DEVELOPMENT AND EXAMINATION OF SEQUENTIAL APPROACHES
FOR APPLICABILITY TESTING OF TREE VOLUME MODELS

By
Yue Wang
B.Sc. (Agri.), Agricultural University Of Guangxi, 1982
M.Sc. (For. Eng.), University Of New Brunswick, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
Doctor of Philosophy

in
THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF FOREST RESOURCES MANAGEMENT

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
December 1994
© Yue Wang, 1994
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of Forest Resources Management
The University of British Columbia
Vancouver, Canada

Date December 20, 1994
Abstract

In forest inventories, an existing standard volume equation may be considered for application to a local area (subpopulation), or to the same species in a different geographic region (new population). In order to use a volume model with confidence under these situations, an applicability test must be carried out to determine the actual accuracy of the model. Procedures based on a predetermined, sufficiently large sample of data (fixed sample size procedures) are available. However, since applicability testing is only used to classify a model as either acceptable or unacceptable, it is likely that a testing decision can be made with a smaller sample size, especially when the actual accuracy of the model is far below or above the user's requirement. This is particularly of concern when data collection for model checking is very expensive, time-consuming or destructive.

As an alternative, three sequential accuracy testing plans (SATP) were developed by extending Freese's (1960) accuracy tests using Wald's (1947) sequential probability ratio tests (SPRT) in this thesis. Observations are taking sequentially, and at each stage of sampling, a decision of whether to accept or reject the model, or to continue sampling, is made. The SATP procedures are potentially superior to a fixed sample size procedure in terms of lower sample sizes. Approximate Operating Characteristic (OC) and Average Sample Number (ASN) equations were also suggested to assist the potential users of the SATP procedures in choosing appropriate tested parameters for a given problem.

The simulation results using normal distribution generators showed that the SATP procedures are reliable for classifying a volume model as either acceptable or unacceptable based on two pre-set limits of the accuracy requirement. Also, on average, the use of the SATP procedures will result in a 40 to 60% of sampling cost-saving compared to an
equally reliable conventional fixed sample size procedure. A detailed example is given to illustrate the application of the SATP procedures.
Table of Contents

Abstract .......................................................... ii

List of Tables ...................................................... viii

List of Figures ...................................................... x

Acknowledgement ................................................... xi

1 Introduction ....................................................... 1

2 Literature Review ................................................ 8

2.1 Tree Volume Estimation ........................................ 8

2.2 Applicability Testing of Tree Volume Models ................. 13

2.3 Fixed Sample Size Procedures for Validating Forestry Models 15

2.3.1 Freese’s procedure of accuracy testing and its modifications 17

2.3.1.1 Freese’s procedure ..................................... 17

2.3.1.2 Underlying assumptions of Freese’s procedure .......... 20

2.3.1.3 Modifications of Freese’s procedure ..................... 24

2.3.1.4 Previous applications of Freese’s procedure ............. 26

2.3.2 Procedures for validating forestry simulation models .... 28

2.3.3 Reynolds’ estimation procedures .......................... 31

2.3.4 Other fixed sample size procedures ........................ 32

2.4 Variable Sample Size Procedures for Testing Forestry Populations 34

2.4.1 Sequential probability ratio tests .......................... 37
2.4.1.1 Wald's general decision rule for SPRT plans ........... 37
2.4.1.2 Wald's general OC function for SPRT plans .......... 40
2.4.1.3 Wald's general ASN function for SPRT plans .......... 42
2.4.1.4 Modifications of Wald's decision boundaries .......... 43
2.4.1.5 Previous forestry applications and studies of Wald's SPRT plans ................................................. 45

2.4.2 Other sequential sampling plans used in forestry .......... 49
2.4.3 Development of sequential analysis theory in statistics .... 51
2.5 Summary ................................................................. 52

3 Extension of Freese's Procedure to Sequential Accuracy Testing Plans 55

3.1 Sequential Accuracy Testing Plan I ............................. 56
  3.1.1 Modification for Freese's original formulation ............ 56
  3.1.2 Sequential decision boundaries ............................. 58

3.2 Sequential Accuracy Testing Plan II ............................ 61
  3.2.1 Modifications of Freese's original formulation .......... 61
  3.2.2 Sequential decision boundaries ............................. 62

3.3 Sequential Accuracy Testing Plan III ............................ 64

3.4 Approximate OC and ASN Equations for the SATP Procedure .... 66

4 Monte Carlo Simulations for Examining Sequential Accuracy Testing Plans 71

4.1 Estimates of Error Parameters of Tree Volume Estimations .... 73

4.2 Reliability of the SATP When Errors Are iid Normal Variables With a Zero Mean and a Constant Variance ................. 76
  4.2.1 Simulation methods ............................................ 79
4.2.2 Reliability of the SATP procedures for classifying the error variance of volume models ........................................ 81
4.2.3 Accuracy of the nominal error probabilities of the SATP .......... 87
4.2.4 Expected sample size of the SATP procedures .......................... 89

4.3 Reliability of the SATP When Errors Are iid Normal Variables With a Non-Zero Mean and a Constant Variance ................. 95

4.4 Reliability of the SATP When Errors Are Normally Distributed With a Zero Mean and Heterogenous Variances .......................... 98
   4.4.1 Simulation Methods .............................................. 98
   4.4.2 Simulation Results and Discussions .............................. 103

4.5 Accuracy of the Approximate OC and ASN Equations of the SATP 104
   4.5.1 Simulation methods ............................................... 107
   4.5.2 Comparisons between the approximate and Monte Carlo OC and ASN equations ........................................ 109

4.6 Effects of Truncation and Group Sampling on the Performance of the SATP110

4.7 Conclusion ............................................................... 118

5 Field Application Procedures And Example 122

5.1 Field Application Procedures ............................................ 122
   5.1.1 Preparation ......................................................... 122
   5.1.2 Sampling and testing process ..................................... 125
   5.1.3 Post-tests for the underlying assumptions ....................... 128
   5.1.4 Making the final testing decision .................................. 129

5.2 Application Example ..................................................... 131
   5.2.1 Problem and data ................................................... 131
   5.2.2 Specification of test parameters ................................... 131
5.2.3 Testing results ................................................. 133
5.2.4 Final testing decision ........................................ 135

6 Conclusions and Recommendations ........................ 138

References Cited .................................................... 144
List of Tables

