predict survival of trees like the Weibull distribution (Somers et al., 1980), the negative exponential (Moser, 1972), the power function (Hett, 1971). A large subclass of nonlinear funcionts restricted between 0 and 1 are the cumulative distribution functions. In the present study only the logistic function was used to compute the probabilty of survival.

Validation of the model

Validation of the model requires the model to be evaluated with a sample not included in the estimation of the parameters of the model (Hamilton,1980). As discussed by Snee (1977) there are different procedures useful in checking the validity of a model:

1. Comparison of the model predictions and coefficients with physical theory.

2. Collection of new data.

3. Comparison of results with theoretical models and simulated data.

4. Reservation of a portion of the available data to obtain an independent measure of the model prediction accuracy.

The first procedure should be examined as soon as a model has been developed. The scientist must check the biological meaning of the parameters obtained in the model as well as the prediction accuracy. The second procedure, that is the collection of new data, is perhaps the most preferred method of model validation. When the model accurately predicts for new data, it is considered validated. The third method, allows us to compare the obtained model with some theory that may exist and in this way enable us to ascertain the accuracy of the model. One can also check if the estimates of the coefficients agree with the coefficients developed by a simulation process.

Although the third method, the collection of new data, is the preferred one, it is sometimes not feasible to apply. In this case a procedure which simulates the collection of new data is needed. One practice that allows for cross-validation (or verification of the accuracy of the model) is data splitting, i.e. the partitioning of the data into two subsamples: a fitting sample, also called estimation data, used to estimate the model coefficients, and the prediction sample, used to measure the prediction accuracy of the model

As Myers (1986) points out, there are some difficulties with data splitting which should be "done by a great deal of common sense". If the data are time dependent, it is reasonable to assign the most recent observations in time to the validation dataset. If there is no obvious criterion such as time to use in order to split the data, there are some other methods that allows us to apply the cross-validation criteria to evaluate the accuracy of the model. One of these methods is the DUPLEX algorithm developed by Snee (1977). It basically consists of dividing the dataset into two datasets which have similar statistical properties. Alternatively, one may randomly divide the dataset using a random distribution generator and certain criteria in order to obtain the two datasets.

Stochastic versus deterministic

An important aspect in the mortality problem is the consideration of a major dichotomy in the mortality models, that is, the distinction between deterministic and stochastic models (Hamilton, 1980). In the first case, a given set of conditions will yield only one definite answer, that is deterministic mortality models predict the probability (or likelihood) that trees with certain characteristics growing in certain environmental conditions will die in some specified time interval. This same mortality model may be used in a stochastic sense to predict whether a specific tree is alive or dead at the end of a specified time interval. In the latter case the relations will be stated in terms of probability distributions and the answers will vary according to random elements. The probability of mortality is estimated for each tree in the stand. Next a uniformly distributed random number between 0 and 1 is selected for each tree and the following rules applied :

- If the probability of live is less than the random number the tree will die.
- If the probability is greater or equal to the random number, the tree will live.

In the past, mortality has often been treated in a deterministic fashion (Hatch, 1971). Some authors, like Paille and Smith (1971) suggest that a stochastic approach is more desirable.

Paille and Smith criticize many measures of competition as being inadequate for predicting mortality because of the variability of individual trees to survive when each is subjected to varying degrees of competition. They suggest using a tree's relative size and position within the forest canopy as a measure to predict mortality from competition. Variables such as the ratio of tree diameter at breast height to average stand diameter, and the ratio of total tree height to stand top height were used by Paille and Smith in simulation models to estimate the probability of tree mortality.

Weber et al. (1986) point out that a limitation of the stochastic approach is that it may require repeated runs and an averaging of results for meaningful analysis. According to them if the tree mortality could be described as a simple Bernoulli process agreement of deterministic and long-run stochastic results would be expected, but it is not clear that the deterministic approach of mortality and the average of stochastic runs will give similar results. Ek (1980) reports a preliminary examination of alternative methods for treating mortality in the Multipurpose Forest Projection System (MFPS) and one of the specific analysis conducted was a comparison of deterministic versus stochastic mortality generation. Projections were made for 11 survey plots from northeastern Minnesota (318 total trees in the input list). The study compared 10 year projections (ten 1 year growth periods) for deterministic mortality for six plots and three stochastic runs for the same plots.

Results from the individual plots and the mean for the deterministic and stochastic trials suggested little if any difference in the treatment of mortality. Also, Weber et al.(1986) note that there appears to be little difference between projections for deterministic and stochastic mortality treatment in terms of basic stand characteristics. Nonetheless the results that they arrived were 'admittedly from a limited trial'. So they suggested repeating the analysis for a much larger range of plots than the ones they used to conduct their study.

Weber et al. (1986) in their study analyzed 10 sample plots using

(i) the deterministic mortality model described by Buchman (1979) which expressed the probability of tree mortality as a function of tree size and growth by species, and

(ii) the stochastic mortality models.

The stochastic algorithm compares the predicted annual mortality rate with a computer generated uniform random number on the open interval 0-1. If the random number is less than the mortality rate, the tree is considered to have died and is removed from subsequent projection periods.

The data analyzed in their study came from the 1976-77 forest survey of northeastern Minnesota by the United States Forest Service. The plots represent a wide range of stand ages, densities, size class distributions, and species composition.

Projections of number of trees, basal area, stem wood volume per hectare and diameter distributions were made for one hundred 1-year growth periods for each of the 10 plots using both approaches. Projections using stochastic mortality were replicated 10 times for each plot, each with a different random start. The results were then averaged for comparison with the deterministic projections. Comparing both results they suggest that the number of stochastic replications was sufficient for meaningful analysis. They found no difference between deterministic and stochastic methods of treating mortality.

Close agreement was found for most of the plots, even for bimodal and trimodal distributions. The largest differences were found for the plots with the most irregular distribution shapes. Weber et al.(1986) suggested that a larger number of replications (more than 10) was probably needed to obtain a reliable average for the stochastic results.

Chapter III

Data

The data used in this study come from the Loblolly Pine Growth and Yield Research Cooperatives regionwide thinning study. During the 1980-81 and 1981-82 dormant seasons, permanent growth and yield research plots were established throughout the native range of loblolly pine. Plots were installed on cutover areas that received typical site preparation treatment for the site conditions and time at which the plantation was established. For each plot the location and stand history were recorded. Stand history included type of stand prior to the current plantation, when clearcut, type of site preparation, when planted, whether released or not released. Number of trees planted and age were also determined. The following data were recorded for all planted pines; dbh to the nearest tenth of an inch, total height to the nearest foot, height to the base of the live crown, crown class and stem quality assessment (Burkhart et al., 1985, Amateis and Burkhart, 1985, Walsh, 1986). To be included in the sample, the plantations had to meet the following characteristics;

- At least eight years in age.

- Unthinned.

- Free of evidence of heavy disease or insect attack, not heavily damaged by ice or windstorms.
- Free of interplanting.
- Unpruned.
- Not fertilized within the last four years.
- Not planted with genetically improved stock.
- Contain a minimum of 200 to 300 planted pine stems per acre which appear free to grow.
- Not more than 25 % of the main canopy composed of volunteer pines.

Initial measurements, plus remeasurements at three and six years after plot establishment are available. Tables 1 to 9 show summary statistics for the 186 plots. Figure 4 shows the location of these plots.

Not all the dead trees were included in this analysis; but rather only those trees that died within the first year following a measurement. The plots were measured at three-year intervals. Thus, characteristics of mortality trees were available at the beginning of an annual growing season only for those that died in the first year of a measurement period. Since the fitted equations represent the probability of mortality in a one-year period, only trees that died in the first year following a measurement could be used. Unthinned and thinned plots were treated the same in that only trees dying in the first year following measurements were included.

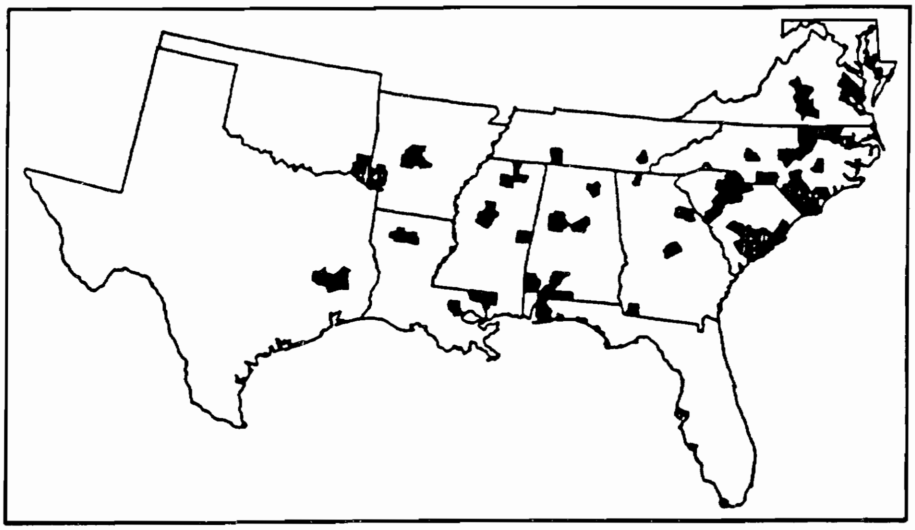


Figure 4. Map showing the location of sample plots by county (Burkhart et al.,1985).

Table 1. Summary of statistics for the 186 unthinned plots. Plot Establishment data.

variable	minimum value	mean	maximum value
AGE(years)	8.00	15.18	25.00
N(trees/acre)	270.00	569.35	1020.00
mdbh(in)	0.50	5.58	14.10
HD(ft)	14.40	41.13	74.40
SI(ft)	32.45	58.24	86.07
ACR(proportion)	0.03	0.46	0.94
BA(sq.ft/acre)	21.70	106.94	232.90

where :

AGE : Years from planting (years).
N : Number of planted loblolly pine (trees/acre).
mdbh : Arithmetic mean dbh of planted loblolly pine (in).
HD : average height dominant and codominant (ft).
SI : Site index (ft, base age 25). Computed using the equation from Burkhart et al. (1987)).
ACR : average crown ratio.
BA : Planted loblolly pine basal area (sq.ft/acre).

Table 2. Summary of statistics for the 186 unthinned plots. First Remeasurement data.

variable	minimum value	mean	maximum value
AGE(years)	11.00	18.18	28.00
N(trees/acre)	270.00	543.97	1020.00
mdbh(in)	0.70	6.20	14.90
HD(ft)	21.70	47.22	80.10
SI(ft)	34.21	58.34	84.34
ACR(proportion)	0.03	0.41	0.93
BA(sq.ft/acre)	45.60	124.36	233.90

where :

Variables are as defined in Table 1.

Table 3. Summary of statistics for the 186 unthinned plots. Second Remeasurement data.

variable	minimum value	mean	maximum value
AGE(years)	14.00	21.18	31.00
N(trees/acre)	90.00	510.57	935.00
mdbh(in)	0.70	6.68	15.30
HD(ft)	26.80	52.18	78.90
SI(ft)	38.12	58.22	80.11
ACR(proportion)	0.03	0.37	0.77
BA(sq.ft/acre)	24.00	134.46	205.20

where :

Variables are as defined in Table 1.

Table 4. Summary of statistics for the 186 light thinned plots. Plot Establishment data.

variable	minimum value	mean	maximum value
AGE(years)	8.00	15.18	25.00
N(trees/acre)	280.00	556.13	938.00
mdbh(in)	0.10	5.58	15.60
HD(ft)	14.80	40.84	72.60
SI(ft)	31.53	57.77	85.68
ACR(proportion)	0.04	0.48	0.97
BA(sq.ft/acre)	24.80	104.64	239.00

where :

Variables are as defined in Table 1.

Table 5. **Summary of statistics for the 186 light thinned plots. First Remeasurement data.**

variable	minimum value	mean	maximum value
AGE(years)	11.00	18.18	28.00
N(trees/acre)	191.00	335.18	633.00
mdbh(in)	1.00	6.89	17.20
HD(ft)	27.80	47.15	76.70
SI(ft)	33.90	58.14	82.61
ACR(proportion)	0.06	0.44	0.97
BA(sq.ft/acre)	30.60	92.23	174.40

where :

Variables are as defined in Table 1.

Table 6. **Summary of statistics for the 186 light thinned plots. Second Remeasurement data.**

variable	minimum value	mean	maximum value
AGE(years)	14.00	21.18	31.00
N(trees/acre)	89.00	328.49	633.00
mdbh(in)	1.40	7.47	18.40
HD(ft)	29.00	52.41	82.00
SI(ft)	37.93	58.39	81.50
ACR(proportion)	0.04	0.41	0.83
BA(sq.ft/acre)	30.70	105.78	193.50

where :

Variables are as defined in Table 1.

Table 7. **Summary of statistics for the 186 heavy thinned plots. Plot Establishment data.**

variable	minimum value	mean	maximum value
AGE(years)	8.00	15.18	25.00
N(trees/acre)	280.00	556.13	938.00
mdbh(in)	0.10	5.58	15.20
HD(ft)	14.80	40.84	72.60
SI(ft)	29.76	57.42	87.35
ACR(proportion)	0.10	0.49	0.95
BA(sq.ft/acre)	24.80	104.64	239.00

where :

Variables are as defined in Table 1.

Table 8. Summary of statistics for the 186 heavy thinned plots. First Remeasurement data.

variable	minimum value	mean	maximum value
AGE(years)	11.00	18.18	28.00
N(trees/acre)	130.00	254.95	479.00
mdbh(in)	1.70	7.18	16.00
HD(ft)	22.90	47.00	78.10
SI(ft)	32.00	57.90	84.08
ACR(proportion)	0.03	0.46	0.89
BA(sq.ft/acre)	20.50	75.94	130.90

where :

Variables are as defined in Table 1.

Table 9. Summary of statistics for the 186 heavy thinned plots. Second Remeasurement data.

variable	minimum value	mean	maximum value
AGE(years)	14.00	21.18	31.00
N(trees/acre)	44.00	250.48	479.00
mdbh(in)	1.80	7.89	17.20
HD(ft)	29.50	52.21	80.40
SI(ft)	37.40	58.15	82.81
ACR(proportion)	0.08	0.44	0.89
BA(sq.ft/acre)	18.90	89.21	145.20

where :

Variables are as defined in Table 1.

Chapter IV

Methods and Procedures

Unthinned plots

The procedures of variable selection as well as the verification or validation of the models were done using the unthinned plot data. The models that performed well for the unthinned plots were then applied to the thinned plots.

Variable selection (SCREEN algorithm).

The first step when dealing with the development of a mortality model as it is described by Hamilton and Edwards (1976) is to screen for the best set of independent variables that explain the probability of survival. The screening algorithm SCREEN was used in order to screen the best predictor variables. In this case, as the dependent variable was mortality, it was desirable to search for variables that were able to explain if the tree would stay alive or would die after a certain period of time. These variables should take into account tree size, tree vigor and tree competition.

As Hamilton (1974) pointed out, variation in mortality can frequently be explained by:

1. A measure of tree size,
2. stand density,
3. individual tree competition, and
4. growth rate (vigor).

It was desirable to have independent variables that show how these important aspects influenced mortality. A variable such as crown ratio (CR) can be used as a surrogate for tree vigor; diameter at breast height (dbh) or height (TH) were possible variables to express the tree size and variables as basal area (BA), ratio between the quadratic mean diameter and diameter at breast height (DD), competition index (CI) or relative spacing (RS) were possible variables that show the tree competition (Avery et al., 1983).

The SCREEN algorithm, specifically designed to be used when the dependent variable is dichotomous, was used to identify those potential variables that appear to explain the greatest amount of variability in the occurrence of mortality.

Information gained from applying the screening algorithm should be combined with an understanding of the process of mortality to develop a series of alternative sets of independent variables in the logistic model which was chosen, as it will be explained, in order to compute the probability of tree survival (or mortality).

Through the SEARCH routine of the SCREEN algorithm some understanding about different interactions among the independent variables can be obtained. The number of variables defined and the number of observations included required that the program read the data one at a time. The continuous variables were converted into discrete classes by dividing their range by 8 because the program allows only eight classes. In these cases it was preferred to define too many discrete classes than to define a few broad classes. This is because the algorithm combine classes for which the relationship is similar.

The algorithm asked the user to introduce the significance level at which the variables should be considered important in order to explain the dependent variable. At each node, SEARCH selected the most significant independent variable if this variable was significant at the user-supplied significance level. In this particular study a significance level of 95 percent was used.

Logistic function

As pointed out before, the SCREEN algorithm only gives the best set of predictors variables when dealing with a dichotomous dependent variable. In this particular study the logistic function was used as a model to predict the probability of remaining alive. This model was chosen because it is a function that is bounded by 0 and 1. There are also in the literature many studies that deal with the logistic function as being suitable to model tree mortality. Computer software is also available that allows parameter estimates for the logistic function. These procedures can also use either a backward procedure or a stepwise technique based on a strategy that requires very few calculations to determine the best variable to be added in the model in any given step. When using the SAS package, maximum likelihood estimates are computed by the Newton-Raphson method. This procedure also computes test statistics for assessing lack of fit of the model. It gives information that allows one to assess the significance of the different variables in the model. These methods were used to select the best set of predictors.

If the stepwise option is specified in the SAS procedure LOGIST, then LOGIST iterates to estimate parameters for variables forced into the model. Next it computes adjusted χ^2 statistics for all variables not in the model.

Every time the model changes because of new variables added or dropped from the model maximum likelihood estimates for the parameters of the variables in the new model are recomputed.

At each step of the stepwise procedure the MLE's (maximum likelihood estimates) and classification tables can be optionally printed. A classification table helps assess the quality of prediction.

The following are concepts used in the classification table when using the SAS package:

- ***Sensitivity*** : is the proportion of true positives that were predicted to be positive.
- ***Specificity*** : is the proportion of true negatives that were predicted to be negative.
- ***False positive rate*** : is the proportion of predicted positives that were actually negative.
- ***False negative rate*** : is the proportion of predicted negatives that were actually positive.

