UNIVERSITI PUTRA MALAYSIA

# DEVELOPMENT OF DIAMETER DISTRIBUTION YIELD PREDICTION MODELS FOR SIMULATION OF *ACACIA MANGIUM* PLANTATIONS

KAMZIAH ABD KUDUS

FSAS 1998 36

# DEVELOPMENT OF DIAMETER DISTRIBUTION YIELD PREDICTION MODELS FOR SIMULATION OF *ACACIA MANGIUM* PLANTATIONS

**By**

**KAMZIAH ABD KUDUS**

**Thesis Submitted in Fulfilment of the Requirements for the Degree of Master of Science in the Faculty of Science and Environmental Studies Universiti Putra Malaysia**

**June 1998**

# DEDICATION

*.... In loving memory of my mother*
*ZAINAB SHAMSUDIN (1944-1970)*

*Always in my thoughts.*

## ACKNOWLEDGEMENTS

First and foremost I would like to express my greatest gratitude to Allah S.W.T. for giving me the strength, courage, and time without which this study could not be completed.

I especially thank my supervisor, Assoc. Prof. Dr. Muhammad Idrees Ahmad for his assistance, constructive advice, commitment and constant encouragement throughout the duration of this study.

I would also like to extend my thanks to the supervisory committee members, Prof. Dr. Nik Muhammad Nik Abd Majid, Assoc. Prof. Dr. Isa Daud and Dr. Awang Noor Abd. Ghani for their review, valuable suggestions and criticisms on the drafts of this manuscript.

Thanks are also due to Assoc. Prof. Dr. Harun Budin, Head of Department of Mathematics and Dr. Jamaluddin Basharuddin for introducing Forest Biometrics to me. Grateful acknowledgements are made to Forest Research Centre, Sandakan, Sabah particularly to Mr. Jaffirin Lapongan, head of the Forest Plantation Section and Mr. Lenim for collections of diameter data of the *Acacia mangium* plantation. I am also grateful to Mr. Ahmad Zuhaidi Yahya, Senior Research Officer of the Forest Plantation Unit, Forest Research Institute of Malaysia (FRIM) for preparation of diameter data from mixed timber species.

My special thanks go to Mr. Arif Idris, Ms. Rosezita Zawawi, Mr. Karim Juned and Mr. Rahmat Maroof of Computer Centre, Universiti Putra Malaysia for providing computer facilities, and to Mrs. Zainaf Kadir for assistance with the typing of this thesis.

I sincerely thank my husband, Abd Talib Dollah, my son Muhammad Amin Asyraf and my father for their love, patience and constant encouragement.

Finally I would like to thank SEAMEO SEARCA for the award of scholarship and Universiti Putra Malaysia for granting me study leave.

## TABLE OF CONTENTS

B ADDITIONAL TABLES

# LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| cdf | cumulative distribution function |
| pdf | probability density function |
| invcdf | inverse cumulative distribution function |
| psp | permanent sample plot |
| dbh | diameter at breast height |
| EDF | empirical distribution function |
| FPP | four percentile points |
| KNB | Knoebel-Burkhart |
| MMLE | modified maximum likelihood |
| LR | linear regression |
| MLE | maximum likelihood estimation |
| NLR | nonlinear regression |
| RMSE | root mean square error |
| $R^2$ | coefficient of determination |
| SE | standard error |
| PPM | parameter prediction method |
| PRM | parameter recovery method |
| PCPM | percentile prediction method |

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Master of Science.