4.1 Descriptive statistics of B. C. tree sectional data used for determining the error parameters of volume estimations. ............................................. 75
4.2 Calculated statistics of the errors for three volume models and B.C. tree sectional data. ................................................................. 77
4.3 Monte Carlo OC estimates of SATP I for testing $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with $\alpha = \beta = 0.05$ when five normal generators with zero means and different variances were used to simulate the errors of volume models. 85
4.4 Monte Carlo OC estimates of SATP I for testing $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with $\alpha = \beta = 0.10$ when five normal generators with zero means and constant variances were used to simulate the errors of volume models. 86
4.5 Monte Carlo estimates of the actual error probabilities when the SATP I is applied to test $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with various $\sigma_0^2$, $\sigma_1^2$ and $\alpha = \beta = 0.05$. ................................................................. 90
4.6 Monte Carlo estimates of the actual error probabilities when the SATP I is applied to test $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with various $\sigma_0^2$, $\sigma_1^2$ and $\alpha = \beta = 0.10$. ................................................................. 91
4.7 Monte Carlo ASN estimates and the associated standard error when $\sigma^2 = \sigma_0^2$ for applying SATP I to test $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with various combinations of $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$. ................................................................. 93
4.8 Comparison of Monte Carlo OC values when $\sigma^2 = \sigma_0^2$ between SATP I and III for testing $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with $\alpha = \beta = 0.05$ and various $\sigma_0^2$ and $\sigma_1^2$ using normal generators with a non-zero mean.

4.9 Monte Carlo estimated probabilities and associated average sample size for accepting $H_0$ when the SATP I is applied to test $H_0 : \sigma^2 = e_0^2/z_{1-\gamma/2}^2$ vs. $H_1 : \sigma^2 = e_1^2/z_{1-\gamma/2}^2$ with $e_0$ and $\gamma$ determined from the mixed normal generated population, $N(0, 0.05x_i)$.

4.10 Monte Carlo estimated probabilities and associated average sample size for accepting $H_0$ when the SATP I is applied to test $H_0 : \sigma^2 = e_0^2/z_{1-\gamma/2}^2$ vs. $H_1 : \sigma^2 = e_1^2/z_{1-\gamma/2}^2$ with $e_0$ and $\gamma$ determined from the mixed normal generated population, $N(0, 0.5x_i^2)$.

4.11 Comparisons of the Monte Carlo OC and ASN curves of the SATP I and III and the approximate equations with $\sigma_0^2 = 0.3, \sigma_1^2 = 0.8, \alpha = \beta = 0.05$.

4.12 Comparisons of the Monte Carlo OC and ASN curves of the SATP I and III and the approximate equations with $\sigma_0^2 = 1.50, \sigma_1^2 = 2.25, \gamma = \alpha = 0.05$ and $\beta = 0.10$.

4.13 Comparison of Monte Carlo estimates of $\alpha, \beta$ and associated ASN values of SATP I for various group sizes with $\sigma_0^2 = 1.221, \sigma_1^2 = 2.747$ and $\alpha = \beta = 0.05$.

4.14 Comparison of Monte Carlo estimates of $\alpha, \beta$ and ASN values of SATP I for various truncation points and different combinations of $\alpha$ and $\beta$ with $\sigma_0^2 = 1.221, \sigma_1^2 = 2.747$.

5.15 Field decision table for applying the SATP procedures.

5.16 Calculated statistics of the lodgepole pine trees collected from FIZ E in B.C.
List of Figures

3.1 Example of an approximate Operating Characteristic (OC) curve of SATP I for testing \( H_0 : \sigma^2 = 1.50 \) vs. \( H_1 : \sigma^2 = 2.25 \) with \( \alpha = 0.05 \) and \( \beta = 0.10 \). 69

3.2 Example of an approximate Average Sample Number (ASN) curve of SATP I for testing \( H_0 : \sigma^2 = 1.50 \) vs. \( H_1 : \sigma^2 = 2.25 \) with \( \alpha = 0.05 \) and \( \beta = 0.10 \). 70

4.3 Comparison of the Monte Carlo OC and ASN curves of the SATP I and III and the Wald’s approximated equations with \( \sigma_0^2 = 0.3, \sigma_1^2 = 0.8, \alpha = \beta = 0.05 \). (a) Comparison of OC curves; (b) Comparison of ASN curves. 113

4.4 Comparison of the Monte Carlo OC and ASN curves of the SATP I and III and Wald’s approximated equations with \( \sigma_0^2 = 1.50, \sigma_1^2 = 2.25, \alpha = 0.05 \) and \( \beta = 0.10 \). (a) Comparison of OC curves; (b) Comparison of ASN curves. 114

5.5 Field decision graph for applying the SATP procedures. 125

5.6 Field decision graph for SATP II when applied to test the applicability of B.C. standard volume model of lodgepole pine for the application to the subregion, FIZ E with specified parameters: \( p_0 = 10 \) and \( p_1 = 15 \), \( \gamma = 0.05, \alpha = \beta = 0.05 \) and \( n_0 = 50 \). 135

5.7 Field decision graph for SATP III when applied to test the applicability of B.C. standard volume model of lodgepole pine for the application to the subregion, FIZ E with specified parameters: \( e_0 = 0.08 \) and \( e_1 = 0.10 \) in logarithmic units, \( \gamma = 0.05, \alpha = \beta = 0.05 \) and \( n_0 = 50 \). 136
Acknowledgement

I am grateful to my major advisor, Dr. Valerie LeMay, for her guidance and encouragement. I would like to thank my committee, Drs. A. Kozak, P. Marshall, D. Tait from Dept. of Forest Resources Management, and Dr. N. Heckman from Dept. of Statistics. Their contributions to my thesis and forest biometrics education were enormous, as was their patience and willingness to help at all times.

I thank Dr. A. Kozak, my former major advisor, for accepting me as a Ph.D. candidate and assisting me in obtaining the data for this research from the Inventory Branch, B. C. Ministry of Forests. Thanks also go to Drs. P. Marshall and V. LeMay for their initial works, which inspired me to carry out this research. I have also profited from working with the other graduate students in the Biometrics Hut.

Funding for this research was supplied by the University of British Columbia through various fellowships.