The BMDP software (1981) was also used to compute the stepwise logistic regression. In the BMDP program the PLR (Stepwise logist regression) procedure selects variables again in a stepwise manner. The independent variables can be continuous or categorical variables. In this latter case, design (or dummy) variables are created by the program. The number of dummy variables is one less than the number of categories of the categorical variable.

The program can estimate the coefficients for a logistic regression using two different methods: the maximum likelihood ratio (MLR) or an approximate asymptotic covariance estimate method (ACE). The latter is considered faster and is recommended for the initial run of large problems.

The maximum likelihood estimates are computed using an iterative process. The algorithm determines that the solution is reached when the relative improvement of the likelihood function is less than the value of the convergence criterion.

In the BMDP procedure the output includes at each step the log likelihood, the change in log-likelihood from the previous step and three goodness of fit chi square statistics. A summary with histograms of predicted probabilities of each group can also be obtained. If the misclassification of an outcome into the other group cost more than the misclassification of the other outcome into the first group, the selection of appropriate cutpoints on the probabilities should be based on a weighted consideration of the types of misclassification. In this software there is a COST function that computes the loss at each cutpoint. This table of correct and incorrect classifications using different cutpoints on the computed probabilities is equivalent at a specific cutpoint to the classification table computed with the SAS procedure.

The use of this package (BMDP) constituted a third way of screening for independent variables which then can be compared with the ones obtained in the SCREEN algorithm and in the LOGIST SAS procedure.

With these two packages one can produce estimates of the coefficients for the logistic regression model as well as information that leads to statistical inference regarding the roles of the variables in the logistic regression.

In order to see how well the logistic model fits the data, "the lack of fit test" was performed. This really is a ratio of likelihoods. This test is called the deviance associated with the fitted logistic regression.

The "lack of fit test" consist of postulating the following hypotheses :

H_0: Logistic model appropriate.

H_a: The logistic is not an appropriate model.

That is, if the null hypothesis is not rejected, then the logistic model is appropriate for the data.

If this deviance is found to be sufficiently large then the logistic model is inappropriate. The deviance (λ) has an approximate χ^2 distribution with n-k- 1 degrees of freedom where

n: is the number of observations

k: is the number of parameters in the model

Thus, it can be easily determined if the logistic model is suitable. The next important aspect is to compute the significance of the different independent variables in the model. In this case the use of deviance can be extended to allow test of other hypotheses, that is, contributions to the deviance can be produced in the same way as it is when working with the residual sum of squares (or regression sum of squares) in the ordinary regression case.

Inferences concerning any regressor or subset of regressors can be computed by determining how much the presence of each regressor contributes to the reduction in deviance. In this case partial χ^2 tests on individual coefficients can be developed.

In the case of only one regressor there are two ways of testing its significance:

- The following ratio $\left[\frac{\text{coefficient}}{\text{standard error}}\right]^2$

  is approximately a χ^2 with 1 degree of freedom

- Computation of the deviance of the model with all the variables and computation of the deviance of the model without the particular regressor of interest. The difference between these two deviances has a χ^2 distribution with k-1 degrees of freedom where k is the number of parameters in the full model.

Validation of the model

In order to validate the model the available dataset was split. Two data splitting methods were applied:

- As there is a sequence in time in the collection of the data one logical way to split the dataset was to use the plot establishment and first remeasurement data in order to estimate the parameters of the model; then the data of the second remeasurement was used to validate the model.

- If the sequence in time is not considered then the dataset can be randomly divided into two datasets. In this case half of the data was used as the estimation data (to estimate the parameters of the model) and the other half of the dataset, the prediction data was used to measure the accuracy of the model.

There are many omitted variables, weather variables being the most notable. Thus the two data splitting methods give an indication of the influence of these omitted variables. If both sets of validation statistics are similar, then the omitted variables are not likely that influential. If, on the other hand, the validation statistics computed using data from different time period are significantly worse than those obtained when the data are randomly subdivided, there is an indication that omitted time-dependent variables (e.g. weather) are important.

The logistic models obtained were compared with the models used in PTAEDA (Burkhart et al., 1987) and in TRULOB (Amateis and Burkhart, 1986).

PTAEDA is a forest stand simulator. It was developed to model growth in loblolly pine plantations on cutover, site-prepared areas. It is a distance-dependent individual tree simulator program. Growth increments are adjusted representing genetic and microsite variability. Subroutines were developed to simulate the effects of hardwood competition, thinning and fertilization on tree and stand development. In the present study there was no consideration to the hardwood competition or fertilization. The independent variables used in this simulator to compute the probability that a particular tree will live are crown ratio (CR) and competition index (CI). The equation has the following form:

$$PLIVE = b_1 CR^{(b_2)} e^{[-b_3 CI^{(b_4)}]}$$

where :

PLIVE : probability that a tree remains alive.

CR : crown ratio.

CI : competition index.

In this model mortality is generated stochastically through Bernoulli trials.

TRULOB is a distance-independent individual tree simulator program. It was developed for loblolly pine in cutover, site-prepared plantations. The dbh increment equation predicts dbh increment from a potential dbh growth function multiplied by a modifier function. The survival equation predicts the probability of survival as a function of crown ratio and the tree's competitive position in the stand (Amateis and Burkhart, 1986). The latter is expressed as a ratio between quadratic mean diameter and dbh. The model used in this case has the following form :

$$PLIVE = b_1 CR^{(b2)} e^{[(b3)(\overline{D}/dbh)^{(b4)}]}$$

where :

PLIVE : probability of living on an annual basis.

$b_1 - b_4$: parameters to be estimated.

CR : crown ratio

$\overline{D}$: quadratic mean diameter of the stand

dbh : initial dbh at the start of the growth period.

The PLIVE equation used in TRULOB is similar to the survival equation used by Daniels and Burkhart (1975). In this case the probability that a tree remains alive in a given year was assumed to be a function of its competitive stress and individual vigor as measured by photosynthetic potential. It can be seen that PLIVE increases with increasing crown ratio and decreases with increasing competition.

The following procedure was used in order to make comparisons among alternative models: the probability that a particular tree remains alive was computed with the three models (Logistic, PTAEDA and TRULOB). This probability was compared with a random number generated from a uniform distribution:

- If PLIVE is greater or equal than the random number the tree was considered alive.
- If PLIVE is less than the random number then the tree was considered dead.

Ten runs with different random starts were made for the different models. Tables of contingency were compiled for comparative purposes.

Thinned plots

In order to model survival for the trees on thinned plots the same available dataset was used. As described before data were available from 186 study locations established in cutover, site-prepared areas throughout much of the natural range of loblolly pine. At each location three plots were established. The three plots were required to be similar in site index, number of planted pine surviving and basal area to insure similar initial conditions. Then no thin (control), light thin (approximately 30 % of the basal area removed) and heavy thin (approximately 50 % of the basal area removed) treatments were randomly assigned to the plots. Up to this point only the control plots (unthinned) were used for variable selection procedures and to study the logistic model as an appropriate model to explain the probability of survival for a tree. The model with the best independent variables for the unthinned dataset was used to evaluate the performance using data from the thinned plots. There are available two sets of data according to the time elapsed since thinning : one of them right after thinning and the other set three growing seasons after thinning.

There are three important aspects to be considered:

1. Is the probability of survival of a tree dependent on the time passed after thinning?.

2. Is there any difference between the unthinned and the thinned plots?.

3. If there are differences between the thinned and unthinned plots, what variables account for the differences?.

A dummy variable was defined to account for the two growth periods following thinning. The significance of the deviance was also studied to see if there were differences between the periods after thinning. If there were differences then the second aspect (difference between thinned and unthinned plots) has to be studied separately in both periods. If not, then the data from both periods can be combined.

If differences are found between the unthinned and the thinned plots, then further study has to be done in the thinned plots. Variables such as time from thinning, age at thinning and proportion of basal area removed needed to be considered as possible independent variables in the logistic model. Once again the SCREEN algorithm was used to screen for the best set of predictors when considering the thinned plots.

Finally it was determined if only one model has to be used for thinned and unthinned plots, changing only the parameters, or if the same model but with different independent variables needs to be used.

Stochastic versus deterministic

Using the program TRULOB (Amateis and Burkhart, 1986) results from stochastic and deterministic approaches mortality predictions were studied. The model has been constructed using equations developed from the first remeasurement data described in Chapter II. As initial input there are two types of data that can be supplied to TRULOB

1. The stand level attributes age, site index and number of trees planted or surviving number of trees can be inputted to the model which will then generate a stand table for projection. or

2. An existing stand table can be inputted directly for projection.

The effects of hardwood competition are also considered in the program.

For this particular study only the type (1) input was used and the hardwood competition option was not used.

When the initial stand attributes are inputted to TRULOB a tree register by tenth-inch diameter classes is recovered from the Weibull distribution. Then the tree register can be projected until a specified target is reached. Among the target criteria are a specified age, basal area, quadratic mean diameter, dominant stand height, merchantable cord volume or mean crown ratio. It is an individual-tree, distance-independent growth and yield simulator program.

As a first step the program was run for different site index and number of trees combinations, without any kind of thinning practice. The following equation for survival probability was used:

$$PLIVE = 1.0537(CR^{(0.03723)})e^{(-0.02706(((BA/N/0.005454)^{0.5})/dbh)^{1.51529})}$$

where :

PLIVE : probability of living on an annual basis.

CR : crown ratio

BA : basal area

N : number of trees

dbh : initial dbh at the start of the growth period.

Once PLIVE was computed it was multiplied times the tree frequency to give an adjusted (for mortality) frequency. Thus mortality (survival) was deterministicaly controlled.

In order to study the stochastic approach the same program was used. A random generator function was introduced. When applying the stochastic approach the survival probability was computed using the same equation but this probability was then compared with the number generated from the uniform random variate between 0 and 1.

- If the probability of living was less than the random number the tree was considered dead.
- If this probability was greater or equal to the random number the tree was considered alive.

In order to make the comparisons between these approaches, several stochastic runs were performed, i.e. the TRULOB distance-independent individual tree growth model was used to make projections of number of trees, basal area, volume and diameter distributions. Then the average of these runs was compared to the deterministic projection. The same combinations of site index and number of trees planted were used for the stochastic projections as for the deterministic projections.

Site index values of 50, 60 and 70 (ft., base age 25 years) were used. Number of trees per acre planted of 400, 800 and 1200 were combined in a factorial fashion with the different values of site index.

This study was done following the same approach as Weber et al (1986). The results obtained from this particular study were compared with those obtained by these authors. It was important to find out if there was any difference in both approaches of the mortality aspect.

Variables such as number of trees after a certain number of growing seasons, stand basal area, total height, as well as the distribution of the live trees or dead trees were analyzed through these two ways of considering mortality. Practical differences in mean stand values for number of trees, basal area, volume or diameter distributions were evaluated.

Plots of the distributions of the dead trees for different number of growing seasons may help to see if there are significant differences between these two ways of treating mortality depending on the number of years simulated.

Chapter V

Results and Discussion

Unthinned plots.

First, the full dataset was split in two ways:

- Using a random generator function and a certain criterion the dataset was split into two sets : the estimation data and the validation data.

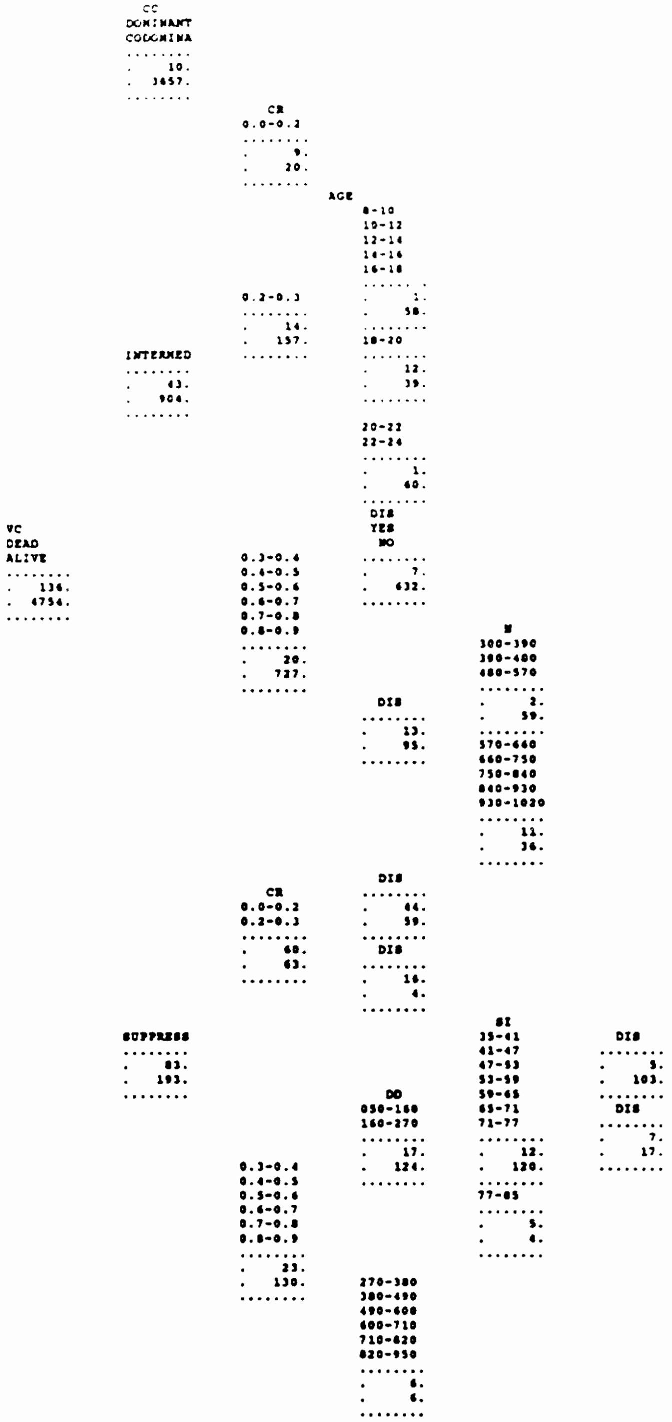
- Using the time (plot establishment, first and second remeasurement, three and six years after plot establishment), the data were split. The plot establishment and first remeasurement constituted the estimation data, while the second remeasurement constituted the validation data.

SCREEN

The SCREEN algorithm was used to select the best set of predictors when the dependent variable was the mortality variable. The GRAPH outputs are shown in Figures 5 and Figure 6.

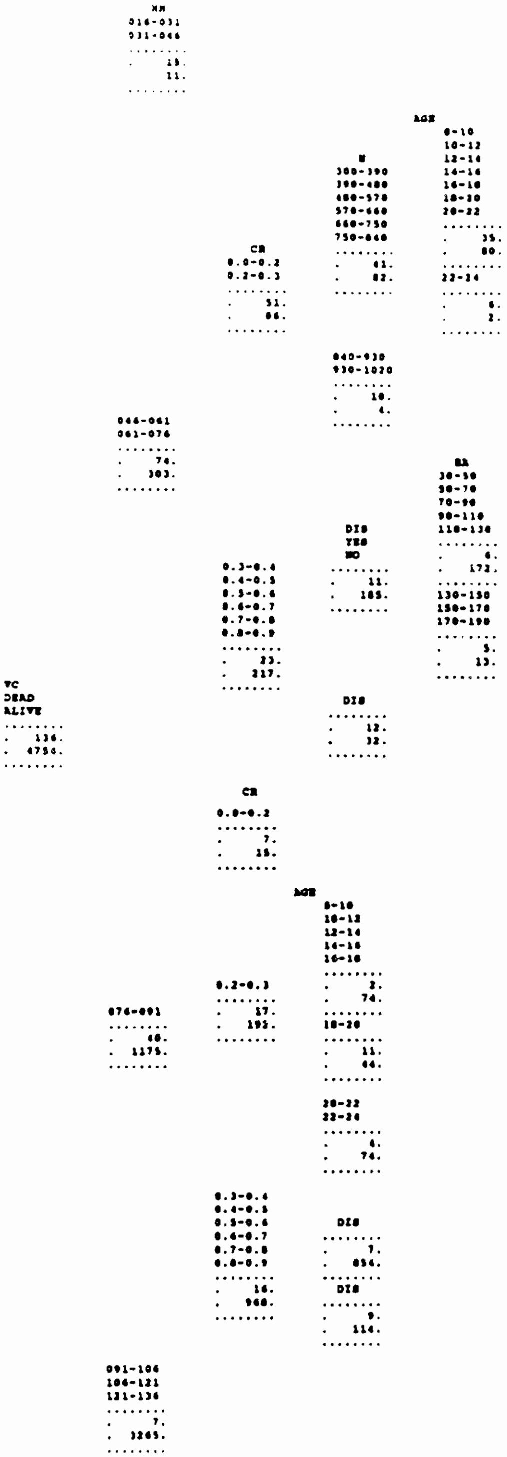
The following variables were used in this algorithm:

- CC : crown class.
- CR : crown ratio.
- DD : quadratic mean diameter/dbh.
- AGE : plantation age.
- SI : site index.
- BA : basal area
- DIS : presence or not of disease or insect.
- T : quality of top.
- S : straight quality.
- ST : stem quality.
- CI : competition index.



```
                 CC
               DOMINANT
               CODOMINA
               ........
               .     10.
               .   3457.
               ........
                                 CR
                              0.0-0.2
                              ........
                              .      9.
                              .     20.
                              ........
                                        AGE
                                             8-10
                                             10-12
                                             12-14
                                             14-16
                                             16-18
                                             ....... .
                              0.2-0.3        .      1.
                              ........       .     58.
                              .     14.      ........
                              .    157.      18-20
               INTERMED       ........       ........
               ........                      .     12.
               .     43.                     .     39.
               .    904.                     ........
               ........
                                             20-22
                                             22-24
                                             ........
                                             .      1.
                                             .     60.
                                             ........
                                               DIS
VC                                             YES
DEAD                                            NO
ALIVE                         0.3-0.4        ........
........                      0.4-0.5        .      7.
.    136.                     0.5-0.6        .    632.
.   4754.                     0.6-0.7        ........
........                      0.7-0.8
                              0.8-0.9                  M
                              ........              300-390
                              .     20.             390-480
                              .    727.             480-570
                              ........              ........
                                               DIS          .      2.
                                             ........       .     59.
                                             .     13.      ........
                                             .     95.      570-660
                                             ........       660-750
                                                            750-840
                                                            840-930
                                                            930-1020
                                                            ........
                                                            .     11.
                                                            .     36.
                                                            ........
                                               DIS
                                 CR          ........
                              0.0-0.2        .     44.
                              0.2-0.3        .     59.
                              ........       ........
                              .     68.        DIS
                              .     63.      ........
                              ........       .     16.
                                             .      4.
                                             ........
                                                              SI
               SUPPRESS                                     35-41          DIS
               ........                                     41-47        ........
               .     83.                                    47-53        .      5.
               .    193.                                    53-59        .    103.
               ........                         DD          59-65        ........
                                             050-160        65-71          DIS
                                             160-270        71-77        ........
                                             ........       ........     .      7.
                                             .     17.      .     12.    .     17.
                              0.3-0.4        .    124.      .    120.    ........
                              0.4-0.5        ........       ........
                              0.5-0.6                       77-85
                              0.6-0.7                       ........
                              0.7-0.8                       .      5.
                              0.8-0.9                       .      4.
                              ........                      ........
                              .     23.
                              .    130.      270-380
                              ........       380-490
                                             490-600
                                             600-710
                                             710-820
                                             820-950
                                             ........
                                             .      6.
                                             .      6.
                                             ........
```

Figure 5. SCREEN output when crown class is one of the independent variables.



```
                    HH
              016-031
              031-046
              .........
              .     15.
                    11.
              .........

                                                      AGE
                                                          0-10
                                                          10-12
                                               H          12-14
                                            300-390       14-16
                                            390-480       16-18
                                            480-570       18-20
                                            570-660       20-22
                                            660-750       .........
                                            750-840       .     35.
                                 CR         .........     .     80.
                             0.0-0.2        .     41.     .........
                             0.2-0.3        .     82.     22-24
                             .........      .........     .........
                             .     51.                    .      6.
                             .     86.                    .      2.
                             .........                    .........

                                            840-930
                                            930-1020
                                            .........
                                            .     10.
                                            .      4.
              046-061                       .........
              061-076
              .........
              .     74.                                      HA
              .    303.                                   30-50
              .........                                   50-70
                                                          70-90
                                                          90-110
                                             DIS          110-130
                                             YES          .........
                                             NO           .      6.
                             0.3-0.4        .........     .    172.
                             0.4-0.5        .     11.     .........
                             0.5-0.6        .    185.     130-150
                             0.6-0.7        .........     150-170
                             0.7-0.8                      170-190
                             0.8-0.9                      .........
                             .........                    .      5.
                             .     23.                    .     13.
                             .    217.                    .........
VC                           .........
DEAD                                         DIS
ALIVE
.........                                   .........
.    136.                                   .     12.
.   4754.                                   .     32.
.........                                   .........

                                 CR
                             0.0-0.2
                             .........
                             .      7.
                             .     15.
                             .........

                                       AGE
                                            0-10
                                            10-12
                                            12-14
                                            14-16
                                            16-18
                                            .........
                             0.2-0.3        .      2.
                             .........      .     74.
              076-091        .     17.      .........
              .........      .    195.      18-20
              .     40.      .........      .........
              .   1175.                     .     11.
              .........                     .     44.
                                            .........

                                            20-22
                                            22-24
                                            .........
                                            .      4.
                                            .     74.
                                            .........

                             0.3-0.4
                             0.4-0.5
                             0.5-0.6         DIS
                             0.6-0.7        .........
                             0.7-0.8        .      7.
                             0.8-0.9        .    854.
                             .........      .........
                             .     16.       DIS
                             .    960.      .........
                             .........      .      9.
                                            .    114.
                                            .........

              091-106
              106-121
              121-136
              .........
              .      7.
              .   3265.
              .........
```

Figure 6. SCREEN output when height ratio is substituted for crown class.

- TH : total height.
- NP : number of pine (trees/acre).
- RS : relative spacing
- HH : total height/height of dominant and codominant.
- VC : status of the tree : dead or alive.

A first run of the SCREEN algorithm was done using all the variables mentioned above except HH (output shown in Figure 5).

As it can be seen the first best predictor, according to this algorithm is crown class. There is no difference between dominant and codominant, but intermediate and suppressed are different from each other and from the dominant-codominant group.

For the crown class dominant-codominant no further significant explanatory variables were found. For intermediate and suppressed, the next best predictor was crown ratio. Because crown ratio is a continuous variable, it was divided into several classes. The algorithm combined some of the classes when there was no differences among them.

For crown ratio less than 0.2 as well as for the crown ratio between 0.2 and 0.3 the next best predictor is plantation age but for crown ratio between 0.3 and 0.9 the next best predictor is the presence or absence of disease or insect damage.

For those trees in the suppressed class and with crown ratio less than 0.3 again the next best predictor is the presence or absence of disease and insect damage. But for those trees with crown ratio between 0.3 and 0.9, the next best predictor is the relationship of the quadratic mean diameter over dbh of the subject tree.

The output also shows some interactions between variables such as crown ratio and age and between crown ratio and the presence or absence of disease or insect damage.

As crown class is a subjective variable dependent on the person doing the field work, some other variables were used to replace it. The variable used was the ratio of the total height of the tree and the average height of dominant and codominant trees (HH). This variable together with some other variables such as CI, BA, AGE, CR, already used in the SCREEN algorithm were used in a discriminant analysis in order to find a suitable variable to replace crown class. When doing this analysis in a stepwise manner HH appears as the best variable to replace CC.

Using HH instead of crown class and all the other variables used with crown class, the SCREEN algorithm was run again. The output is shown in Figure 6.

In this case the best predictor appears to be the ratio of the heights. As this variable (HH) is continuous, it was divided into several discrete classes. There is no difference between the classes 0.16 to 0.46 and in this case no further significant explanatory variables (at the user's level) were found.

For those trees with HH between 0.46 and 0.76 crown ratio was the next best predictor. For crown ratio less than 0.3 the next best predictor was number of pine trees per acre. For those trees with HH between 0.46 and 0.76 and CR between 0.3 and 0.9 the next best predictor was the presence or absence of disease or insect damage.

According to this screening, variables such as HH and CR can be used as independent variables. When CC is replaced by HH, the latter variable appeared as the best predictor. Thus, HH can be used as independent variable replacing CC.

CI did not appear in this screening. Nonetheless it was used as an explanatory variable that accounted for the tree competition.

Then, CR, CI and HH are potential independent variables for the logistic model. Another variable, DD is used together with CR and HH in the logistic model. The way these two sets of variable are used allows comparison of these models with the models used to compute probability of survival in PTAEDA (when using CI) and with the model used in TRULOB (when using DD).

The best set of independent variables found through the SCREEN algorithm were compared with the best predictors variables found through the stepwise logistic regression. In this particular case the results were similar. The first procedure (SCREEN) constitutes a good alternative for those cases where the assumptions (normality and homoscedasticity) do not held.

Logistic function

Once the variables were selected, two different statistical software packages, SAS (SAS, 1983) and BMDP (BMDP, 1981), were applied in order to estimate the equations. For BMDP two methods to estimate the coefficients were applied: the maximum likelihood ratio (MLR) and the asymptotic covariance estimates (ACE).

Using half of the data randomly selected as fitting data and SAS software the results were as follow:

$$\hat{p} = 1/[1 + e^{-(-5.1936 + 9.0182CR + 7.3015HH - 0.5628CI)}]$$

where;

$\hat{p}$ = estimated probability of survival in a specific time interval $0 \leq \hat{p} \leq 1$

Using the BMDP package (MLR method) this was the result:

$$\hat{p} = 1/[1 + e^{-(-5.17601 + 9.0081CR + 7.2863HH - 0.5633CI)}]$$

The results obtained with the BMDP package using the ACE method are similar to those obtained with this package using the MLR method. As these results are also similar to those obtained with the SAS package only the latter was used for further study.

Using the DD variable the model following estimates were obtained:

$$\hat{p} = 1/[1 + e^{-(-5.3011 + 10.2728CR + 6.6588HH - 0.5931DD)}]$$

where;

DD: ratio of quadratic mean diameter to dbh

In these models the trends were as expected. The probability of survival increased with increasing crown vigor measured by CR. This probability also increased with increasing HH, that is the higher the height of a tree with respect to the height of dominant and codominant trees the greater the probability of survival for a tree. The probability of survival decreased with increasing competition measured by either CI or DD.

In order to ascertain if there was any improvement due to adding more variables (all of them found to be significant through the SCREEN algorithm) once again similar procedures were performed as before.

The results were :

$$\hat{p} = 1/[1 + e^{-(-4.8455 + 7.4251CR + 8.8110HH - 0.5382CI - 0.9377DIS + 0.0423AGE - 0.0193BA + 0.012N)}]$$

where;

DIS: presence or absence of disease or insect damage.

AGE: plantation age.

BA: basal area.

N: number of trees.

$$\hat{p} = 1/[1 + e^{-(-4.23 + 8.1681CR + 8.5736HH - 0.5724DD - 0.8337DIS + 0.0678AGE - 0.0247BA + 0.0002N)}]$$

where;

$\hat{p}$ = estimated probability of survival in a specific time interval $0 \leq \hat{p} \leq 1$

CR, HH, DIS, AGE, BA, N are as defined before.

DD: ratio of quadratic mean diameter to dbh.

Using the plot establishment and first remeasurement data as the fitting dataset the results, with CI as one of the independent variables, were as follows :

$$\hat{p} = 1/[1 + e^{-(-5.4191 + 9.3846CR + 7.2552HH - 0.4838CI)}]$$

Using DD instead of CI this was the result:

$$\hat{p} = 1/[1 + e^{-(-5.9652 + 10.4028CR + 7.1591HH - 0.4103DD)}]$$

Using CI and all the other variables this was the result:

$$\hat{p} = 1/[1 + e^{-(-5.2047 + 7.8851CR + 8.3963HH - 0.5122CI - 1.1567DIS + 0.0339AGE - 0.0162BA + 0.0017N)}]$$

Using DD as one of the independent variable instead of CI this was the result:

$$\hat{p} = 1/[1 + e^{-(-4.7128 + 8.4870CR + 8.3707HH - 0.5011DD - 1.07DIS + 0.0509AGE - 0.0211BA + 0.0008N)}]$$

where;

$\hat{p}$ = estimated probability of survival in a specific time interval $0 \leq \hat{p} \leq 1$

CR, HH, DIS, AGE, BA, N are as defined before.

DD: ratio of quadratic mean diameter to dbh.

In order to make comparisons with the models already used in PTAEDA and TRULOB , these models were also fitted in the same way as was done with the logistic function. The results when using half of the dataset as fitting data follow.

PTAEDA MODEL: the equation obtained was:

$$PLIVE = 1.1111CR^{0.1385}e^{[-0.0098CI^{2.4872}]}$$

where :

PLIVE : probability that a tree remains alive.

TRULOB MODEL: the results were:

$$PLIVE = 1.1931CR^{0.1367}e^{[-0.0827(\overline{D}/dbh)^{1.7505}]}$$

where :

$\overline{D}$: quadratic mean diameter of the stand

dbh : initial dbh at the start of the growth period.

Using the plot establishment and first remeasurement as fitting data the results were:

PTAEDA MODEL : The equation had the following form:

$$PLIVE = 1.1026CR^{0.1340}e^{[-0.0085CI^{2.6016}]}$$

TRULOB MODEL: the results were:

$$PLIVE = 1.1768CR^{0.1321}e^{[-0.0734(\overline{D}/dbh)^{1.7413}]}$$

The appropriateness of the logistic model obtained using as independent variables CR, HH and CI was studied. The model deviance was equal to 1126.93. This value of deviance did not reject the null hypothesis that the logistic was an appropriate model. Thus, the logistic model was a good model for these data. The value for the rest of the models are in Table 10. All of them led to the same conclusion (i.e., the logistic was an appropriate model) when considering other independent variables such as DD, AGE, N, DIS or BA.

Table 10. Logistic model (using half of the dataset).

Variables in the model	Deviance	Model χ^2 [a]
CR,HH,CI	1126.93	798.22*
CR,HH,DD	1173.58	751.57*
CR,HH,CI,DIS,AGE,N,BA	1080.70	838.03**
CR,HH,DD,DIS,AGE,N,BA	1114.23	804.50**

* χ^2 with 3 degrees of freedom
** χ^2 with 7 degrees of freedom
[a] Significance of all the variables considered together in the logistic model (apart from the intercept).

where :

CR,HH,CI,DD,DIS,AGE,N,BA are as defined before.

The logistic model with the variables CR, HH and CI was further studied to see if all the variables considered together were significant (apart from the intercept). The χ^2 for the model was found to be significantly different of zero; thus all the variables were important when considered altogether. Similar studies were done for all the other models, obtaining the same results.

The last step was to study different subsets of parameters (Table 11). It can be seen that the different subsets of parameters lead to the conclusion that the variables, CR, CI and HH were important in the logistic model as predictors variables.

Table 11. Study of the logistic model (half dataset).

H_0	Deviance	p-value
$\beta_1 = 0$	161.09	0.000*
$\beta_2 = 0$	164.17	0.000*
$\beta_3 = 0$	54.39	0.000*
$\beta_1 = \beta_2 = 0$	408.06	0.000**
$\beta_1 = \beta_3 = 0$	278.15	0.000**
$\beta_2 = \beta_3 = 0$	323.69	0.000**

* χ^2 with 1 degree of freedom.

** χ^2 with 2 degrees of freedom.

Logistic model of the following form :

$$\hat{p} = 1/[1 + e^{-(b_0 + b_1 CR + b_2 HH + b_3 CI)}]$$

where :

CR,HH,CI are as defined before.

When the same study was done for the logistic model including the variable DD similar results were obtained.

Validation of the model

The validation of the models obtained using the logistic function was done in two different ways:

- First, half of the dataset (the validation data) was used, using only the equations obtained with the SAS software.
- Then, the dataset corresponding to the second remeasurement was used as the validation data.

In order to compare the logistic models (considering the DD, CI and all the other variables) with the models already used in PTAEDA and TRULOB ten stochastic runs were made using a different random number for each run but the same for each model within each run. In each run, the probability of survival of a certain tree was compared with a random number generated from a uniform random function.

The average of the ten runs for each of the models (PTAEDA, TRULOB, logistic(CI) and logistic (DD)) are shown in Table 12.

Table 12. Contingency table when half of the dataset is used as validation data.

		Predicted		
		0	1	Total
true	0	10 (1) 9 (2) 16 (3) 16 (4) 17 (5) 17 (6)	32 (1) 33 (2) 26 (3) 26 (4) 25 (5) 25 (6)	42
	1	207 (1) 194 (2) 123 (3) 130 (4) 130 (5) 128 (6)	4478 (1) 4491 (2) 4562 (3) 4555 (4) 4555 (5) 4557 (6)	4685
	Total			4727

where :

(1) PTAEDA model

(2) TRULOB model

(3) Logistic model (CR,HH,CI)

(4) Logistic model (CR,HH,DD)

(5) Logistic model (CR,HH,CI,DIS,AGE,N,BA)

(6) Logistic model (CR,HH,DD,DIS,AGE,N,BA)

CR,HH,CI,DD,DIS,AGE,N,BA are as defined before, and 0: dead tree; 1:live tree

According to these tables there is no significant improvement of the logistic model over the models already used in PTAEDA and TRULOB. Even when all the variables found significant through the SCREEN algorithm were introduced there was no significant improvement of the logistic model.

Another strategy used to test the performance of the models was done through the COST function in BMDP program and the classification table in SAS package (Tables 13, 14, 15 and 16). In these particular tables the logistic model with CR, HH and CI as independent variables was used. Table 15 shows the results for the misclassification table when the BMDP software is used. Different cutpoints are shown with the corresponding "cost" for the misclassification of the different values. Table 16 shows the misclassification table obtained when using the SAS package. In SAS only one cutpoint is used. This corresponds to 0.5 of the BMDP package.

Table 13. Misclassification matrix (BMDP software)

Each cutpoint yields a misclassification matrix,containing counts of the form of this table

	Predicted as dead	Predicted as live
Cases in dead	A (•)	B
Cases in live	C	D

(•) where A,B,C and D are as shown in table 15

Table 14. Cost matrix (BMDP software)

```
 The 'cost' matrix times the classification table yields the

gain/loss function

                                   Cost matrix

                                    Predicted

                         Dead                Live

                Dead     0.00               -1.00


                Live    -1.00                0.00


   Loss=+0.0000A - 1.0000B - 1.0000C +0.0000D
```

Table 15. Results for the misclassification table (BMDP software).