# DEVELOPMENT OF DIAMETER DISTRIBUTION YIELD PREDICTION MODELS FOR SIMULATION OF *ACACIA MANGIUM* PLANTATIONS

By

**KAMZIAH ABD KUDUS**

**June 1998**

**Chairman:** **Associate Professor Dr. Muhammad Idrees Ahmad**

**Faculty:** **Science and Environmental Studies**

The purpose of this study is to develop diameter distribution yield prediction models for predicting probability density function parameters when age, spacing and number of trees per hectare planted are known. Five distributions, Weibull, Gamma, Johnson $S_B$, Lognormal and Generalised Normal were compared in terms of their ability to model diameter data in uneven-aged and even-aged forest stands. The classical moments were applied as a measure of flexibility of the distribution in regard to their changes in shape. Diameter data were obtained from 16 uneven-aged stands of mixed timber species located at Bukit Lagong Forest Reserve, Kepong, Selangor and 14 even-aged stands of *Acacia mangium* located at Segaliud Lokan Project, Sandakan, Sabah. The stands were all plantations and the ages range from 2 to 22 years. The diameter data were fitted to the five distributions by the maximum likelihood estimation method. The Johnson $S_B$ distribution showed the best performance in

terms of quality of fit to the diameter data based on relative ranking of the log likelihood criterion. The estimation of Johnson $S_B$ distribution was further investigated and the nonlinear regression method was proposed for the estimation of the $S_B$ parameters. This method was compared to five other estimation methods; namely the four percentile points method, Knoebel-Burkhart method, linear regression method, maximum likelihood method, and modified maximum likelihood method through simulation. The performance of the nonlinear regression was confirmed by using the real diameter data. Goodness-of-fit tests based on empirical distribution function (namely the Kolmogorov-Smirnov statistic, Cramer-von Mises statistic and the Anderson-Darling statistic) were used in selecting the most superior parameter estimation method. The results suggested that the nonlinear regression method was superior for estimating parameters of the Johnson $S_B$ distribution for the diameter data. In order to simulate the stand characteristics, equations were developed for predicting average height, basal area per hectare, and number of trees per hectare surviving when age, spacing and number of trees per hectare planted were known. The predicted stand characteristics were then related to the estimated parameters of the Johnson $S_B$ and solving the resulting set of equations for the scale and shape parameters. This study revealed that the parameter prediction method yields reliable prediction equations of the stand characteristics, but the prediction equations of the scale and shape parameters suggested that further research is needed to improve the model.

Abstrak yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains.

**PEMBENTUKAN MODEL RAMALAN HASIL TABURAN PEREPANG BAGI SIMULASI PERLADANGAN *ACACIA MANGIUM***

Oleh

**KAMZIAH ABD KUDUS**

**Jun 1998**

**Pengerusi :** **Profesor Madya Dr. Muhammad Idrees Ahmad**

**Fakulti :** **Sains dan Pengajian Alam Sektiar**

Tujuan kajian ini adalah untuk membentuk sebuah model ramalan hasil taburan perepang bagi meramalkan parameter sesebuah fungsi ketumpatan kebarangkalian apabila umur, jarak dan bilangan pokok ditanam per hektar diketahui. Lima taburan terdiri daripada taburan Weibull, Gamma, Johnson $S_B$, Lognormal dan Normal Teritlak dibandingkan dari segi keupayaannya untuk membentuk model data perepang dari dirian hutan tak seumur dan seumur. Momen klasikal digunakan sebagai satu ukuran kelenturan berhubung dengan perubahan dalam bentuk sesebuah taburan itu. Data perepang diperolehi daripada 16 dirian tak seumur dari pelbagai spesis terletak di Hutan Simpan Bukit Lagong, Kepong, Selangor dan 14 dirian seumur dari spesis *Acacia mangium* terletak di Projek Segaliud Lokan, Sandakan, Sabah. Kesemua dirian tersebut adalah hutan ladang dan umumnya di dalam lingkungan 2 hingga 22 tahun. Kelima-lima taburan yang digunakan disesuaikan kepada data perepang menggunakan kaedah penganggar kebolehjadian maksimum.