Finally, I would like to thank my wife, Wan Ju, for her patience and encouragement.
Chapter 1

Introduction

Although the growth of trees involves a complex biological process, and the forest is a dynamic ecosystem, timber management shares the general character of other production processes. That is, given levels of inputs, the process will result in certain outputs, which mean a profit or loss for the owners. Intervening in, or modifying the production process may help the owners to maximize profit. The goal of timber management is to maintain or increase the yield of the timber resources. To achieve this goal, forest managers must make many decisions. Examples include selecting the optimum age to harvest (rotation age), the levels of planting density, and the timing of thinnings. Whether these decisions can be made intelligently depends on the availability of accurate and up-to-date information about the managed forest, especially quantitative information on the measurable characteristics of forest. Since the basic management unit of the forest is a stand, and any stand is an aggregation of trees, the basic information for timber management decision-making is measurements of tree or stand characteristics (e.g., diameter, height, volume and density).

To help forest managers obtain quantitative information, forest mensurationists and biometricians have developed many mathematical functions for various purposes. For example, to sample a forest population, the relationship between a main variable and an auxilliary variable may be represented as a regression model. Based on this relationship, the parameter of the main variable can be efficiently estimated using ratio/regression estimators, since the measurements of the auxilliary variable are usually easier to obtain.
or previously available.

To estimate the total or merchantable volume of a stand or forest, a volume equation may be developed that relates tree volume to a few easily measured variables (e.g., tree diameter and height). The difficulty of estimating tree volume is reduced. To obtain information on future forest volumes, a forest growth or yield model may be developed for making predictions. Based on this model, a yield table can be established to provide the information required for forest management decision-making. As the intensity of forest management increases, the use of forest growth and yield simulation models, which include a system of forestry prediction models, becomes more and more common in forest management activities.

In the history of applying mathematical functions, much attention in forestry has been focused on the following: (1) suggesting appropriate sampling methods for model development (optimal sampling design for model estimation) (e.g., Penner 1989, Marshall and Wang 1993); (2) proposing improved approaches for estimating the model parameters (model estimation) (e.g., Burkhart 1986, Gregoire 1987, Kozak 1988, LeMay 1990); and (3) validating the ability of the developed models to represent the real systems (model validation), or to assess the ability of the developed models to predict the dependent variable (accuracy, or prediction error testing) (e.g., Freese 1960, Reynolds 1984, Reynolds et al. 1988, Gregoire and Reynolds 1988). The research on suggesting new sampling and estimation methods provides forest researchers with better models, which give more accurate predictions, a better representation of real systems, or models with superior properties required for making statistical inferences about model parameters (i.e., constructing confidence intervals or conducting hypothesis tests). Model validation builds confidence for using a developed model.

The specific problem addressed in this dissertation is the assessment (testing) of the applicability of forestry models, particularly tree volume estimation models. Applicability
testing of forestry models is an extension of model validation; it focuses on assessing the predictive ability of a forest model when applied to a population other than that for which it was developed. The predictive ability (accuracy) of a model is defined as the closeness of the predicted or estimated values to their target values.

Applicability testing of forestry models is a common practice in forest management. Examples can be found from the early stages of checking a tree volume table established by hand-drawn curves (e.g., Bruce 1920) to the validation of the complicated computer simulation models of today (e.g., Goulding 1979; Reynolds et al. 1981). The importance of this problem is based on the fact that the procedures of model specification, parameter estimation, and model validation are usually based on data collected from a specific forest population (e.g., a particular tree species) over a large geographic region of forest lands. Developed in this way, a forestry model is valid only for making estimates or prediction for the same population for which it was developed. However, in many situations, an existing forestry model may be considered for application to a subpopulation (e.g., subregion), a new population (e.g., the same species in a different geographic region or a similar species), or the same population following some change (e.g., following a forest fire or silvicultural treatment). Under these situations, the above assumption for applying regression models may not hold, and the accuracy of the developed models becomes uncertain. Using an existing model in these situations may result in inaccurate estimations; the user may then make incorrect management decisions (e.g., annual allowable cut levels too low or too high).

To test the applicability of a forestry model, a new sample set is required from the population to which the model will be applied (application population). The accuracy of the model is then determined by comparing the predicted or estimated values with the actual observed values. This research is concerned with the appropriate statistical procedures that can be used to draw conclusions about the accuracy of models based
on comparing the predicted and actual values. In forestry, a number of papers have been published to suggest such statistical procedures (e.g., Freese 1960, Reynolds et al. 1981, Reynolds 1984, Reynolds et al. 1988, Gregoire and Reynolds 1988). These procedures treat the differences between the predicted or estimated values and the actual values (model errors) as random variables. Statistical hypothesis testing or estimation procedures are conducted to infer the parameters (mean or variance) of these random variables.

Although these statistical procedures can be used for detailed evaluations of the predictive ability (accuracy) of forestry models, they must be based on a sample set with a predetermined sufficiently large size from the application populations (i.e., they are fixed sample size procedures). In many situations, a decision to accept or reject an existing model may be made based on data from a smaller sample set, especially when the actual accuracy of the model is far below or above the accuracy requirement of the user. A fixed sample size procedure is then considered to be inefficient for applicability testing. This is particularly true when data collection is highly expensive, time-consuming or destructive, as with data for volume model development. Therefore, alternatives to fixed sample size procedures should be suggested for applicability testing of forestry models. The desired criteria for the alternative procedures were:

1. The accuracy determination of a model should be made with a minimum sample size, or at least a smaller sample size than required by an equally reliable fixed sample size procedure;

2. The accuracy of a model should be explicitly tested;

3. The use of the alternative procedure should not preclude the use of the available fixed sample size procedures; and
4. The testing procedures should be easy to apply under field conditions.

The principal objectives of this research were: (1) to review forestry and other literature, and from this literature to propose alternative procedures which meet the desired criteria for applicability testing of forestry models, particularly tree volume estimation models; (2) to test the reliability of the alternative procedures for various common errors of tree volume models; and (3) to suggest and illustrate appropriate procedures for applying the alternative procedures.

To meet the first objective, a literature search was conducted, first by examining forestry literature, and then extending this to biometrics, technometrics, industrial quality control and statistics literature. In order to limit the scope of this research, only model validation procedures based on comparing the estimated and observed values (prediction error testing or estimation procedures) were examined. No procedures were found in the literature which satisfied all of the desired criteria of applicability testing of forestry models. However, the information from the literature was used to develop three new testing procedures, labelled sequential accuracy testing plans (SATP) I, II and III. The developed SATP procedures are extensions of Freese's (1960) procedure of accuracy testing using Wald's (1947) sequential probability ratio test (SPRT) plan. Since Freese's (1960) idea of accuracy testing was adopted, the SATP procedures can be used for explicitly testing the accuracy of a model. As sequential sampling plans, the characteristic feature of the SATP procedures is that the number of observations required is not preset in advance, but is a function of the observations themselves. This feature of the SATP procedures allows applicability testing of forestry models to be realized with a minimum sample size, especially when the actual accuracy of the tested model is very high or very low. In order to assist potential users of the SATP procedures, an approximate Operating Characteristic (OC) equation for determining the probability of making a correct
Chapter 1. Introduction

decision and an approximate Average Sample Number (ASN) equation to indicate the expected average sample size required to reach a terminal decision were also derived.