Cut	Correct pred.			Percent correct			Incorrect pred.			Cr.Pr.(1)	Gain or
Point	Dead	Live	Total	Dead	Live	Total	Dead	Live	Total	Ratio	Loss (2)
	A	D	E=A+D	A/(A+B)	D/(C+D)	E/(E+F)	B	C	F=B+C	AD/BC	
0.008	228	2822	3050	97.02	55.56	57.40	7	2257	2264	40.73	-2264
0.025	215	3924	4139	91.49	77.26	77.89	20	1155	1175	36.52	-1175
0.042	208	4286	4494	88.51	84.39	84.57	27	793	820	41.64	-820
0.058	202	4480	4682	85.96	88.21	88.11	33	599	632	45.78	-632
0.075	194	4610	4804	82.55	90.77	90.40	41	469	510	46.51	-510
0.092	184	4696	4880	78.30	92.46	91.83	51	383	434	44.24	-434
0.108	172	4765	4937	73.19	93.82	92.91	63	314	377	41.43	-377
0.125	165	4803	4968	7021	94.57	93.49	70	276	346	41.02	-346
0.142	156	4840	4996	66.38	95.29	94.02	79	239	318	39.99	-318
0.158	149	4868	5017	63.40	95.85	94.41	86	211	297	39.97	-297
0.175	142	4891	5033	60.43	96.30	94.71	93	188	281	39.72	-281
0.192	137	4919	5056	58.30	96.85	95.14	98	160	258	42.98	-258
0.208	132	4935	5067	56.17	97.16	95.35	103	144	247	43.92	-247
0.225	128	4948	5076	54.47	97.42	95.52	107	131	238	45.18	-238
0.242	126	4959	5085	53.62	97.64	95.69	109	120	229	47.77	-229
0.258	124	4971	5095	52.77	97.87	95.88	111	108	219	51.42	-219
0.275	121	4977	5098	51.49	97.99	95.94	114	102	216	51.79	-216
0.292	117	4990	5100	49.79	98.11	95.97	118	96	214	51.47	-214
0.308	110	4983	5100	46.81	98.25	95.97	125	89	214	49.34	-214
0.325	105	5003	5108	44.68	98.50	96.12	130	76	206	53.17	-206
0.342	98	5006	5104	41.70	98.56	96.05	137	73	210	49.05	-210
0.358	94	5010	5104	40.00	98.64	96.05	141	69	210	48.41	-210
0.375	90	5013	5103	38.30	98.70	96.03	145	66	211	47.14	-211
0.392	83	5016	5099	35.32	98.76	95.95	152	63	215	43.48	-215
0.408	80	5022	5102	34.04	98.88	96.01	155	57	212	45.74	-212
0.425	74	5026	5100	31.49	98.96	95.97	161	53	214	43.59	-214
0.442	71	5030	5101	30.21	99.04	95.99	164	49	213	44.44	-213
0.458	66	5037	5103	28.09	99.17	96.03	169	42	211	46.84	-211
0.475	62	5038	5100	26.38	99.19	95.97	173	41	214	44.04	-214
0.492	60	5043	5103	25.53	99.29	96.03	175	36	211	48.03	-211
0.508	58	5045	5103	24.68	99.33	96.03	177	34	211	48.62	-211
0.525	55	5048	5103	23.40	99.39	96.03	180	31	211	49.76	-211

(1) Cross product ratio

(2) Gain or loss function

Table 16. Table of misclassification (SAS software).

		Predicted		
		Negative	Positive	Total
True	Negative	59	176	235
	Positive	36	5043	5079
	Total	95	5219	5314

Sensitivity: 99.3% Specificity: 25.1% Correct: 96.0%

False positive rate: 3.4% False negative rate: 37.9%

where in this particular study,
negative refers to dead trees and
positive refers to live trees.

It can be seen that according to these functions used in both softwares no significant difference were found with the results already obtained through the contingency table.

The validation of the logistic models was also done using the second remeasurement dataset as validation data. That is, the time aspect was used to split the dataset. This way of splitting also gives an opportunity to evaluate the factor 'time' and its influence in the prediction of the probability of survival of a tree. No significant differences were found with the results shown in Table 12. This mean that variables such as the weather conditions through time apparently can not be considered as significantly influential for this particular dataset.

Thinned plots

The following aspects were studied when dealing with thinned plots.

- Are there any differences in the thinned plots because of the period passed after thinning?.
- Are there any differences between the unthinned and thinned plots?.
- If differences are found in the latter aspect, which variables can be used in the logistic model to explain this difference?.

In order to study these aspects the logistic models including the variables CR, HH and CI and CR, HH and DD were used. A dummy variable indicated that there are differences in the thinned plots because of the time after thinning. The question of differences between thinned and unthinned plots was studied in one particular time, the period right after thinning. The unthinned and thinned plots were found to differ significantly.

Thus, the next step was to introduce variables that account for thinning such as: proportion of residual basal area, age at thinning, time after thinning. The SCREEN algorithm was used again with the thinned plots in a similar way as it was done with the unthinned plots. The same variables, CR, HH, CI and DD were found significant in these case together with the variables that account for thinning. The logistic model was fitted for the thinned plots, but the three variables used to explain the thinning aspect do not improve the model. Thus, an analysis was conducted with the variable CR, which was found to be important for distinguishing the thinned and unthinned plots. The analysis showed that for those trees with crown ratio less than 0.3 the probability of survival for the second three-year growth period after thinning was higher than for those trees with crown ratio less than 0.3 for the period immediately following thinning. So, in order to capture this increased level of survival for the thinned plots CR was raised to a power. Several values were tried. A value of 1.5 was found to give the smallest model deviance. Thus the final model for the thinned plots, when CI is one of the independent variables, was the following:

$$\hat{p} = 1/[1 + e^{-(0.5065 + 7.8838(CR^{1.5}) + 1.8340HH - 0.4114CI)}]$$

where;

$\hat{p}$ = estimated probability of survival in a specific time interval $0 \leq \hat{p} \leq 1$

CR: crown ratio.

HH: ratio of the total height to height of dominant and codiminant

CI: competition index

The same analysis was conducted with the variable crown ratio when DD was one of the independent variables. Once again the smallest deviance was obtained when crown ratio was raised to the 1.5 power. Thus, the final model for the thinned plots, when DD is one of the independent variables, is the following:

$$\hat{p} = 1/[1 + e^{-(1.4106 + 8.6588(CR^{1.5}) + 0.9372HH - 0.6406DD)}]$$

where;

$\hat{p}$ = estimated probability of survival in a specific time interval $0 \leq \hat{p} \leq 1$

CR: crown ratio.

HH: ratio of the total height to height of dominant and codominant

DD: ratio of quadratic mean diameter to dbh of the subject tree.

Stochastic versus deterministic

Using the TRULOB program in a deterministic way 50 years projections were obtained for the following combinations of number of trees per acre planted (N) and site index (SI) measured in feet at base age 25 years :

N = 400
= 800
= 1200

and

SI = 50
= 60
= 70

Using the same program but introducing a uniform random generated number between 0 and 1, the same combinations of number of trees and site index were used. For each combination, ten stochastic runs were done. In order to do a comparison between both approaches the ten stochastic runs were averaged. Table 17 shows the results for number of live trees, average height, total volume and basal area.

Table 17. Stochastic runs and deterministic results (50 years projections).

Conditions (1)	Type of mortality model (2)	Number of live trees	Average height	Total volume	Basal area
SI=50	D	103.1	64.82	1855.6	60.58
N=400	S	102.7	66.72	1922.5	61.03
SI=50	D	174.9	63.92	2402.2	79.04
N=800	S	173.0	65.95	2520.3	80.57
SI=50	D	236.2	63.41	2720.9	89.89
N=1200	S	233.3	65.12	2806.8	90.38
SI=60	D	92.4	78.79	2378.4	64.07
N=400	S	93.5	81.03	2477.1	64.88
SI=60	D	156.5	77.96	3004.3	81.43
N=800	S	153.4	79.94	3074.1	81.32
SI=60	D	210.9	77.21	3376.9	92.10
N=1200	S	207.3	79.48	3488.7	92.50
SI=70	D	84.9	93.10	2959.9	67.59
N=400	S	81.5	96.14	3008.2	66.65
SI=70	D	142.2	91.91	3733.0	86.02
N=800	S	140.7	94.56	3870.7	86.72
SI=70	D	191.7	91.18	4207.4	97.44
N=1200	S	193.0	93.82	4451.9	100.28

(1)SI=site index
N=Number of trees

(2)D,deterministic;S,stochastic(average 10 runs)

No practical differences between both types of projections were observed. Plots for the number of dead trees were made. Figure 7 shows this distribution per tenth inch diameter class. For the other combinations plots can be seen in Appendix A.

No important differences between the deterministic projection and the average of ten stochastic runs were observed.

The same procedure was applied but using as a target age 25 growing seasons. Figure 8 shows the results for this particular case for live trees. Again no significant differences are observed between the two approaches. Plots for other combinations of site index and number of trees can be seen in Appendix B.

These results agree with the ones found by Weber et al. (1986), using a different simulator program, different mortality equation and the same number of stochastic runs (10).

Deterministic versus averaged stochastic projections show no practical differences in mean stand values for number of trees, basal area, volume or diameter distributions. Deterministic projection has the advantage that multiple runs are not needed as in the stochastic approach.

While additional replication could be performed, it is thought that they would confirm what it is found with 10 runs.

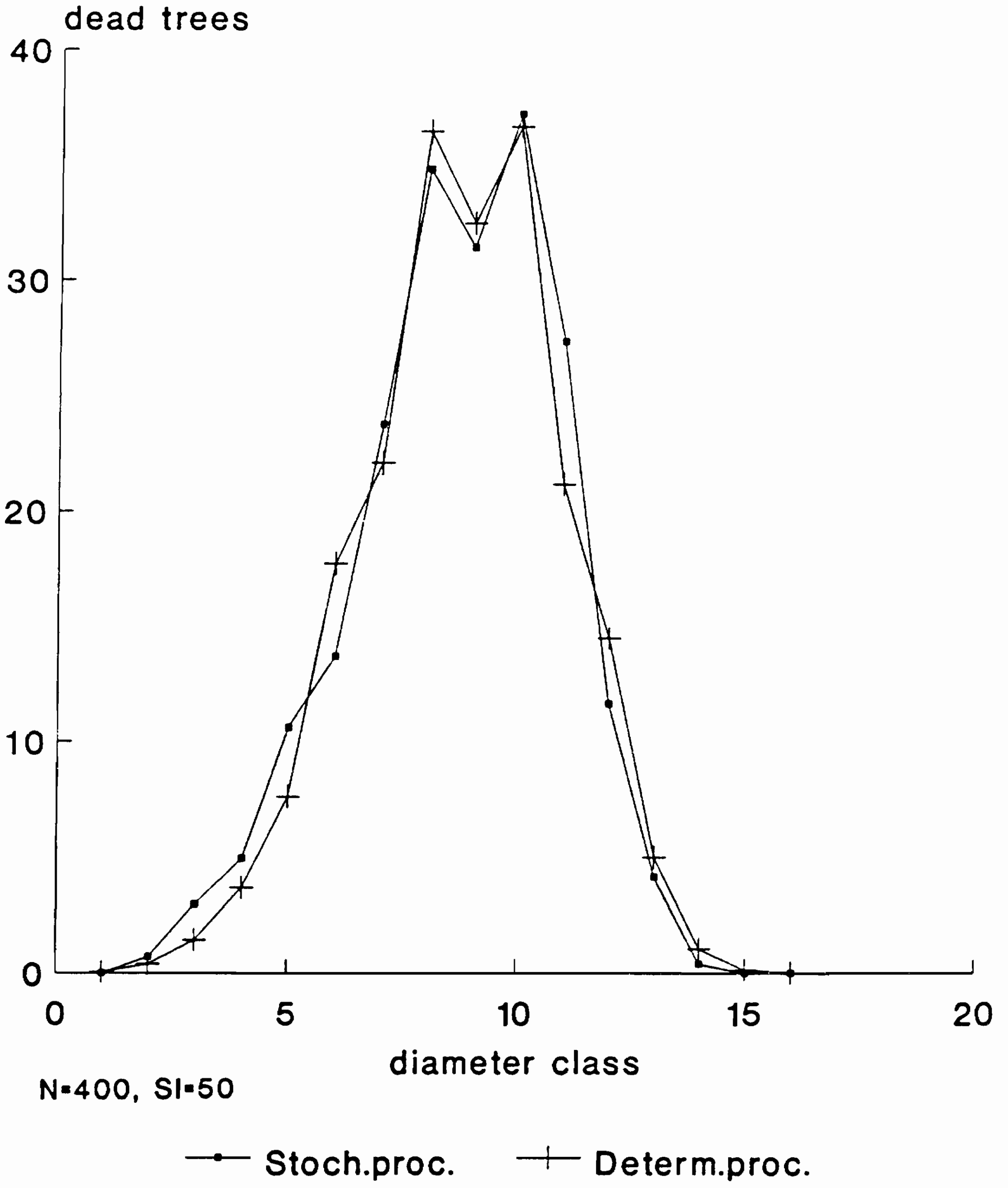


Figure 7. Diameter distribution of dead trees.

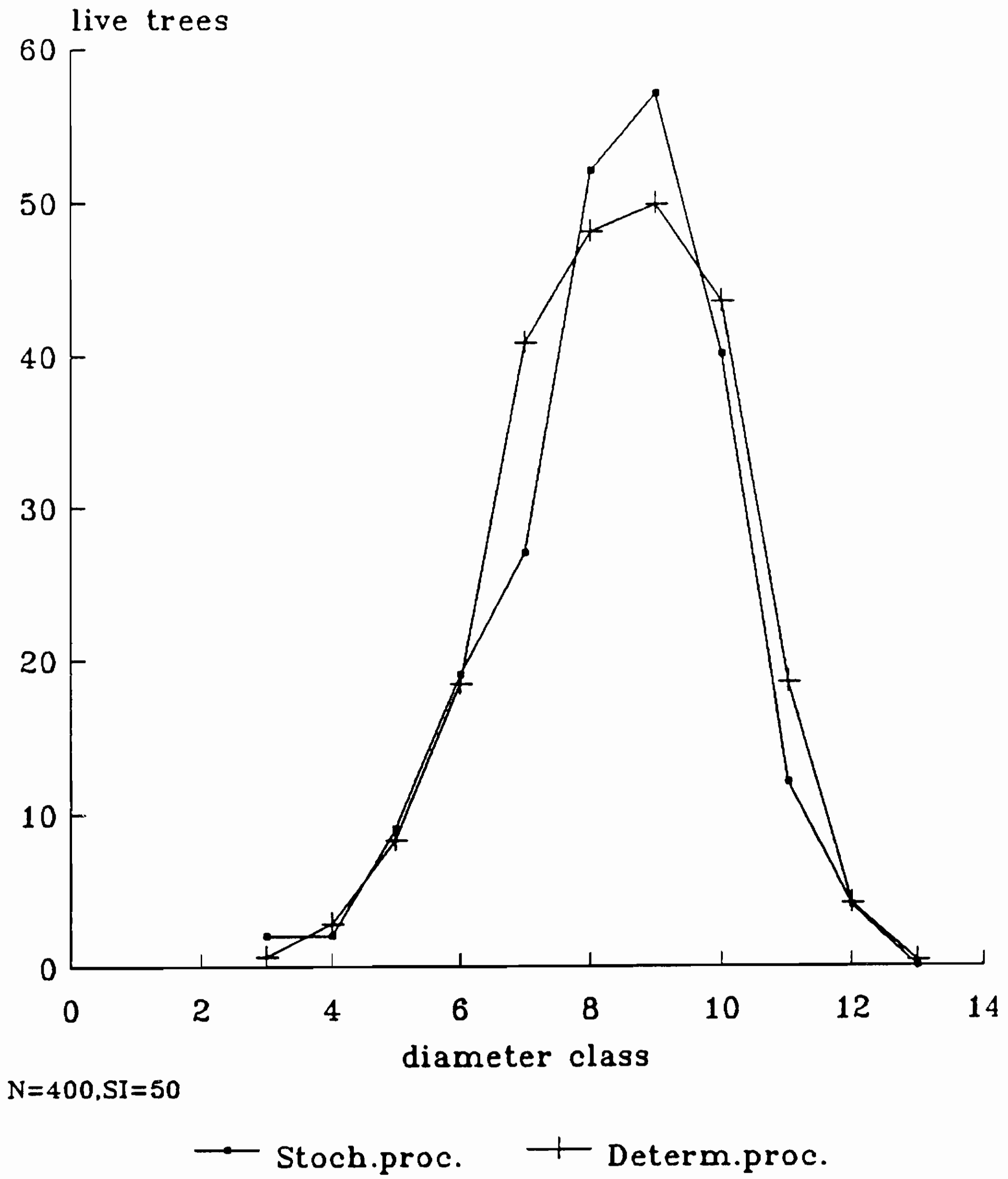


Figure 8. Diameter distribution of live trees.

Chapter VI

Summary and Conclusions

Different approaches were taken to obtain a model that predicts probability of survival for an individual tree growing under certain conditions with particular characteristics. The logistic model with the variables CR (crown ratio), HH (total height/height of dominant and codominant trees) and CI (competition index) was used and compared with the model used in the PTAEDA simulation program. No significant improvement was obtained. A logistic model using the variables CR, HH and DD (quadratic mean diameter/dbh) was compared with the model used in TRULOB. Once again no significant difference was found between these two models.

The logistic model with the variables CR, HH, CI, AGE (plantation age), BA (basal area), N (number of trees), DIS (presence or absence of disease or insect damage) was used to see if there was any improvement due to these latter variables, which were found significant in the SCREEN algorithm. No improvement was found when compared the logistic model with only the first three variables (CR, HH, CI) in it with the logistic model with all the variables (CR, HH, CI, AGE, BA, N, DIS) in it.

Once the models for the unthinned plots were determined, they were used with the thinned plots. The logistic model using CR raised to the 1.5 power had a better performance in the thinned plots. Thus, for the thinned plots the variables HH, CI and CR raised to 1.5 were used as independent variables in the logistic model.

The logistic models used in this work do no show improvement when compared to the models already published (TRULOB < PTAEDA). Further study should be done using other models that may improve the prediction of probability of a tree.

A particular study was conducted to see what differences, if any, exist when treating mortality in a deterministic or stochastic way. No practical differences were found in these two approaches of computing the probability of mortality for a particular tree. This study confirms the results found in previous work done by Weber et al.(1986). Nonetheless, more study involving different models to predict mortality as well as different number of stochastic runs could be useful to further augment results to date.

References

Amateis, R.L. and Burkhart, H.E. 1985. "Site index curves for loblolly pine plantations on cutover site-prepared lands." *South.J.Appl.For.*, 9: 166-169.