Taburan Johnson $S_B$ mempamerkan prestasi terbaik dari segi kualiti kesesuaian terhadap data perepang berdasarkan kepangkatan relatif kriteria kebolehjadian log. Penganggaran taburan Johnson $S_B$ disediliki seterusnya dan kaedah regresi tak linear dicadangkan bagi menganggar parameter $S_B$. Kaedah ini dibandingkan dengan lima kaedah penganggaran yang lain, iaitu, kaedah empat titik persentil, kaedah Knoebel-Burkhart, kaedah regresi linear, kaedah kebolehjadian maksimum dan kaedah kebolehjadian maksimum yang diperbaiki melalui satu kajian simulasi. Prestasi regresi tak linear telah diperkukuhkan dengan menggunakan data perepang yang sebenar. Ujian kebagusan penyuaian berdasarkan fungsi taburan empirik (yang dinamakan statistik Kolmogorov-Smirnov, statistik Cramer-von Mises dan statistik Anderson-Darling) telah digunakan untuk memilih kaedah penganggaran parameter yang terbaik. Keputusan mencadangkan kaedah regresi tak linear adalah yang terbaik untuk menganggar parameter taburan Johnson $S_B$ berdasarkan data perepang. Untuk mensimulasi cirian dirian, persamaan dibentuk untuk meramal purata ketinggian, luas pangkal per hektar dan bilangan pokok terus hidup per hektar apabila umur, jarak dan bilangan pokok ditanam per hektar diketahui. Cirian dirian yang telah diramalkan akan dikaitkan dengan parameter Johnson $S_B$ yang telah dianggarkan dan seterusnya menyelesaikan persamaan yang terhasil untuk mendapatkan parameter bentuk dan skala. Kajian ini menunjukkan bahawa kaedah ramalan parameter menghasilkan persamaan ramalan cirian dirian yang boleh dipercayai, tetapi persamaan ramalan bagi parameter bentuk dan skala mencadangkan kajian seterusnya diperlukan untuk memperbaiki model ini.

# CHAPTER I

# INTRODUCTION

The evolution of biometry in particular, and statistical practice in general has been derived by the needs of biologists to solve problems and to answer questions arising from day-to-day applications in the discipline in which they labour. This certainly has been true in forestry where reliable quantitative information is required to manage forestlands. The quantitative information includes growth and yield, area logged and tree volume. This study gives an overview of statistical procedures used to describe diameter distribution, estimate of parameters based on diameter and predict the parameters of the distribution for a plantation stand at given values of age, spacing and density. All of the three areas of discussion are long-standing subjects of concern in forestry and are likely to remain so wherever forests are actively managed and regulated on a sustainable basis. Because of the needs for increasingly precise information about forest resources and their growth pattern, improved statistical analysis and understanding remain an area of priority research and development globally.

## Research and Development on Modelling Diameter Distribution

Many statistical distributions have been used to describe diameter distributions in forest stands with varying degrees of success. Virtually, in all applications, the diameter distributions have been characterised by continuous, unimodal pdf's. The distributions used include the normal, lognormal, exponential, gamma, beta, Weibull, and Johnson $S_B$ distribution, with the beta, Weibull, and $S_B$ distributions being most widely applied (Cao, 1997). The lognormal distribution was first applied in describing diameter distribution in forest stands by Bliss and Reinker (1964). Nelson (1964) applied the gamma distribution while Clutter and Bennet (1965) used the beta distribution to describe diameter distributions. Bailey and Dell (1973) proposed the Weibull distribution and Hafley and Schreuder (1977) found the $S_B$ distribution to be appropriate for describing both diameter and height distributions. Where later on these two distributions became the driving force in numerous growth and yield systems (Smalley and Bailey, 1974a; Feducia *et al.*, 1979; Matney and Sullivan, 1982; Baldwin and Feduccia, 1987; Brooks *et al.*, 1992; Hafley *et al.*, 1982).