To meet the second objective, Monte Carlo simulations were conducted to examine the reliability of the SATP procedures using normal, or mixed normal distribution generators to simulate different errors found in tree volume models. These simulated error populations included normally distributed errors with (1) a zero mean and a constant variance; (2) a non-zero mean and a constant variance; and (3) a zero mean and non-constant variances. Two modifications of Wald’s (1947) SPRT, taking more than one observation at each stage (group sequential sampling) and setting an upper limit to avoid the situation of a very large sample size (truncation), were also examined.

Finally, to meet the third objective, field procedures for applying SATP procedures were suggested, and an example was constructed to illustrate the suggested application procedures.

The contributions of this research are:

1. This research lays the foundation for future research in application of sequential analysis to the applicability testing of forestry models and thereby brings to forest researchers’ attention the concept of sequential analysis for forestry model evaluation;

2. The results of this research provide an alternative to fixed sample size procedures for applicability testing of forestry models, especially for circumstances where data collection is highly expensive, time-consuming, or destructive; and

3. Although tree volume models were addressed, the methods and results of this research are applicable for determining the applicability of other forestry models with the same distributional error assumptions. The results of this research will also benefit other scientific fields where mathematical models are developed for estimation
or prediction purpose.

The dissertation has been organized according to the following format. The background information for this research is presented in Chapter 2. The extension of Freese's (1960) procedure into a sequential testing procedure is presented in Chapter 3. The methods and results of Monte Carlo simulations for testing the reliability of SATP procedures are presented in Chapter 4. The proposed field application procedures and application example are in Chapter 5. The conclusions and recommendations for further research can be found in the last chapter.
Chapter 2

Literature Review

To achieve the first objective of this research, to propose alternatives to fixed sample size procedures for applicability testing of tree volume models, a literature search was carried out, first by examining forestry literature, and then by extending it to the biometrics, technometrics, industrial quality control, and statistics literature. A summary of this literature search is presented in this chapter. The first section is a presentation of the methods for tree volume estimation. A brief review of the problem and history of applicability testing of tree volume models is presented in the second section. In section three, some widely used statistical procedures for validating forestry models using a fixed sample size are reviewed and discussed in detail. Variable sample size (sequential sampling) procedures for sampling forestry populations are presented in section four. Finally, an overview of the results of this literature search can be found in the last section.

2.1 Tree Volume Estimation

Tree volume information is essential information required for timber management decision-making. Any tree is composed of a bole or stem, a root system, and branches and leaves that collectively make up the crown. However, a forester speaking of individual tree volumes is commonly referring to either the total volume, or the commercially marketable volume of the tree bole. The total tree volume includes the volume of the main bole of a tree from ground to tree tip, whereas the merchantable tree volume includes the volume of the merchantable part of the main bole only. Methods for estimating tree volumes
in forestry have been well researched, and many mathematical functions (tree volume models) have been proposed for this purpose (see, Clutter et al. 1983, p.8; and Loetsch and Haller 1973, p.154).

Historically, the relationship between tree volume and other tree measurements was represented as a volume table. A volume table provides the average volume of standing trees of various sizes and species (Avery 1975, p.92). The basis for the use and construction of tree volume tables was given by Cotta in the 19th century (see Husch et al. 1982 p.157). Such tables are based on mathematical relationships between tree volume and a few easily measured tree attributes (e.g., diameter, height) for a given species and region. These relationships can be expressed as

\[ V = f(\text{dbh}, H, F) \]  

(2.1)

where \( V \) is either total tree volume or merchantable volume; \( \text{dbh} \) is tree diameter outside bark at breast height (1.3 metres above ground); \( H \) is the total tree height or the height to some specified upper-stem merchantability limit of tree diameter inside bark (e.g., 6 cm, 8 cm etc.); and \( F \) is tree form. The commonly used measures of tree form are ratios of diameters at specified heights to \( \text{dbh} \), called form quotients (Spurr 1952). Also, since other variables, besides \( V \), \( H \) and \( F \), may affect the volume of a tree, an additive or multiplicative error, \( \epsilon \), will be associated with the function above in representing tree volume (i.e., a random error term). A tree volume table derived from this function is known as a form class volume table. Because of the difficulty of measuring the upper diameter of a tree, the construction of this type of table is mainly for research purposes. Functional forms including only \( \text{dbh} \), or \( \text{dbh} \) and height are more common. Tree volume tables derived from these functional forms are called local tree volume tables and standard tree volume tables, respectively. Today, volume tables are rarely used in forestry practice; instead, calibrated volume functions (tree volume models) are directly used for estimating
Chapter 2. Literature Review

Some commonly used tree volume models are:

1. Schumacher and Hall (1933)

\[ V = \beta_1 (dbh)^{\beta_2} H^{\beta_3} \epsilon \]  

(2.2)

2. Logarithmic Schumacher and Hall

\[ \log(V) = \beta_0 + \beta_1 \log(dbh) + \beta_2 \log(H) + \epsilon \]  

(2.3)

3. Combined variable (Spurr 1952)

\[ V = \beta_0 + \beta_1 (dbh)^2 H + \epsilon \]  

(2.4)

where \( \log \) is logarithm (base 10); and \( \beta_0, \beta_1, \beta_2, \beta_3 \) are unknown parameters that need to be estimated.

To develop a tree volume model, a large sample of trees is selected from a given species and forest region where the volume model will be applied. Besides dbh and the total height, the diameters at intervals along the stem of each sampled tree are accurately measured using a precise optical instrument, or sample trees are felled and measured. Using the measurements of sectional diameters, standard formulae (e.g., Huber's or Smalian's formulae, see Husch et al. 1982) are then applied to determine the volume of each section, and then summed to obtain the stem volume. Using this large data set, the parameters of a tree volume model can be estimated. In the early years of volume table construction, graphical methods, such as the harmonized-curve method (e.g., Baker 1925) and alignment-chart methods (e.g., Reineke and Bruce, 1932), were used to relate the measured volume to dbh and height of the sampled trees. However, all tree volume models used today are fitted using least-squares regression.