Amateis, R.L and Burkhart, H.E. 1986. "TRULOB tree register updating for Loblolly Pine.Prototype version P.1." Not published.

Arabatzis, A.A. 1989. "Qualitative response model theory and its application to forestry." Ph.d Thesis. School of For. and Wild. Res., Va. Polytech. Inst. and State Univ.,Blacksburg.

Avery T.E and Burkhart H.E. 1983. "Forest measurements. " 3rd edition. McGraw-Hill Book Company, New York.

BMDP. 1981. " Biomedical computer programs. " *University of California Press, Berkeley, California.*

Buchman, R.G. 1979. "Mortality functions."In A Generalized Forest Growth Projection System Applied to the Lake States Region. *USDA Forest Service.* Gen.Tech.Rep.NC-49 :47-55.

Buchman, R.G.,Pederson, S.P.,and Walters, N.R. 1983. "A tree survival model with application to species of the Great Lakes region." *Can.J.For.Res.*, 13: 601-608.

Burkhart H.E. 1987. " Data collection and modeling approaches for forest growth and yield prediction. " In *'Predicting Forest Growth and Yield: Current Issues, Future Prospects'*. College of Forest Resources, University of Washington, Institute of Forest Resources. Contribution 58: 3-16.

Burkhart H. E., Farrar, K.D., Amateis R.L. and Daniels R.F. 1987. " Simulation of individual tree growth and stand development in loblolly pine plantations on cutover, site-prepared areas. " *FWS-1-87. School of Forestry and Wildlife Resources. Virginia Polytechnec Institute and State University.*

Burkhart, H.E.,Cloeren, D.C. and Amateis, R.L. 1985. "Yield relationships in unthinned Loblolly Pine plantations on cutover, site-prepared lands." *South.J.Appl.For.*, 9: 84-91.

Daniels, R.F.,Leuschner, W.A.,Zarnoch, S.J.,Burkhart, H.E. and Hicks, R.R. 1979. "A method for estimating the probability of southern pine beetle outbreaks." *Forest Science,* 25: 265-269.

Daniels, R.F. and Burkhart H.E. 1975 " Simulation of individual tree growth and stand development in managed loblolly pine plantations. " *FWS-5-75. Division of Forestry and Virginia Polytechnic Institute and State University.*

Dudek, A. and Ek, A.R. 1980. "A bibliography of worldwide literature on individual tree based forest stand growth models." Univ.Minn.,Coll.For.,Dep.For.Resour., *Staff Paper Series No 12.*

Ek, A.R. 1980. "A preliminary trial of alternative methods for treating mortality in the multipurpose forest projection system (MFPS) model." Univ.Minn.,Coll.For.,Dep.For.Resor., *Staff Paper Series No 8.*

Gleser, M.A. and Collen, M.F. 1972. "Towards automated medical decisions." *Compt. and Biomed. Res.,* 5: 180-189.

Glover, G.R. and Hool, J.N. 1979. "A basal area ratio predictor of loblolly pine plantation mortality." *Forest Science,* 25: 275-282.

Halperin, M.,Blackwelder, W.C. and Verter, J.I. 1971. "Estimation of the multivariate logistic risk function: a comparison of the discriminant function approach and maximum likelihood approaches." *J.Chron.Dis.,* 24: 124-158.

Hamilton, D.A. 1974. "Event probabilities estimated by regression." *USDA Forest Service.* Research Paper INT-152.

------------ 1980. "Modeling mortality: a component of growth and yield modeling." In *Forecasting Forest Stand Dynamics,* proceedings of the workshop (Kenneth M.Brown and F.Richard Clarke,eds) School of Forestry, Lakehead Univ., Thunder Bay, Ontario : 82-99.

------------ 1986. "A logistic model of mortality in thinned and unthinned mixed conifer stands of northern Idaho." *Forest Science,* 32: 989-1000.

------------ 1988. "SCREEN(F): a Fortran77 program to identify predictors of dichotomous dependent variables." *USDA Forest Service.* Intermt.Res.Station.Research Note INT-383.

Hamilton, D.A. and Wendt, D.L.R. 1975. "SCREEN: a computer program to identify predictors of dichotomous dependent variables." *USDA Forest Service.* General Technical Report INT-22.

Hamilton, D.A. and Edwards, M.B. 1976. "Modeling the probability of individual tree mortality." *USDA Forest Service.* Intermt. Forest and Range Exp. Station.Research Paper INT-182.

Harrell, F.E., Lee, K.L. and McKinnis, R.A. 1980. "Procedures for large regression problems requiring maximum likelihood estimation " *Proceedings of the fifth annual SAS users group international conference, Cary, NC: SAS Institute Inc., 199-202.*

Hatch, C.R. 1971. "Simulation of an even-aged red pine stand in northern Minnesota." Ph.D. Thesis.Univ. Minnesota, St.Paul.

Hett J.M. 1971. "A dynamic analysis of age in sugar maple seedlings. " *Ecology,* 52: 1071-1074.

Hosmer D.W. and Lemeshow S. 1989. " Applied logistic regression. " John Wiley and Sons, New York.

Keister, T.D. 1972. "Predicting individual tree mortality in simulated southern pine plantations." *Forest Science,* 18: 46-51.

Lee, J.Y. 1971. "Predicting mortality for even-aged stands of lodgepole pine." *The Forest Chronicle,* 47: 29-32.

Lenhart D.J., 1972. "Predicting survival of unthinned, old-field loblolly pine plantations." *Journal of Forestry.* 70: 754-755.

Lowell, K.E. and Mitchell, R.J. 1987. "Stand growth projection: simultaneous estimation of growth and mortality using a single probabilistic function." *CanJ.For.Res.,* 17: 1466-1470.

Myers R.H., 1986. "Classical and Modern regression with applications. " Duxbury Press. Boston, Massachusetts.

Myers R.H., 1990. "Classical and Modern regression with applications. "2nd edition. Duxbury Press. Boston, Massachusetts.

Monserud, R.A. 1975. "Methodology for simulating Wisconsin northern hardwood stand dynamics." Ph.D.Thesis. Univ Wisconsin-Madison

------------ 1976. "Simulation of forest tree mortality." *Forest Science,* 22: 438-444.

Moser J.W. 1972. "Dynamics of an uneven-aged forest stand. " *Forest Science,* 18: 184-191.

Nerlove, M. and Press, S.J. 1973. "Univariate and multivariate log-linear and logistic models." Rand Corporation Technical Report R-1306-EDA/NIH, Santa Monica,California

Neter, J. and Maynes, S.E. 1970. "On the appropriateness of the correlation coefficient with a 0,1 dependent variable." *J. Am. Stat. Assoc.,* 65(350): 501-509.

Paille, G. and Smith, J.H.G. 1971. "Uses of simulation in forecasting stand growth and mortality." Proc. of a meeting of a Working Group of the International Union of Forest Research Organizations.Fourteenth Congress,Economics Section, Section 31.London,*Forestry Commission Bulletin 44* :11-17

Payandeh, B. 1983. "Some applications of nonlinear regression models in forestry research" *The Forest Chronicle,* 59: 244-248.

Pregibon D., 1981 "Logistic regression diagnostics. " *The Annals of Statistics* 1981, 9: 705-724.

Reed, D.D.,Burkhart, H.E.,Leuschner, W.A. and Hedden, R.L. 1981. "A severity model for southern pine beetle infestations." *Forest Science,* 27: 290-296.

SAS Institute ,INc.1983. *SUGI supplemental library user's guide.* 1983 Edition SAS Institute Inc., Cary, NC.

Snee, R.D. 1977. "Validation of regression models; methods and examples." *Technometrics,* 19: 415-428.

Somers G.L., Oderwald R.G., Harms W.R. and Langdon G.O., 1980. " Predicting mortality with a Weibull distribution. " *Forest Science.* 26: 291-300.

Stage, A.R. 1973. "Prognosis model for stand development. " *USDA Forest Service* Intermt. Forest and Range Exp Station. Research Paper INT-137.

Sterling, T.D., Randall, G.B., Shelby, H., and Seymour, P.V. 1969. "Robot data screening-a ubiquitous automatic search technique." in; Milton, R. and Nelder, J.(Ed) *Statistical Computation,* : 319-333. Academic Press, New York.

Strub, M.R., Green, E.J., Burkhart, H.E., and Pirie, W.R. 1986. "Merchantability of loblolly pine-an application of nonlinear regression with a discrete dependent variable." *Forest Science,* 32: 254-261.

Walker, S.H. and Duncan, D.B. 1967. "Estimation of the probability of an event as a function of several independent variables." *Biometrika,* 54: 167-179.

Walsh, T.A. 1986. "Diameter/basal area increment equations for loblolly pine trees in cutover, site-prepared plantations." M.S. Thesis, School of For. and Wild. Res.,Va. Polytech. Inst. and State Univ.,Blacksburg.

Weber, L.A, Ek, A.R. and Droessler, T.D. 1986. "Comparison of stochastic and deterministic mortality estimation in an individual tree based stand growth model." *Can.J.For.Res.,* 16: 1139-1141.

Yang Y.C., 1988. " Generalized logistic mortality model for natural forests and plantations in Taiwan. " *Forest Simulation Systems. Proceedings of the IUFRO Conference.* Berkeley, California. November 2-5,1988.

Appendix A

Plots of distribution of dead trees

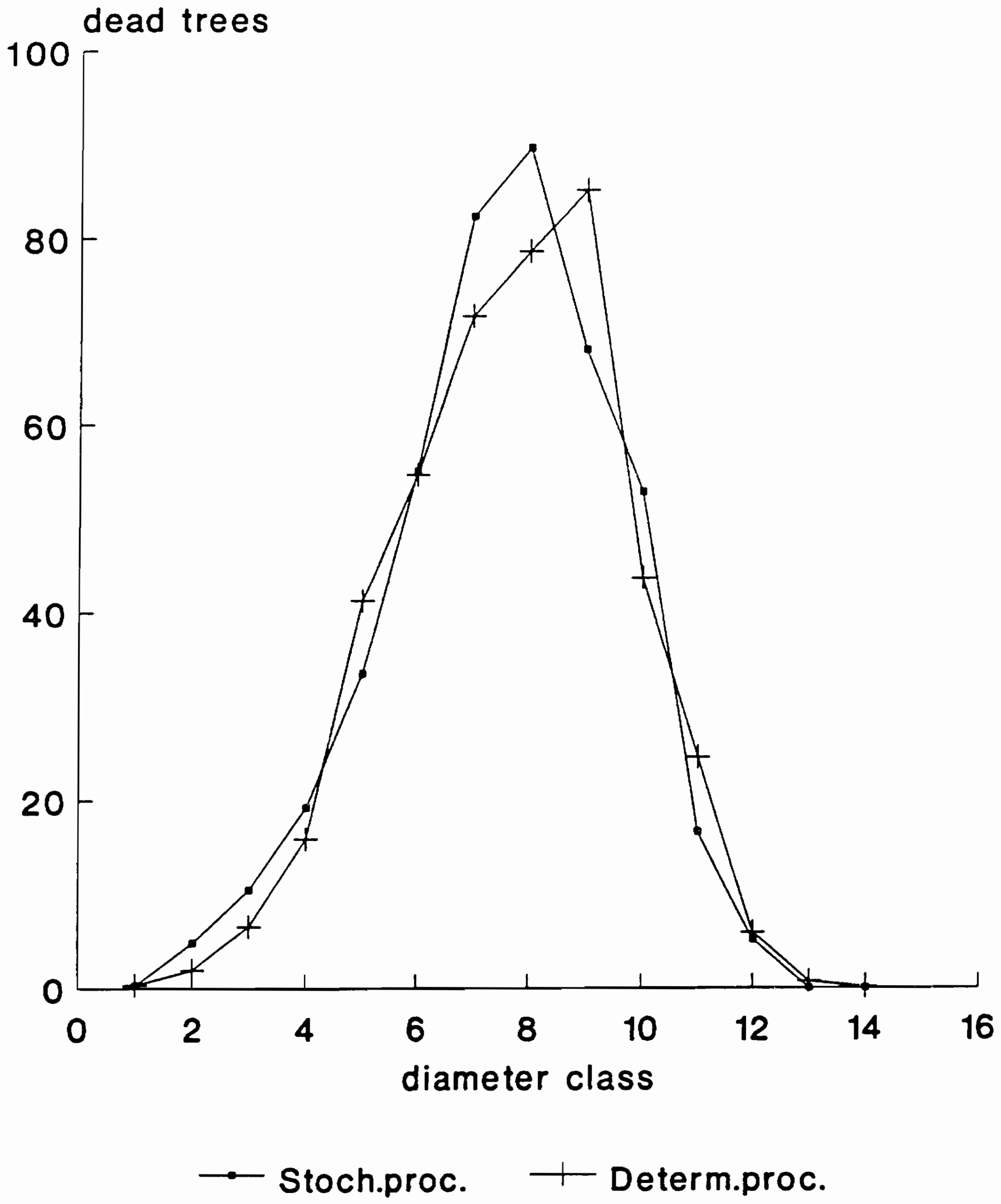


Figure 9. Distribution of dead trees. (SI = 50, N = 800)

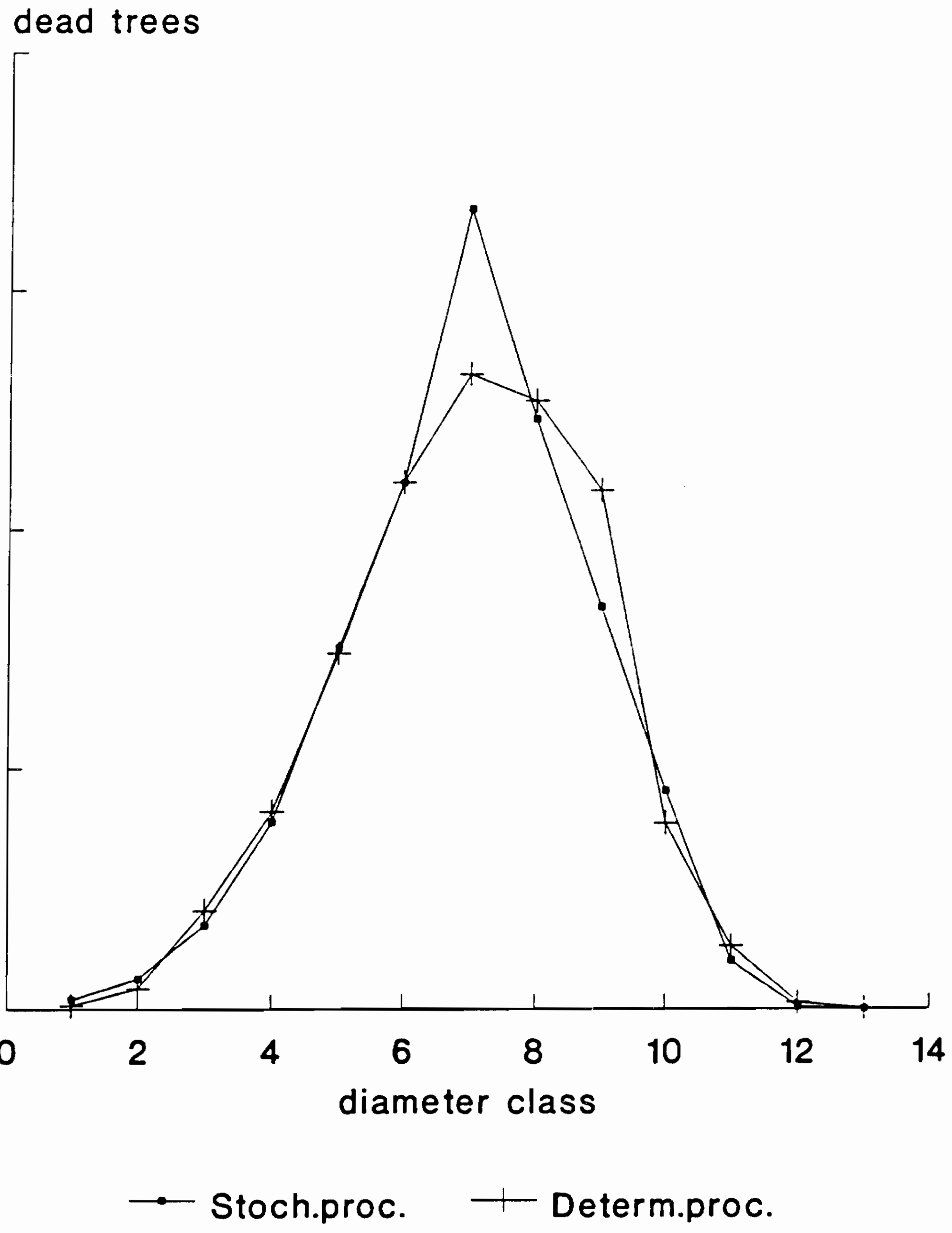


Figure 10. Distribution of dead trees. (SI = 50, N = 1200)