This study emphasizes on the Johnson $S_B$ distribution for development and comparison of parameter estimation techniques. Many methods have been developed to estimate the parameters of the Johnson $S_B$ distribution, such as the four percentile points method (Slifker and Shapiro, 1980), the Knoebel-Burkhart method (Knoebel and Burkhart, 1991), the maximum likelihood estimation method (Johnson, 1949), mode method (Hafley and Buford, 1985), and linear regression method (Zhou and Mc Tague, 1996). This study proposes a new method based on nonlinear regression to estimate the parameters of the Johnson $S_B$ distribution based on diameter data of a plantation forest.

When probability distributions are used to construct diameter distribution models, the parameters of the probability distribution are usually related to stand variables by a relatively complicated fitting process. Henceforth, regression equations are developed to relate these parameter estimates to stand characteristics such as age, spacing and number of trees per unit area. The parameter prediction method (Reynolds *et al.*, 1988) is applied to develope the regression equations.

## Objectives

Growth and yield models assist forest researchers and managers in many ways including the ability to predict future yields and to explore silvicultural options. Models provide an efficient way to prepare resource forecasts, but more importantly their ability to explore management options and silvicultural alternatives. Foresters are able to predict the long-term effect on the forest or future harvests or both, or a particular silvicultural decision, such as changing the cutting limits for harvesting. Hence, a growth model could assist foresters to examine the likely outcomes, both with the intended and alternative cutting limits, and can make their decision objectively.

This study concentrates on describing the frequency distribution of diameter measurements in plantation forest using pdf, estimating the parameters of the distribution function based on diameters and predict the pdf parameters at given values of stand variables.

The objectives of this study are to

1. Model diameter distribution in forest stand and to verify the distribution that gives the best performance in terms of quality of fit to the variety of sample distributions based on diameter data of mixed species gathered from Forest Research Institute of Malaysia and diameter data of *Acacia mangium* plantation collected from Forest Research Center, Sandakan, Sabah.

2. Find a more reliable and stable method of parameter estimation based on diameter data of *Acacia mangium* plantation in Sabah.

3. Estimate the average height per unit area at any given values of age and spacing. The basal area and number of trees per unit area are then estimated from given values of age, spacing and average height per unit area. Subsequently, to develop regression equations via the parameter prediction method by relating the estimates of the parameters to stand characteristics such as the age, spacing, average height, basal area and number of trees per unit area.

4. Determine alternative method for prediction of the parameters at given stand characteristics.

Therefore, the results of this study are to help forest managers to estimate yield of a forest stand with various combinations of age, spacing and average height. Thus, yield prediction models can assist forest managers to minimize time, costs and the risk of obsolescence since field experiments are exposed to the risk and uncertainties of environmental and economic constraints.

## Thesis Planning

Chapter II emphasizes on developing diameter distribution models based on pdf of Gamma (GM), Generalised Normal (GN), Johnson $S_B$ ($JS_B$), Lognormal (LN) and Weibull (WB) distributions. The qualities of fit of the diameter distribution models are compared based on the log likelihood criterion.

In Chapter III, six parameter estimation methods (namely the four percentile points (FPP), Knoebel-Burkhart (KB), maximum likelihood estimation (MLE), modified maximum likelihood estimation (MMLE), linear regression (LR), and nonlinear regression (NLR) methods) are applied to estimate the parameters at the best fitting diameter distribution model. Simulation experiments are conducted to identify the most efficient method of parameter estimation. The efficiency of the methods is compared on the basis of bias and root mean square error.

The parameter estimation methods are also applied on diameter data of *Acacia mangium* from Sabah. Several goodness-of-fit tests based on empirical distribution function (EDF); namely the Kolmogorov-Smirnov (D) statistic, Cramer-von Mises ($W^2$) statistic, and Anderson-Darling ($A^2$) statistics are applied to evaluate the performance of the estimation methods.

Chapter IV highlights the regression equations developed in this study, which is used to estimate the stand characteristics from a given age, spacing and number of trees per hectare

to the stand characteristics to predict the scale and shape parameters.

Chapter V gives the summary of this study and some suggestions for further research.

Figure 1 illustrates basic concept of the development of diameter distribution yield prediction models.