In developing and using tree volume models, some common problems are encountered. First, since the variation of tree volume usually increases as the tree size (dbh and height)
increases, the assumption of regression that the variances of the dependent variable will be equal across all levels of the independent variables (homogeneous variances) is violated. In this situation, the least-squares estimated coefficients are still unbiased. Since tree volume models are developed mainly for the purpose of estimation, many tree volume models with heterogeneous variances of the estimated volumes are commonly used in forestry. However, statistical inferences based on the standard error of estimate \((SE_E)\) and the standard errors of estimated coefficients of these models will be invalid. This can be improved through using weighted least-squares regression to estimate the coefficients of the model, if the appropriate weight is known or it can be estimated.

Cunia (1964) studied the variance patterns of tree volumes for some common tree species of North America (black spruce \((Picea mariana\) (Mill.) B.S.P.), balsam fir \((Abies balsamea\) (L.) Mill.), jack pine \((Pinus banksiana\) Lamb.) and yellow birch \((Betula alleghaniensis\) Britton \((Betula lutea\) Michx.f.)). He found that the variance of tree volumes could usually be estimated using one of following models:

\[
\sigma^2_V = k_1(ddbh)^{k_2} \quad (2.5)
\]
\[
\sigma^2_V = k_1(ddbh)^2 \times \bar{H} \quad (2.6)
\]
\[
\sigma^2_V = k_1 + k_2(ddbh) + k_3(ddbh)^2 \quad (2.7)
\]

where \(\sigma^2_V\) is the estimated variance of total tree volume; \(\bar{H}\) is the average tree height for given dbh class; and \(k_1, k_2\) and \(k_3\) are unknown constants. Clutter et al. (1983) tabulated the assumed variance patterns for some common volume equations.

Logarithmic tree volume models (e.g, logarithmic Schumacher and Hall's (1933) model) were found to meet the assumptions of the constant variance and normality for estimated logarithmic volumes better. This model form is widely used for tree volume estimation in North America. For example, the standard tree volume equations suggested for the main species in the province of British Columbia (B.C.) are all logarithmic
forms (Watts 1983, p.431-432). Two problems exist in using the logarithmic models: (1) the $SE_E$ associated with a logarithmic model is not comparable to the $SE_E$ of other models with the dependent variables in tree volume units; and (2) by taking logarithms the estimated arithmetic means are automatically replaced by the estimated geometric means, and because the first is always larger than the second, the volumes are consistently underestimated. To solve the first problem, Meyer (1938) applied a theorem from the calculus of errors. If the standard error ($\sigma$) of a statistic $x$ is known, the standard error, $\sigma_f$, of any function of $x$ ($f(x)$) can be approximated as:

$$\sigma_f \cong \left| \frac{df(x)}{dx} \right| \times \sigma$$ (2.8)

where $\frac{df(x)}{dx}$ is the first derivative of the function of $x$ with respect to $x$. When this method is applied, the standard error of estimate in volume units ($S_v$) for logarithmic Schumacher and Hall’s (1933) model can be approximated as follows. For $\log V = x$, and $V = 10^x$, this becomes:

$$f(x) = 10^x$$

$$\frac{df(x)}{dx} = 10^x \ln 10.$$  

Therefore, the approximate standard error of estimate in volume unit is:

$$S_v \cong 10^x \times \ln 10 \times SE_E$$

$$S_v \cong 2.3026 \times 10^x \times SE_E$$ (2.9)

where ln is natural logarithm and $SE_E$ is the standard error of estimate associated with the logarithmic model. The percentage standard error of estimate in volume units ($S_v(\%)$) is then:

$$S_v(\%) \cong 100 \left( \frac{2.3026 \times 10^x \times SE_E}{10^x} \right) = 230.26 \times SE_E$$ (2.10)
Alternatively, Meyer (1938) showed that $S_v(\%)$ could also be approximated as:

$$S_v(\%) \cong 100 \left(10^{SE_E} - 1\right) \tag{2.11}$$

Replacing $10^{SE_E}$ in Equation 2.11 by the first two terms of the expansion of Taylor’s series, Equation 2.11 will also be equal to 230.26 $SE_E$. Meyer indicated that Equation 2.10 was appropriate when $SE_E$ was less than 0.10. However, in case $SE_E$ was greater than 0.10, Equation 2.11 should be used.

For solving the problem of underestimated volumes using logarithmic models, Meyer (1938) indicated that the underestimate depends upon $SE_E$. He suggested correction factors for different values of $SE_E$ (see Spurr 1952, p.74 for a detailed discussion). Other methods based on the properties of the lognormal distribution were discussed by Flewelling and Pienaar (1981).

### 2.2 Applicability Testing of Tree Volume Models

Applicability testing of tree volume models (tables) appears to have been first mentioned by Bruce (1920). He stated that the accuracy checking of a volume table was applicability testing when applied to a population other than that for which it was developed. Spurr (1952, p.122) indicated that applicability testing of tree volume tables was used to answer the question of whether a volume table should be constructed locally for each species, or whether a single volume table could be applied to a given species wherever it occurs. Using data from 1021 red spruce, black spruce and balsam fir trees, he investigated the effect of species, type of growth, site, and locality on estimating total cubic-foot volume using standard tree volume tables. He concluded that “the locality, type of growth, and site where a tree grows apparently do not affect the total cubic-foot volume sufficiently to justify the development of more than one volume table for a given species. Different species, however, do have different volumes for trees of the same dimensions,
but this difference may usually be expressed as a constant percentage correction which is independent of tree size. One volume table, therefore, may be used for a number of species, with the final volume estimate being corrected by a single percentage to adjust for the species being estimated." Spurr's conclusion implies that a well established standard tree volume model may be applied to a larger geographic region, and for different tree species, if the source of the model errors can be identified and a correction factor can then be used to modify the model estimates.

In applicability testing of tree volume models, the volume of sample trees should be compared with the estimated volume from the volume table to be tested (Husch et al. 1982). In order to do this, the sample trees first must be selected appropriately from the population where the tree volume model will be applied and the volume of each sample tree must be accurately determined. Second, appropriate criteria or statistical procedures should be used to determine the accuracy of the estimation of the model (i.e., how close the estimated volume is to the actual volume). Husch et al. (1982, p.156) suggested the following conditions for selecting sample trees for testing the applicability of a tree volume table:

1. Sample trees for a given species, or species group, should be well distributed through the population to which the volume table will be applied;

2. No size, type, or growing conditions should be unduly represented in the sample; and

3. If a sample of cut-trees is used, this sample, if not representative of the timber, should be supplemented by a sample of standing trees.