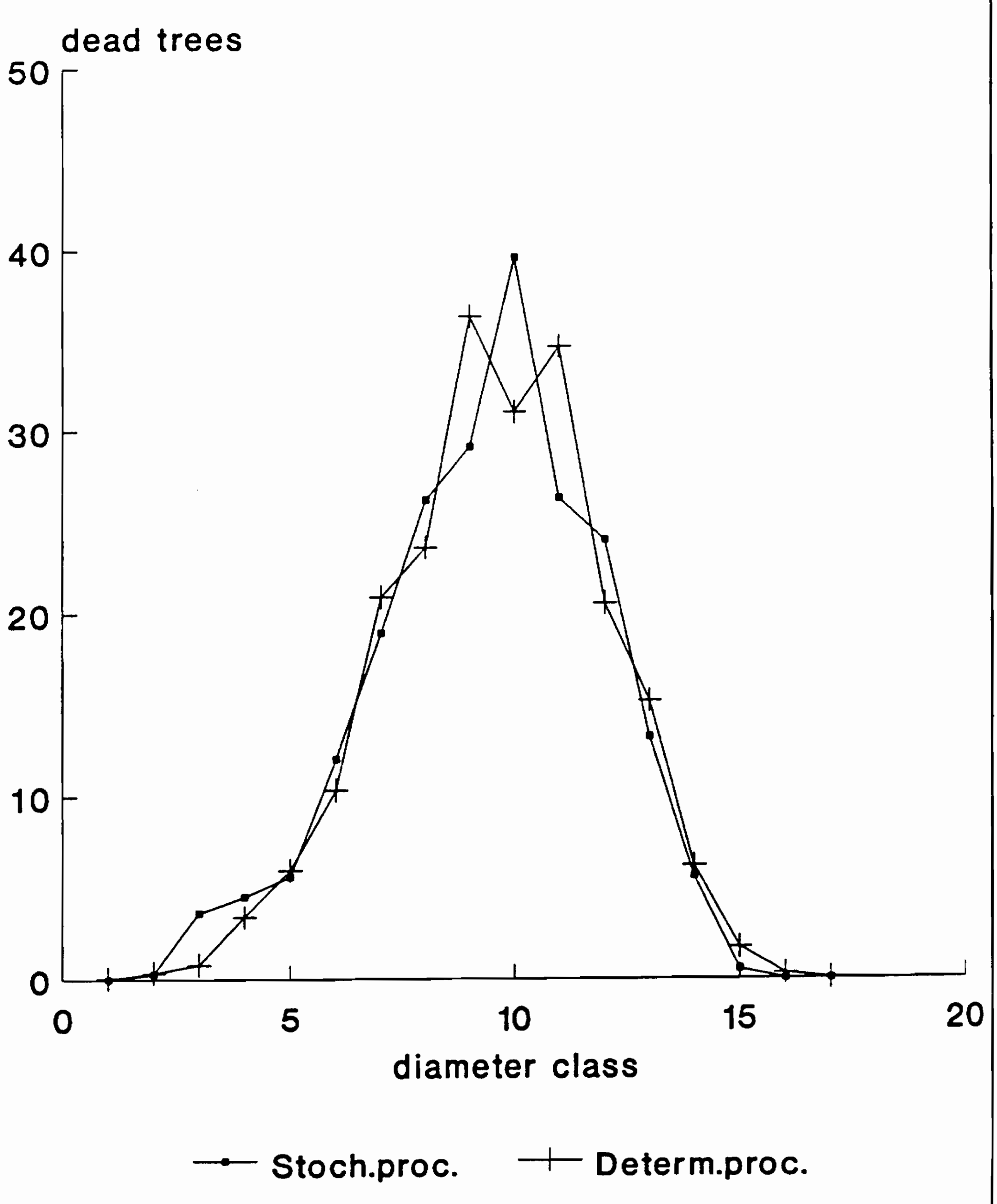


Figure 11. Distribution of dead trees. (SI = 60, N = 400)

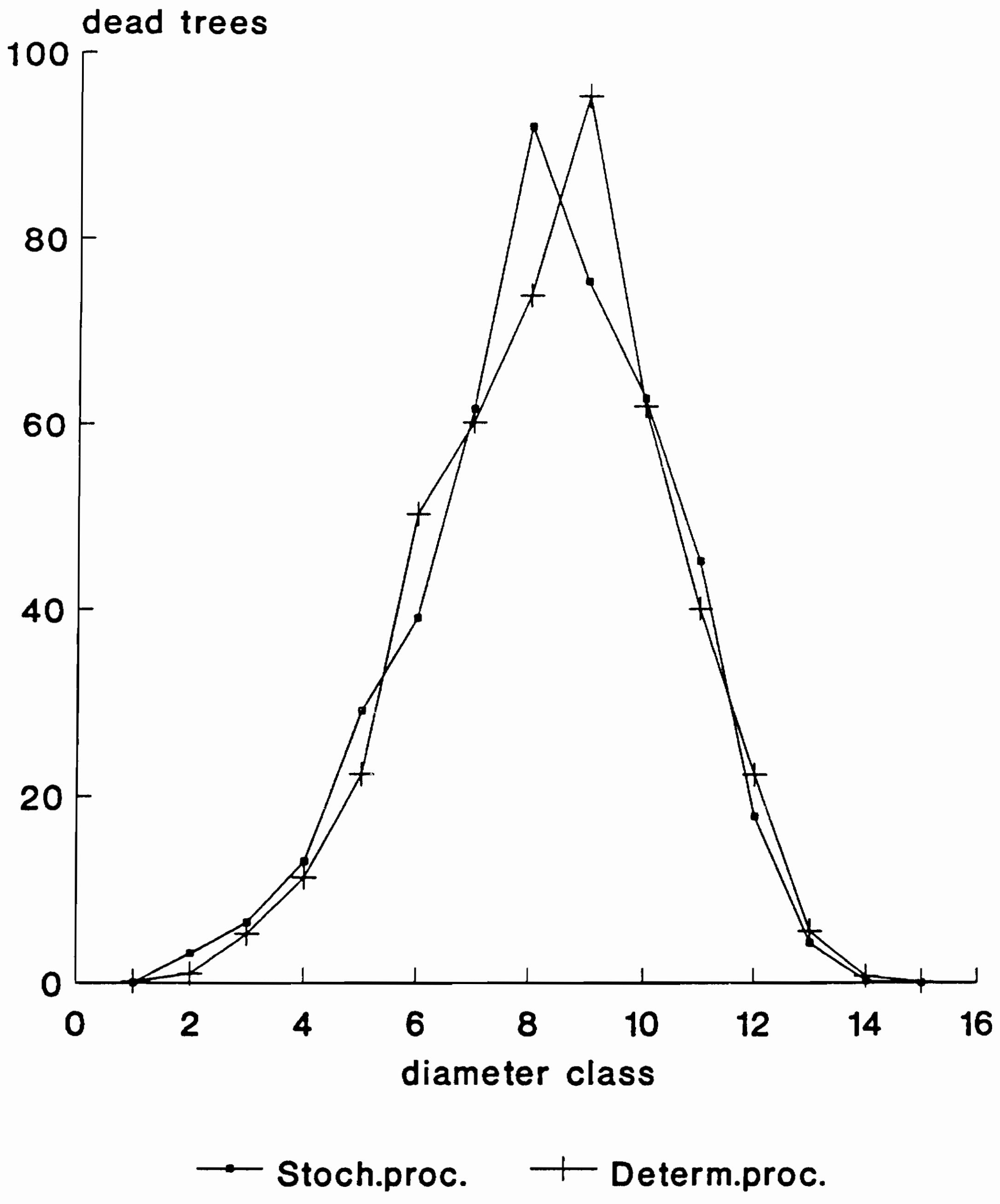


Figure 12. Distribution of dead trees. (SI = 60, N = 800)

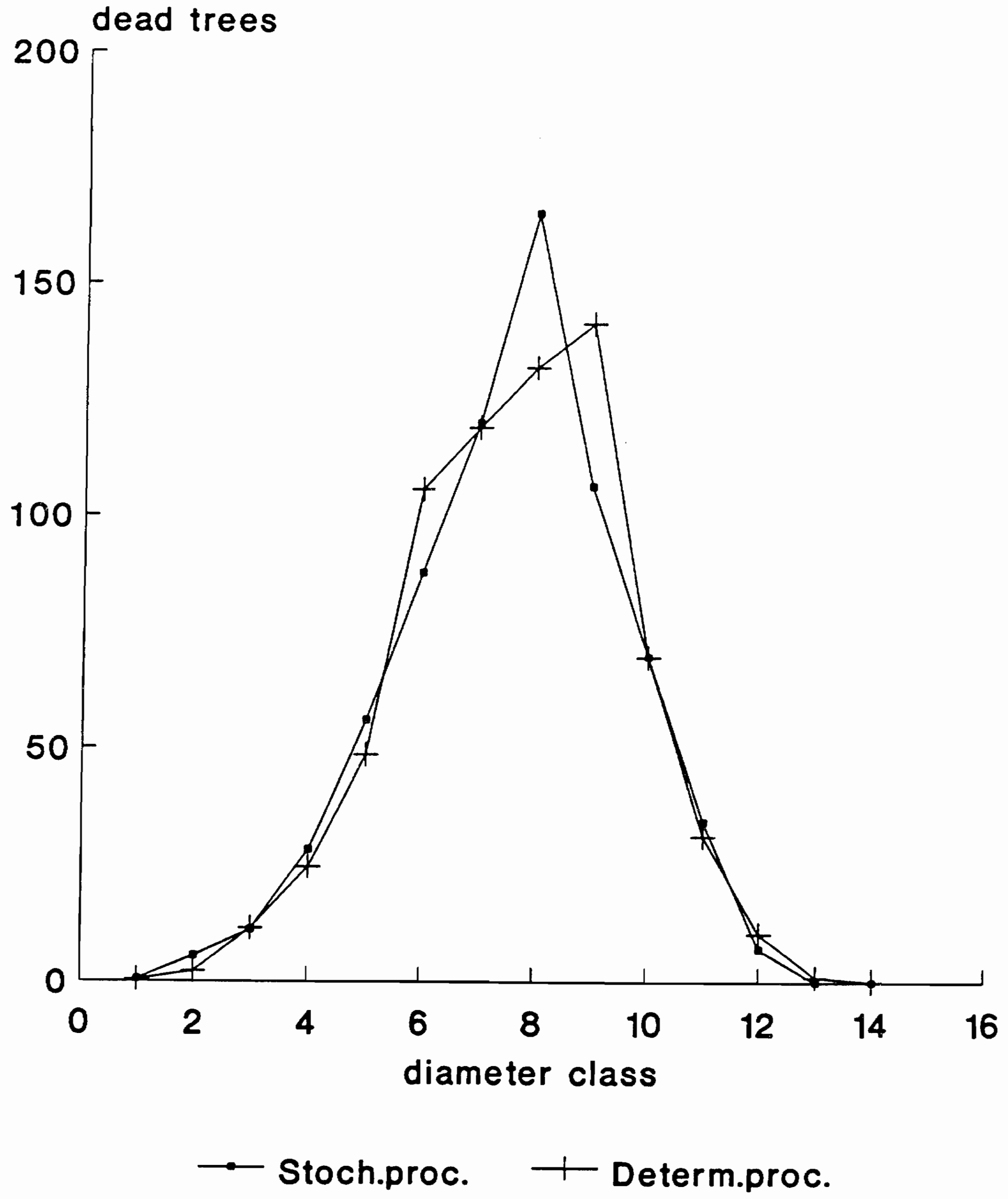


Figure 13. Distribution of dead trees. (SI = 60, N = 1200)

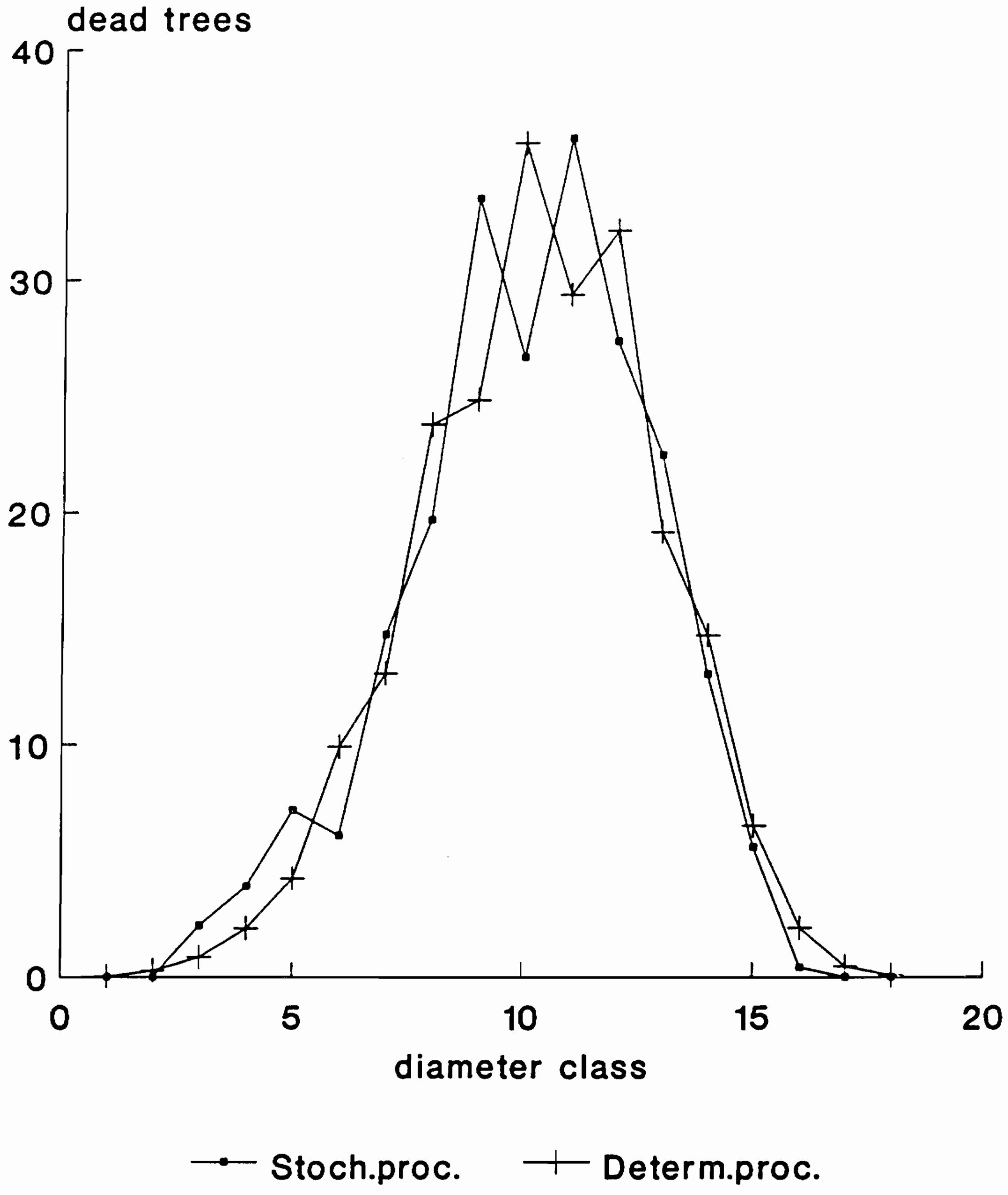


Figure 14. Distribution of dead trees. (SI = 70, N = 400)

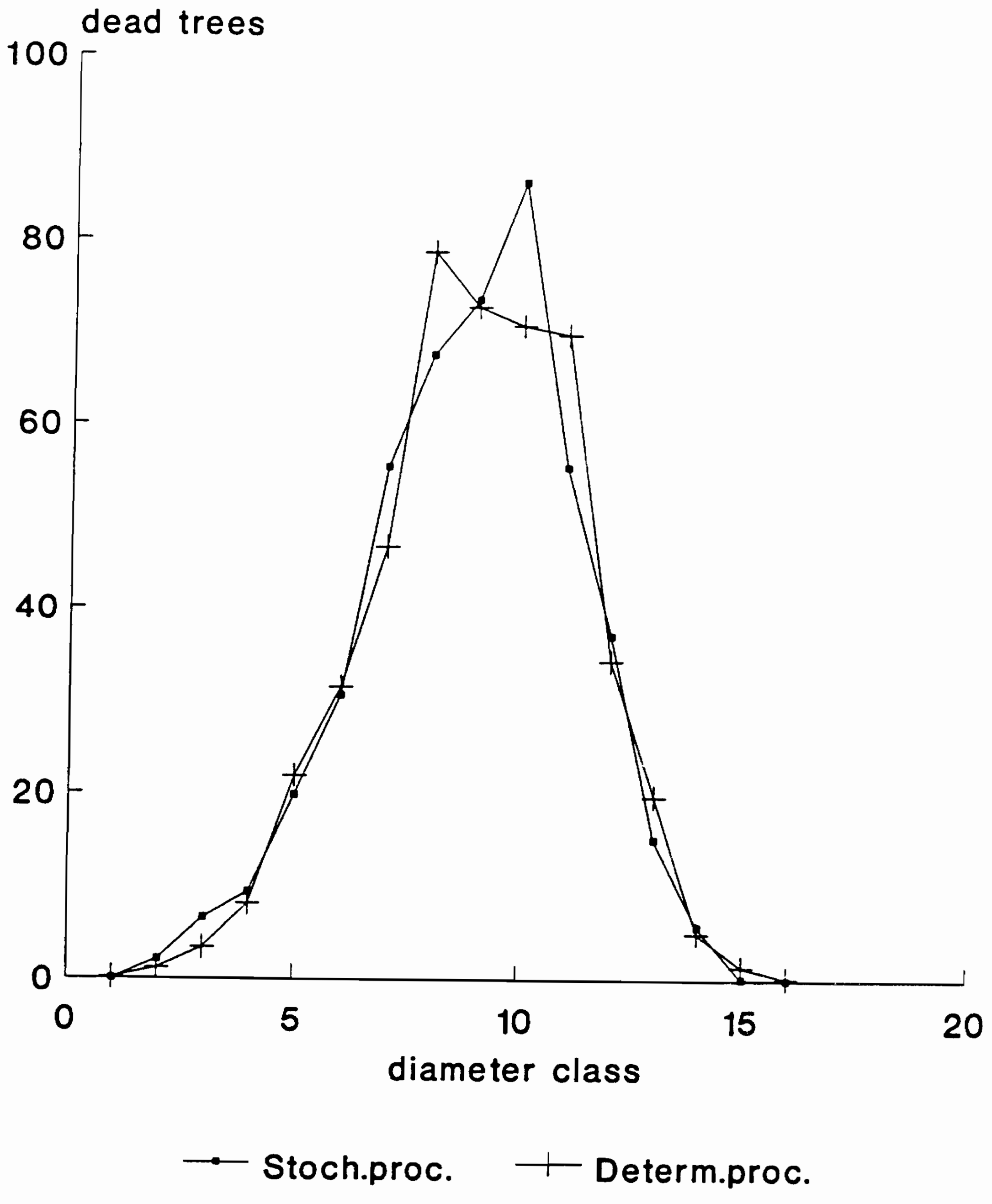


Figure 15. Distribution of dead trees. (SI = 70, N = 800)