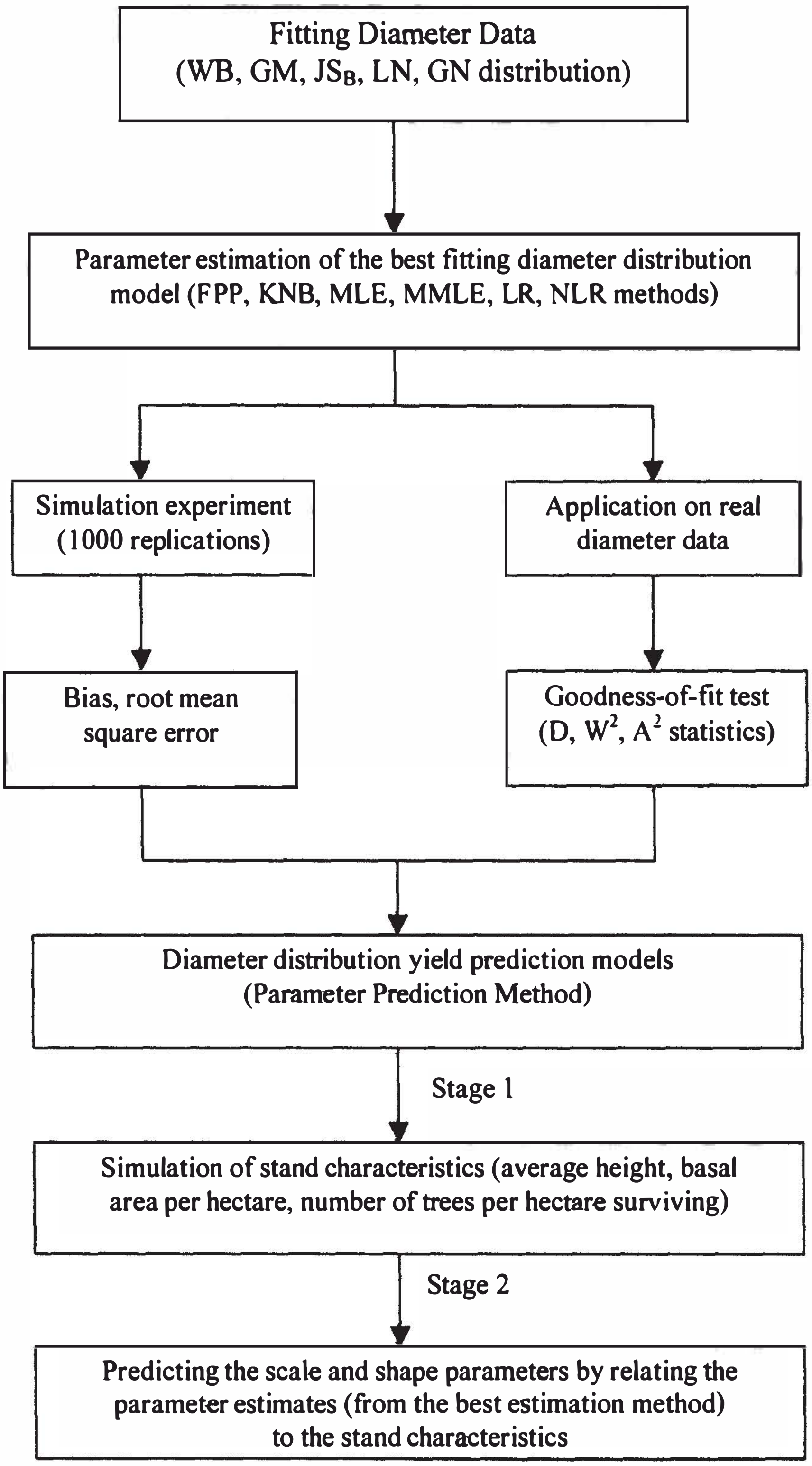


**Figure 1 : Basic concept of the development of diameter distribution yield prediction models**

# CHAPTER II

# STATISTICAL DISTRIBUTIONS FOR FITTING DIAMETER DATA

The distribution of diameter is the most potent simple factor to describe the properties of forest stands. Other variables such as volume, value, conversion cost, and product specifications are well correlated with diameter. Its relationship to site, stand composition, age, and density is often valuable for economic and biological purposes. These quantitative informations are helpful for managing forestlands on a sustainable manner.