Bruce (1920) suggested that the aggregate deviation and the average deviation between the estimated and actual volumes should be used to evaluate the accuracy of tree
volume tables. He defined the aggregate deviation as the difference between the sum of the actual volumes and the sum of the estimated volumes, expressed as a percentage of the latter, and the average deviation as the arithmetic sum of the absolute values of the differences between actual and estimated volumes expressed as a percentage of the sum of the estimated volumes. Spurr (1952, p. 75) indicated that aggregate deviation is an indication of the freedom of the volume table from bias and should ordinarily not exceed one percent, and average deviation indicates primarily the variability inherent in the data used, and may often be as high as 10 percent. He also indicated that these two measures are valid tests of precision, especially in the graphic and alignment-chart methods where the curves are fitted on the basis of the aggregate and average deviations. However, if the tree volume tables are constructed using the least-squares method, then it is advisable to assess precision in terms of the squares of the deviations rather than in terms of the deviations themselves. Husch et al. (1982, p. 156) stated that the aggregate deviation between the measured and estimated volumes of sample trees in an applicability check of tree volume tables, for practical purposes, should not exceed $2CV/\sqrt{n}$, where $CV$ is the coefficient of variation of the volume table being tested, and $n$ is the number of trees used in the test. If this criterion is met, the table is applicable without correction.

2.3 Fixed Sample Size Procedures for Validating Forestry Models

The applicability testing of forestry models is an extension of forestry model validation. Van Horn (1971) defined model validation as “the process of building an acceptable level of confidence that an inference about a simulated process is a correct or valid inference for the actual process”. In other words, a model validation process involves testing the “usefulness” or “validity” of a model. Testing the usefulness or validity of a model implies that (1) we have established a set of criteria for differentiating between those
models which are "useful" or "valid" and those which are "not useful" or "not valid", and (2) we are able to apply these criteria to any given model. These two aspects of model validation require an understanding of the nature of the problem including how the model estimation or prediction will be used and the impact of errors on these uses, as well as the availability of statistical procedures that are designed to fit the conditions of the problem (Reynolds et al. 1981, p.350).

Model validation may be considered to be an essential component of the model building process, which provides users with some indication about how well the model will perform in the real population. In this situation, model validation can be carried out by splitting the available data into an estimation and a validation set, and no new data set is needed (i.e., cross-validation approach, see Snee 1977 for detailed discussions). However, since applicability testing of forestry models is used when a model is considered for application to a population other than that for which it was developed, new data must be collected from the application population.

In forestry, a number of papers have been published on the philosophy and methodology of model validation (e.g., Freese 1960; Reynolds et al. 1981; Reynolds 1984; Gregoire and Reynolds 1988; and Marshall and LeMay 1990). In general, the method for validating an estimation or prediction model is to compare the data \( y \) obtained from the actual population of interest with corresponding data \( \hat{y} \) generated from the model. The essence of this comparison is to examine the model errors \( d \ (d = y - \hat{y}) \), or relative errors, \( d/y \). Widely used statistical procedures for this purpose include:

1. Freese's (1960) procedure of accuracy testing and its modifications (i.e., Rennie and Wiant 1978; Reynolds 1984; and Gregoire and Reynolds 1988);

2. Reynolds et al.'s (1981) procedures for validating stochastic simulation models;

3. Reynolds' (1984) estimation procedures; and
4. Other procedures (i.e., paired t-test, analysis of variance and regression procedure etc.).

The use of these procedures requires a data set from the application population with a predetermined, sufficiently large sample size; these procedures are then called fixed sample size procedures.

2.3.1 Freese's procedure of accuracy testing and its modifications

2.3.1.1 Freese’s procedure

Freese (1960) developed a statistical procedure for determining the accuracy of a new estimating or measuring technique against an accepted standard. His procedure used a standard chi-square test for a hypothesized variance. Freese's idea of accuracy testing was based on comparing estimated and observed values against an established standard. This idea was realized by first stating the accuracy required (or the inaccuracy that would be tolerated) by the user of the technique, then estimating the accuracy obtained by the technique, and finally determining whether the accuracy of the technique met the accuracy required through statistical testing of a hypothesis.

Freese's (1960) procedure of accuracy testing included three chi-square test statistics. The first test statistic was designed to determine the accuracy of the technique in the measurement units of the variable of interest. This test statistic is:

\[ \chi^2_{n} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sigma_h^2} = \sum_{i=1}^{n} d_i^2 / \sigma_h^2 \]  

(2.12)

where \( \chi^2_{n} \) represents a chi-square distributed variable with \( n \) degrees of freedom under the assumptions explained in section 2.3.1.2; \( y_i \) and \( \hat{y}_i \) are the \( i^{th} \) pair of observed and estimated values, respectively; \( d_i \) are the differences between observed and estimated values (also called the errors of a model in this dissertation); \( n \) is the sample size; and \( \sigma^2_h \) is the hypothesized variance representing the precision required by the user.
In applying this test statistic, the potential user of the technique is first required to specify an allowable error limit \( e \) with a desired probability level, \( 1 - \gamma \). These values specify that for acceptance of the model, \( 100(1 - \gamma) \) percent of the estimated values \( \hat{y}_i \) must be within \( e \) units of the actual values \( y_i \). The specified values are then used to calculate a hypothesized variance as:

\[
\sigma^2_h = \frac{e^2}{z^2_{1-\gamma/2}}
\]

where \( z_{1-\gamma/2} \) is the \( 100(1 - \gamma/2) \) percentile of the standard normal distribution. Using a specified probability of a type I error, \( \alpha \), the model will be considered to be acceptable if the computed \( \chi^2_{(n)df} \) is below \( \chi^2_{1-\alpha}(n) \), the \( 100(1 - \alpha) \) percentile of the chi-square distribution with \( n \) degrees of freedom. If the model is accepted, it means that the user's accuracy requirement specified by \( e \) and \( \gamma \) is met. Since the sum of squared errors \( \sum_{i=1}^{n} d_i^2 \) is a measure of the total error, Freese (1960) indicated this statistic would reject inaccurate techniques regardless of the sources of inaccuracy (large bias, lack of precision, or both).