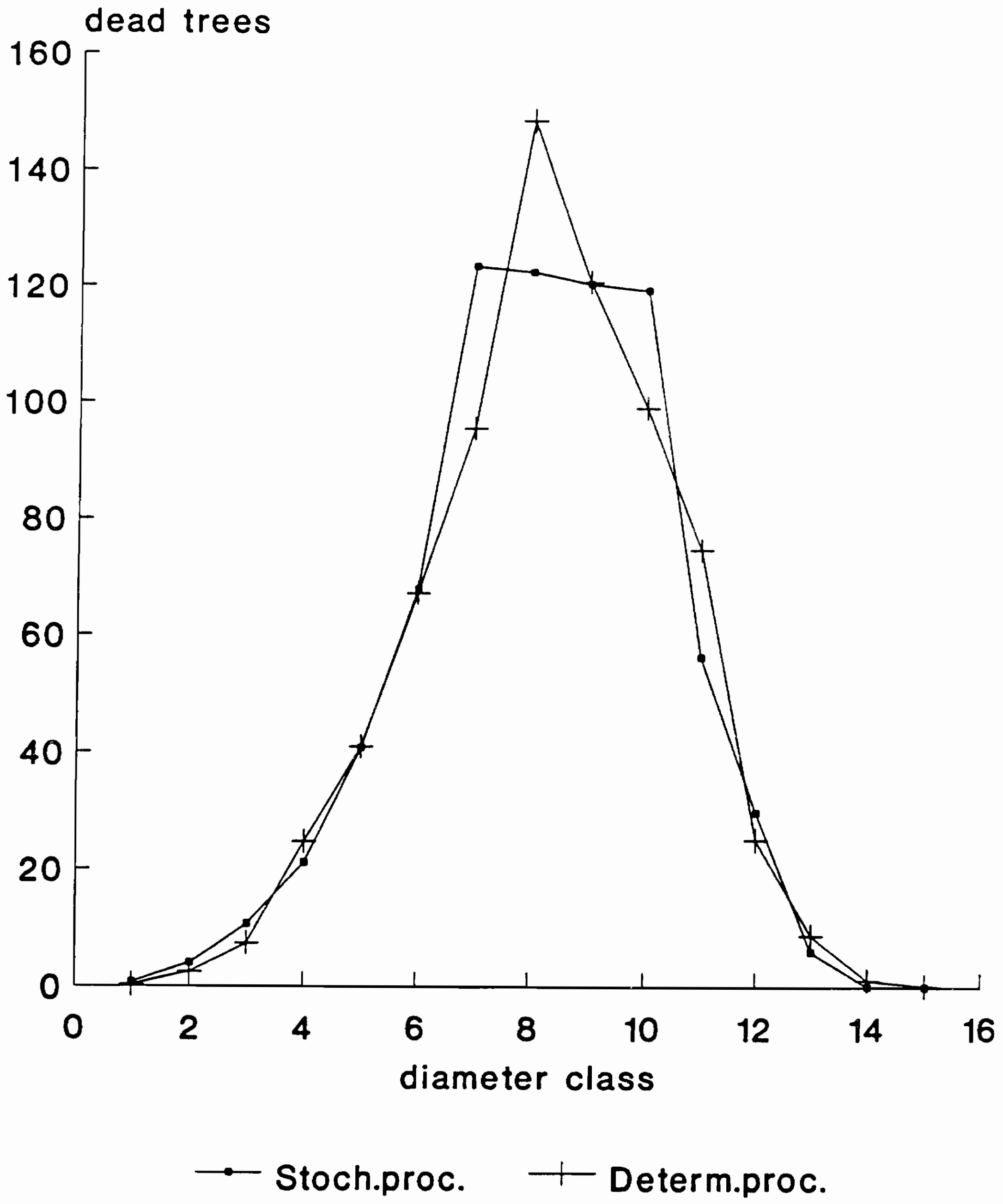


Figure 16. Distribution of dead trees. (SI = 70, N = 1200)

Appendix B

Plots of distribution of live trees

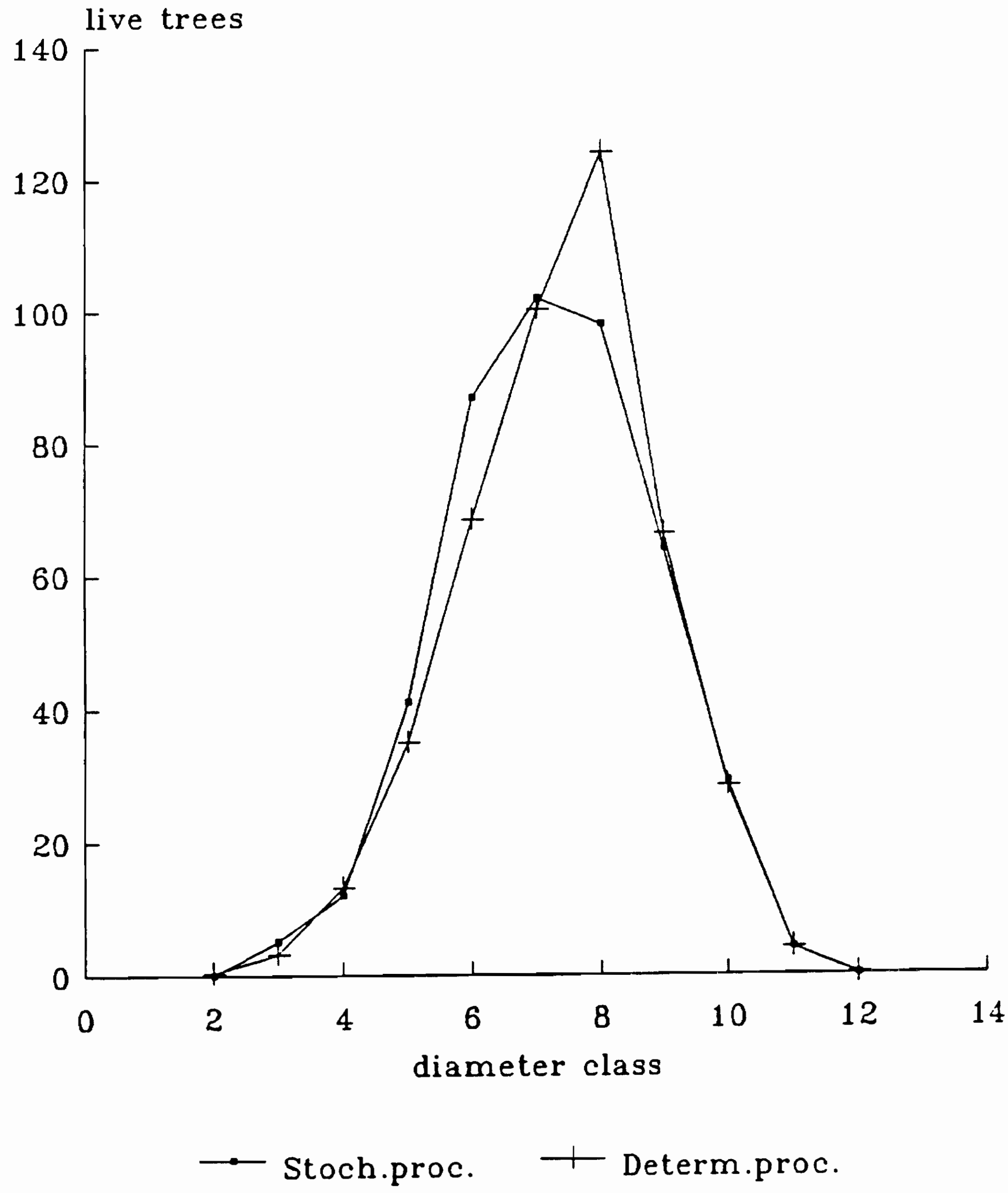


Figure 17. Distribution of live trees. (SI = 50, N = 800)

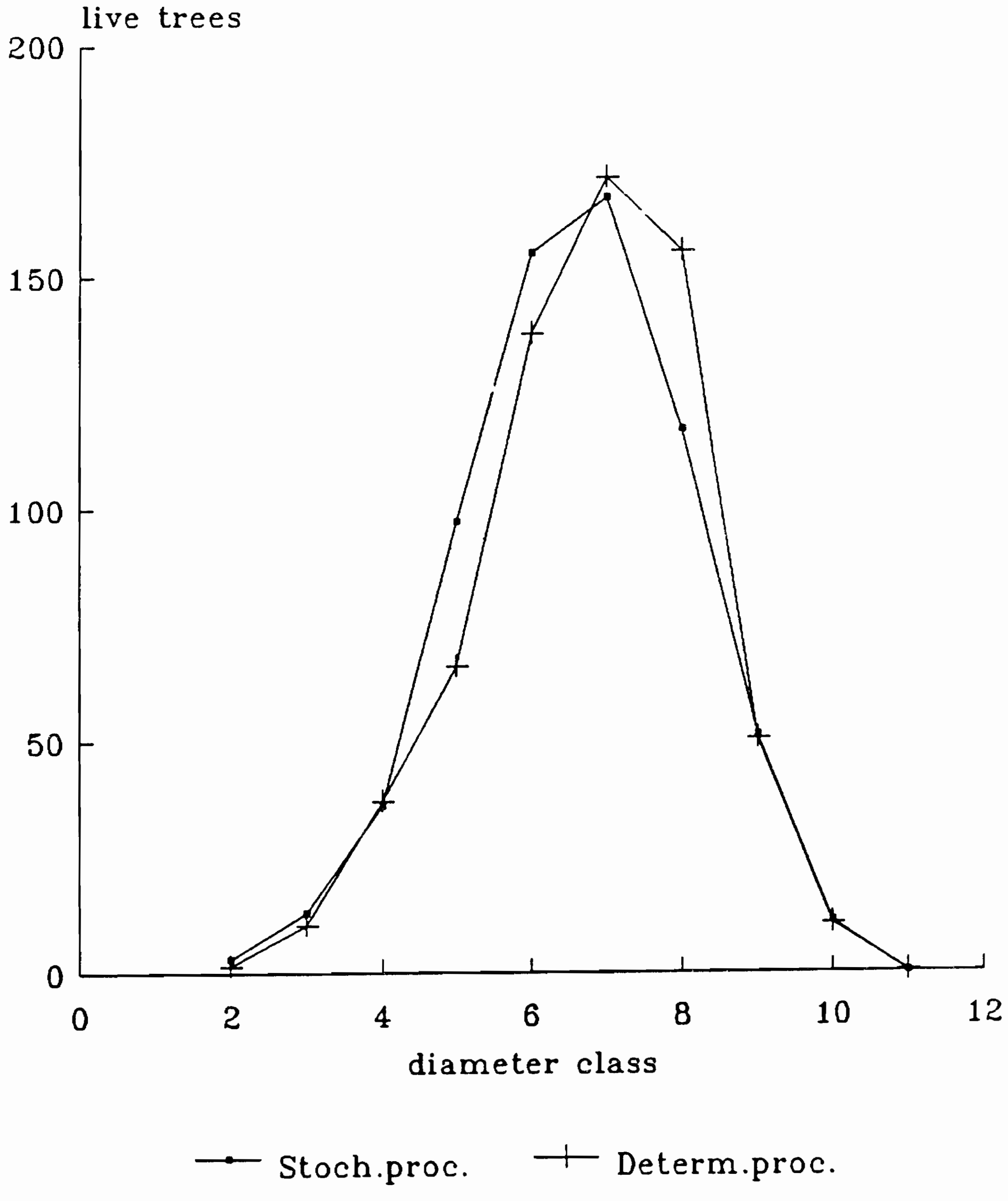


Figure 18. Distribution of live trees. (SI = 50, N = 1200)

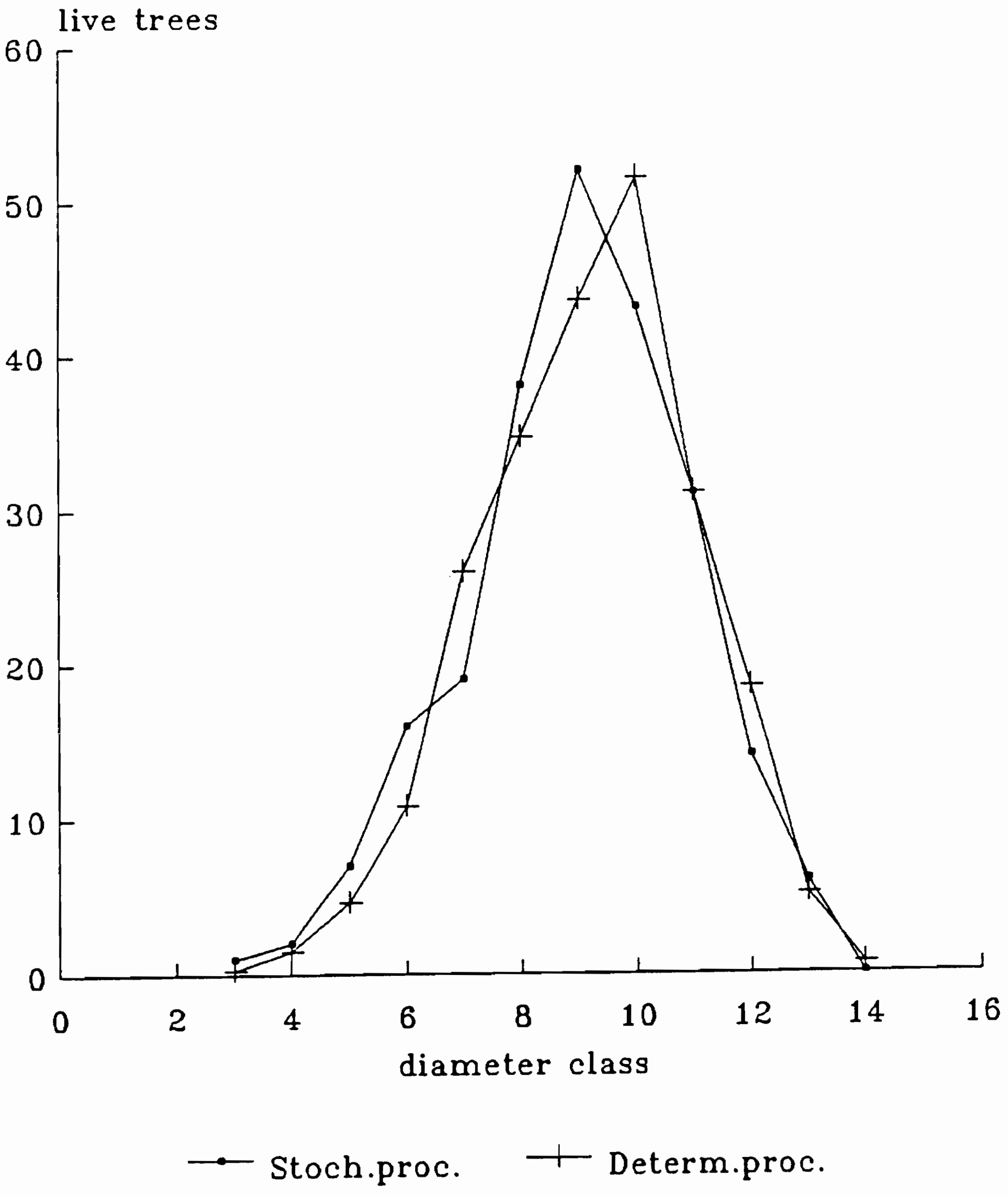


Figure 19. Distribution of live trees. (SI = 60, N = 400)

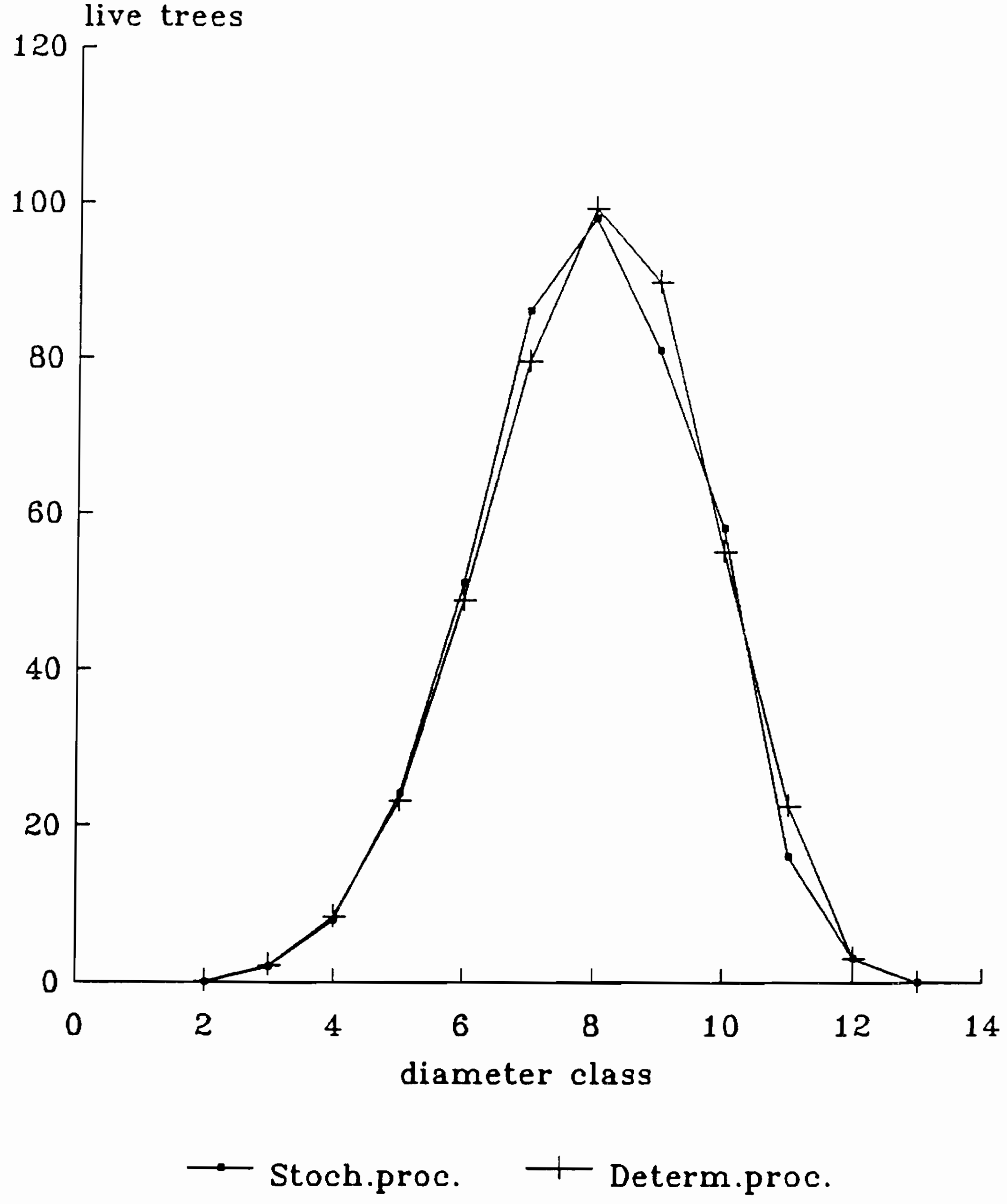


Figure 20. Distribution of live trees. (SI = 60, N = 800)