There are many approaches to developing diameter distribution models. The most common approach is to represent the distribution of diameter in a stand based on the use of a probability distribution such as the Johnson $S_B$ or Weibull. Most of the probability distributions used in the model are actually a family of distributions, and they are indexed by several parameters. The model is constructed so that the parameters are related to aggregate stand variables such as age, density and site index. In applying the model on stand of interest, the stand variables are specified and the model determines a member of the family of the distribution that can be used as the model for the diameter distribution of the stand (Reynolds *et al.*, 1988).

For many years researchers have put considerable interest in describing the frequency distribution of diameter measurements in forest stands using probability density functions (pdf). Researchers have used various distributions for both even-aged and mixed-aged stands with varying degrees of success (Bailey and Dell, 1973 and references therein, Clutter and Allison, 1974, Zohrer, 1972).

According to Meyer and Stevenson (1943), in 1898 de Liocourt constructed a model based on the geometric progression for diameter distributions from uneven-aged forests and later on, they applied this general model, the exponential distribution, to forest of mixed species in Pennsylvania (Bailey and Dell, 1973). Bailey and Dell (1973) noted that other systems and distributions which broaden the consideration to include mound shape are Gram-Charlier series (Meyer, 1930), the Pearl-Reed growth curve (Nelson, 1964; Osborne and Schumacher, 1935), the gamma distribution (Nelson, 1964), and the three-parameter logarithmic-normal (Bliss and Reinker, 1964). The beta distribution, which is essentially a reparameterization of Pearson's more general Type 1, was applied to diameter distributions by Clutter and Bennet (1965) and later on McGee and Della-Bianca (1967) and Lenhart and Clutter (1971) subsequently developed the applications of models based on the beta distribution (Bailey and Dell, 1973).

The main problem in fitting distributions has been the choice of statistical distribution function for describing the probabilities of interest (Hafley and Schreuder, 1977). Hafley and Schreuder (1977) revealed that the criteria for choosing a distribution appear to be that the distribution be relatively simple to fit in terms of parameter estimation, sufficiently flexible to fit a relatively broad spectrum of shapes, lend itself easily to simple integration techniques for estimating proportions in various size classes, and fit any given set of observations well.

The pdf should cover shapes of either positive or negative skewness. Bailey and Dell (1973) emphasized that any constant in the model should be easily related to shape and locations features of the distribution and thus vary in a consistent manner with stand characteristics. The function should provide a promising base for advanced development and the function should also be easily fitted to observed data using parameter estimators that have desirable statistical properties.

## Statistical Distributions

Several studies have been conducted to describe the frequency distribution of diameter measurements in forest stands using pdf (Hafley and Schreuder, 1977). In this chapter, five distributions, Weibull, Gamma, Johnson $S_B$, Lognormal and Generalised Normal are compared in terms of their ability to model diameter data in uneven-aged and even-aged forest stands.

The cumulative distribution function (cdf), probability density function (pdf), inverse cdf or quantile function (invcdf), likelihood function and the loglikelihood function for the five distributions are given as follows.

### Weibull Distribution

Bailey and Dell (1973) showed that the Weibull function has the ability to assume a variety of curve shapes and easy to use by simplicity of algebraic manipulations. Other growth and yield systems based on the Weibull distributions were developed by Smalley and Bailey (1974b), Feduccia *et al.* (1979), Matney and Sullivan (1982), Baldwin and Feducia (1987) and Brooks *et al.* (1992).