The second test statistic was designed to test the accuracy of a technique in relative units. The explanation for using this test is that specifying a percent error limit \( p \) is more convenient than specifying the error limit in absolute units \( e \) in many practical applications.

\[
\chi^2_{(n)df} = \sum_{i=1}^{n} \left[ \frac{(y_i - \hat{y}_i)^2}{\sigma^2_h y_i^2} \right]
= \sum_{i=1}^{n} \left[ \frac{100^2 z_{1-\gamma/2}^2 (y_i - \hat{y}_i)^2}{p^2 y_i^2} \right]
= \left[ \frac{100^2 z_{1-\gamma/2}}{p} \right] \sum_{i=1}^{n} \left[ \frac{d_i}{y_i} \right]^2
\]

where the hypothesized variance, \( \sigma^2_h = p^2/(100z_{1-\gamma/2})^2 \); and \( \gamma \) is specified to define an acceptable model that at least \( 100(1 - \gamma) \) of the relative errors, \( d/y \), must be within
p/100. With a significance level of $\alpha$, if the computed $\chi^2_{(n)df}$ is below $\chi^2_{1-\alpha}(n)$, the accuracy requirement of the user specified as $p$ and $\gamma$ is met, and the technique will be acceptable. This test statistic also rejects an inaccurate technique regardless of the sources of inaccuracy.

The last test of Freese’s (1960) procedure is a “bias-free” test. This test is considered to be useful since some techniques may be capable of providing excellent accuracy if bias (systematic errors) can be removed. For different assumptions of the bias, two possible test statistics may be used for this test.

1. If the bias is assumed to be the same for all values of $y_i$, and the magnitude of the bias can be estimated by the mean error ($\bar{d}$) of the technique, the test statistic is:

$$\chi^2_{(n-1)df} = \frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{\sigma_h^2} = \frac{\sum_{i=1}^{n} d_i^2 - n\bar{d}^2}{\sigma_h^2}$$  \hspace{1cm} (2.15)

where $\chi^2_{(n-1)df}$ is a chi-square distributed variable with $n - 1$ degrees of freedom under assumptions explained in section 2.3.1.2.

2. If the bias is assumed to increase or decrease directly with the true values in a linear fashion, the bias-free accuracy test requires fitting a linear regression of the estimated values on the true values, and computing sum of squared errors ($SSE$) about this regression. In this case, the test statistic is:

$$\chi^2_{(n-2)df} = \frac{SSE}{\sigma_h^2}$$  \hspace{1cm} (2.16)

where $\chi^2_{(n-2)df}$ is a chi-square distributed variable with $n - 2$ degrees of freedom under assumptions explained in section 2.3.1.2; and $\sigma_h^2$ in Equations 2.15 and 2.16 are defined to be the same as Equation 2.13.

The application of this bias-free test is the same as the first test. For convenience, the three chi-square test statistics (i.e., Equations 2.12, 2.14 and 2.15) will be denoted as Freese’s chi-square test I, II and III throughout this dissertation.
2.3.1.2 Underlying assumptions of Freese’s procedure

Since it was proposed, Freese’s (1960) procedure of accuracy testing has been widely used for evaluating the accuracy of forestry models (e.g., Evert 1981; Hazard and Berger 1972; Moser and Hall 1969). However, the basic assumptions underlying this procedure were not explicitly stated by Freese. Consequently, users of the procedure have not always known when the assumptions of the procedure have been violated. Realizing this problem, Reynolds (1984), and Gregoire and Reynolds (1988) presented discussions on the basic underlying assumptions and derivation of Freese’s procedure. The main points of these discussions are summarized in this section.

Freese’s (1960) procedure to determine how well a model will perform involves formulating statistical hypothesis testing for the distribution of the errors, \( d \), or the relative errors, \( d/y \). In discussing the distributions of \( d \), Reynolds (1984, p.455-457) indicated that two types of distributions may relate to \( d \). These are the conditional distribution of \( d \) for a given input vector, \( x \), (i.e., the values of independent variables), and the unconditional distribution of \( d \) where \( d \) is generated from input vectors that are randomly selected from all possible \( x \). The accuracy related to the unconditional distribution of \( d \) is called the average accuracy, a measure of the “average” accuracy over all values of \( x \). The basic assumption underlying Freese’s chi-square test I is that the \( n \) differences are a random sample from the unconditional distribution of \( d \), and the \( d_1, d_2, \ldots, d_n \) are independent and identically distributed (iid) normal variables.

The derivation of Freese’s (1960) test I as shown by Reynolds (1984 p.457) requires the user to specify values of \( e \) and \( \gamma \) such that if:

\[
P(|d| \leq e) \geq 1 - \gamma
\]  

(2.17)

the accuracy of a model will be acceptable. Since \( \gamma \) is usually small, the acceptance of the model means that the differences between the estimated and observed values should
be less than \( e \) with a high probability. In other words, if at least \( 100(1-\gamma) \) percent of the absolute values of model errors, \( d \), are less than or equal to \( e \), the model is acceptable. If the normally distributed variable, \( d \), can be further assumed to have a zero mean \((\mu_d = 0)\) and a variance, \( \sigma^2 \), Reynolds showed that the accuracy statement represented by Equation 2.17 will be satisfied if:

\[
e^2 \geq \sigma^2 \chi^2_{1-\gamma}(1)
\]  

(2.18)

where \( \chi^2_{1-\gamma}(\nu) \) is the \( 1-\gamma \) percentile of chi-square distribution with \( \nu \) degrees of freedom. Equation 2.18 results since \( d^2/\sigma^2 \) has a chi-squared distribution with 1 degree of freedom under these assumptions, and thus:

\[
P\left( \frac{d^2}{\sigma^2} \leq \chi^2_{1-\gamma}(1) \right) = 1 - \gamma
\]  

(2.19)

Also, since \( \chi^2_{1-\gamma}(1) = z^2_{1-\gamma/2} \), Equation 2.18 can be rewritten as:

\[
\sigma^2 \leq e^2/z^2_{1-\gamma/2}
\]  

(2.20)

This inequality represents a variance bound of the errors, \( d \), for satisfying the accuracy requirement given as Equation 2.17. Freese’s test I is designed to test whether the variance of \( d \) exceeds this bound by setting \( \sigma^2 = e^2/z^2_{1-\gamma/2} \), and testing the hypothesis:

\[
H_0 : \sigma^2 \leq \sigma^2_H \text{ vs. } H_1 : \sigma^2 > \sigma^2_H
\]  

(2.21)