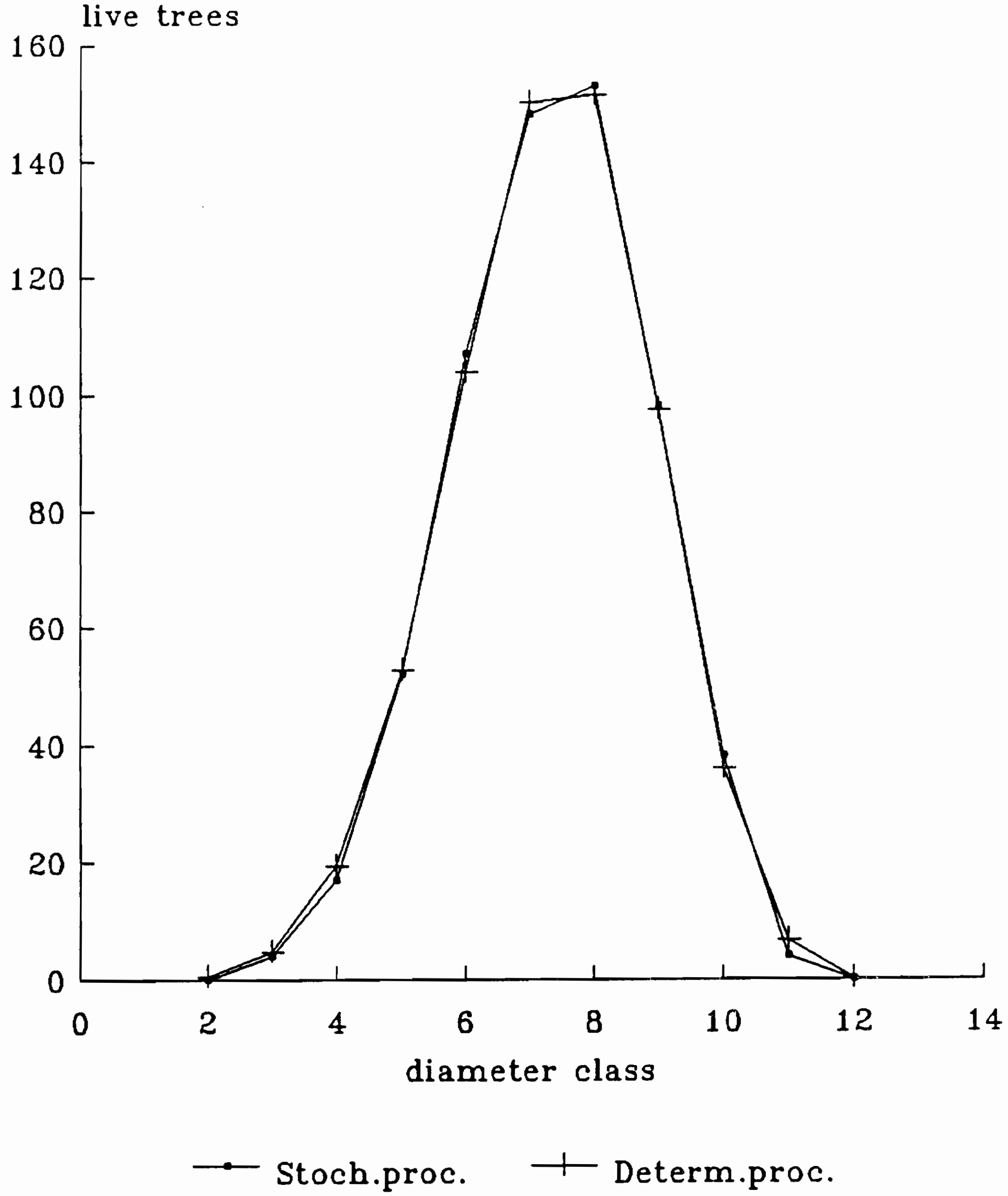


Figure 21. Distribution of live trees. (SI = 60, N = 1200)

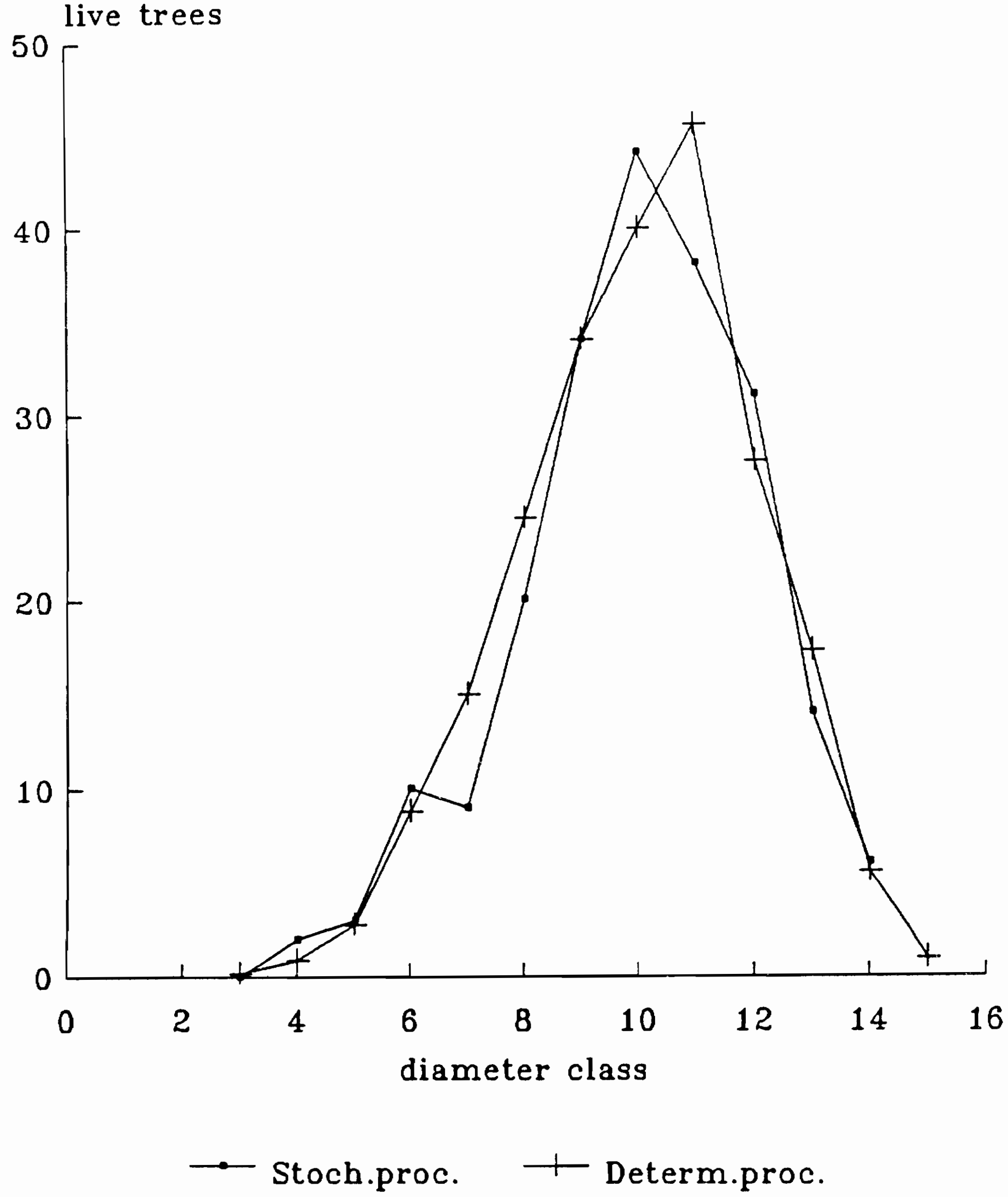


Figure 22. Distribution of live trees. (SI = 70, N = 400)

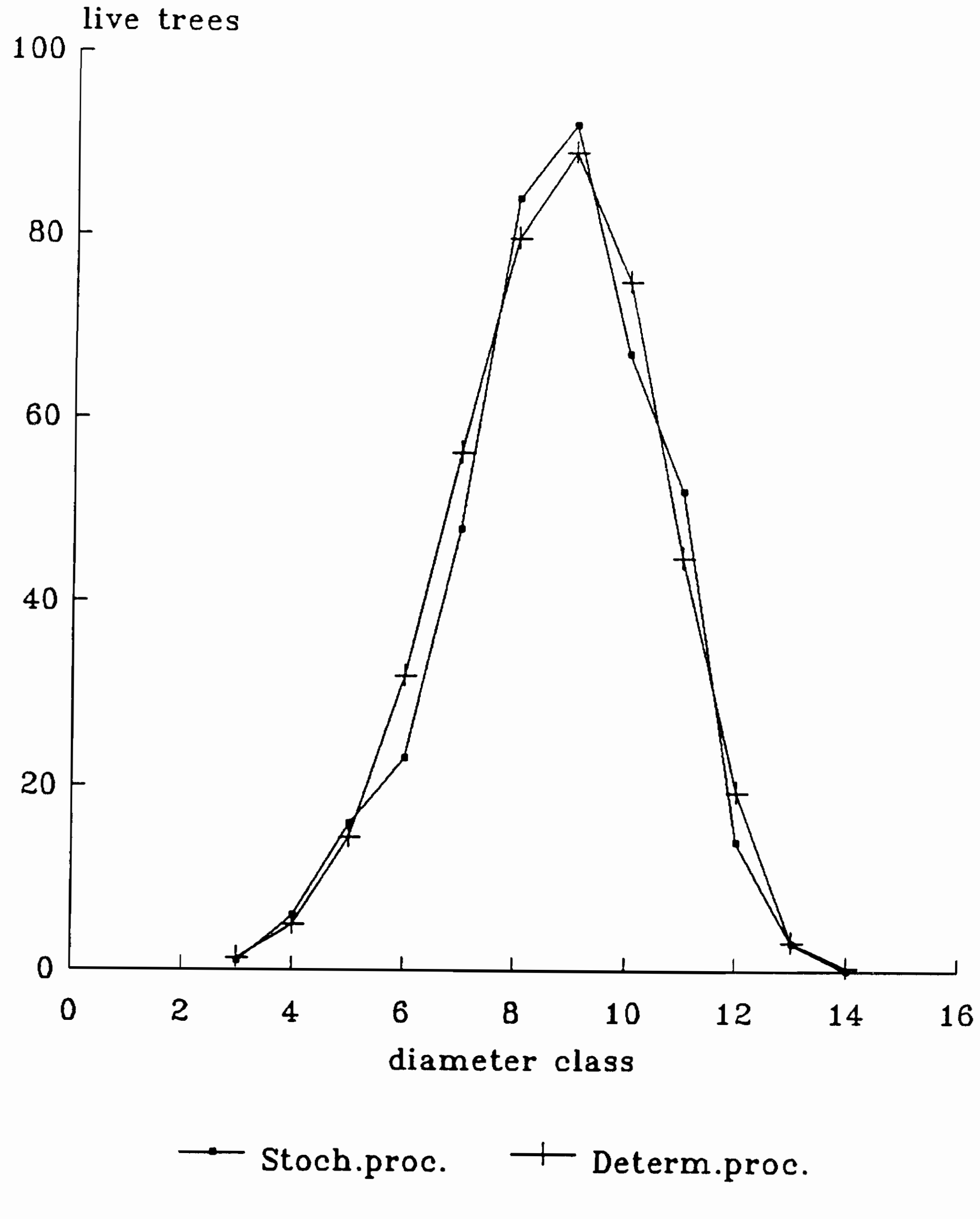


Figure 23. Distribution of live trees. (SI = 70, N = 800)

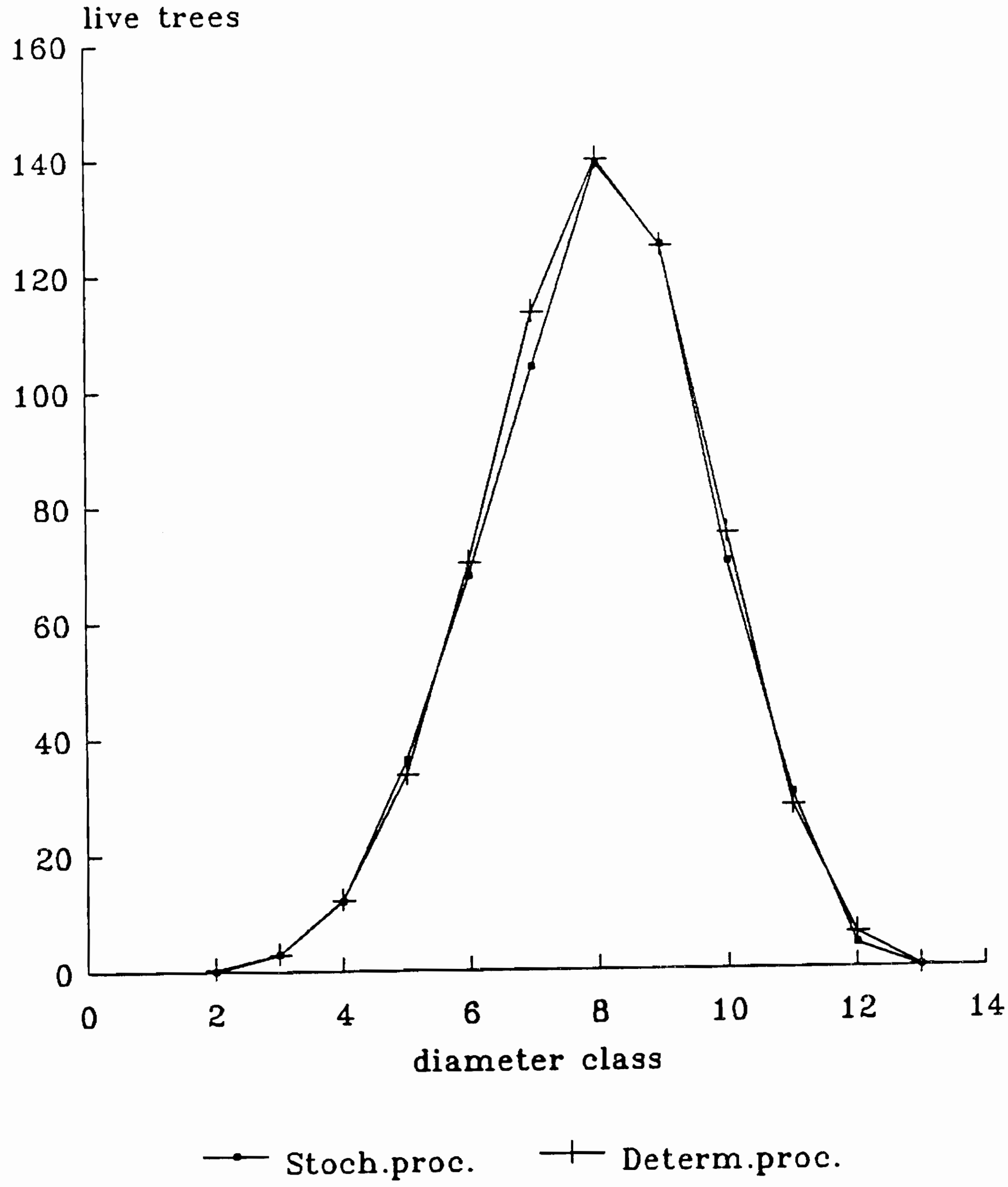


Figure 24. Distribution of live trees. (SI = 70, N = 1200)

Vita

Olga Beatriz Avila was born on February 25, 1960 in Buenos Aires, Argentina. She attended the National University of Santiago del Estero, Argentina where she received a Bachelor of Science degree in Agronomical Engineering in December 1983. Upon graduation she was hired by the National University of Santiago del Estero where she worked as an instructor and research engineer in statistics applied to biology for three years. During 1985 the National Institute of Statistical Research of Tucuman, Argentina granted her a fellowship to take graduate courses in applied statistics. In June 1988 she entered Virginia Polytechnic Institute and State University where she earned her Master of Science degree in Forest Biometrics in August 1990.

OB Avila de Spies