Although not explicitly stated by Freese, the derivation of this procedure assumes that the model estimation is not biased \((i.e., \mu = 0)\). If the model estimation is biased, in order to satisfy Equation 2.17 for a given pair of \( e \) and \( \gamma \), the true variance of \( d \) would need to be even smaller than the bound (Equation 2.20). If the bias is large enough, for given values of \( e \) and \( \gamma \), it may not be possible to meet the accuracy requirement represented by Equation 2.17 no matter how small the true variance of \( d \).
The underlying assumptions and derivation of Freese's (1960) chi-square test II are similar to those of test I. The difference is that the variable being examined is the relative error, \( d/y \), instead of \( d \), where \( y \) is the actual observed value of the dependent variable. The underlying assumption of this test is that the unconditional distribution of \( d/y \) is normal with a zero mean and a variance, \( \sigma^2_{d/y} \). Similar to test I, Reynolds (1984) indicated that the specified values, \( p \) and \( \gamma \) were used to establish the accuracy requirement:

\[
P(100|d/y| \leq p) \geq 1 - \gamma \tag{2.22}
\]

Under the assumptions above, and similar to test I, the following relationship is true:

\[
P\left( \frac{(d/y)^2}{\sigma^2_{d/y}} \leq \chi^2_{1-\gamma}(1) \right) = 1 - \gamma \tag{2.23}
\]

The accuracy requirement (Equation 2.22) will be satisfied if:

\[
p^2 \geq \sigma^2_{d/y} \chi^2_{1-\gamma}(1)100^2 \tag{2.24}
\]

or if:

\[
\sigma^2_{d/y} \leq \left[ \frac{p}{(100 \chi_{1-\gamma/2})^2} \right]^2 \tag{2.25}
\]

To confirm the accuracy requirement (Equation 2.22), Freese's (1960) procedure is used to determine whether the variance of the unconditional distribution of \( d/y \) exceeds the variance bound (Equation 2.25) by testing the hypotheses:

\[
H_0 : \sigma^2_{d/y} \leq \sigma^2_p \text{ vs. } H_1 : \sigma^2_{d/y} > \sigma^2_p \tag{2.26}
\]

where \( \sigma^2_p = \frac{p^2}{(100 \chi_{1-\gamma/2})^2} \). Reynolds (1984, p.458) also indicated that the distributional assumptions between Freese's test I and II were contradictory, since if the distribution of \( d \) is normal then the distribution of \( d/y \) will not be normal and vice versa. Therefore,
when applying these two tests, the examination of normality should be conducted for the appropriate variable, $d$ or $d/y$.

The normality assumption provides the crucial link to the $\chi^2$ distribution, which is central to Freese's (1960) procedure of accuracy testing. Gregoire and Reynolds (1988) studied the sensitivity of Freese's procedure to non-normality through simulation experiments. In their study, three non-normal distributions were used. These were a uniform distribution with the interval from -1 to 1, a double exponential distribution, and a mixture of two normal distributions, $N(0,1)$ and $N(0, \sigma^2_a)$, which was mixed with a ratio of 9-to-1 favoring the $N(0,1)$, where $\sigma^2_a = 2, 3, 4,$ and $5$. They concluded that nonnormality would very seriously distort the operating characteristic and power of Freese's chi-square tests, and the deterioration increased with increasing sample size.

From the derivations of Freese's (1960) tests I and II, the derivation of Freese's chi-square test III is straightforward. The maximum likelihood estimate for the variance of $d$ is $\sum_{i=1}^{n} (d_i - \overline{d})^2/n$, instead of $\sum_{i=1}^{n} d_i^2/n$ if the expected value of $d$ is not zero and unknown. Therefore, the measure of the accuracy of a model should be $\sum_{i=1}^{n} (d_i - \overline{d})^2$, instead of $\sum_{i=1}^{n} d_i^2$. However, Reynolds (1984, p.457) indicated that this test was only applicable when the bias of a model is removable. For the bias to be removable, the user of a model must understand the nature of the model bias (e.g., negative or positive), and the magnitude of the model bias must be estimated using an appropriate estimate (two situations indicated by Freese). The simplest situation is when the bias is the same for all estimated values, and the mean of errors is the appropriate estimate of the bias.

The difference between Freese's (1960) test III and the others is that this test is used to determine the accuracy of a model after the bias is eliminated. The source of errors cannot be separated when using Freese's test I and II; the total errors are used to compare with the specified requirement of accuracy (hypothesized variance). If the model is unbiased, then the results using Freese's test I and III would be the same under
the assumptions of normality and constant variance.

2.3.1.3 Modifications of Freese's procedure

Three modifications to Freese's (1960) procedure have been suggested to improve its usefulness for evaluating the predictive ability of forestry models. First, the specification of the error or percent error limit, \( e \) or \( p \), is sometimes not straightforward, and it depends on the knowledge of the user about the desired level of accuracy. When different users are involved, one user's specified value may be too accurate or not accurate enough for others (Rennie and Wiant 1978). To overcome this problem, some researchers (Bell and Groman 1971; Boehmer and Rennie 1976) inverted Freese's testing procedure to calculate the maximum anticipated error (\( E^* \)), and percent error (\( E^*(\%) \)), which are the maximum \( e \) and \( p \), respectively that will result in rejecting the model. Rennie and Wiant (1978) gave the formulae for calculating these values. These formulae were:

\[
E^* = \left[ \frac{\sum_{i=1}^{n} d_i^2}{\chi^2_{1-\alpha(n)}} \right]^{1/2}
\]  
\[2.27\]

\[
E^*(\%) = \left[ \frac{\sum_{i=1}^{n} (d_i/y_i)^2}{\chi^2_{1-\alpha(n)}} \right]^{1/2}
\]  
\[2.28\]

where \( \gamma \) is the specified probability level associated with the allowable error limit, \( e \), or \( p \); and \( \alpha \) is the specified probability of the type I error for testing the hypotheses given as Equations 2.21 or 2.26. If the user-specified \( e \) or \( p \) is larger than the calculated \( E^* \) or \( E^*(\%) \), the tested model should be acceptable. If the user-specified \( e \) or \( p \) is less than the calculated \( E^* \) or \( E^*(\%) \), the model will be unacceptable. Ek and Monserud (1979) used these calculated maximum anticipated errors as criteria to evaluate two stand growth models; the model with smaller \( E^* \) and \( E^*(\%) \) was considered to be better.

Reynolds (1984) stated that since the null hypothesis of Freese's (1960) original procedure was \( H_0 : \sigma^2 \leq \sigma^2_h \), the model would be judged to be adequate unless there was
strong evidence to show that it was not. This formulation may result in favor of the acceptance of a model. Reynolds then suggested modifying Freese's null hypothesis to $H_0: \sigma^2 \geq \sigma^2_h$, which would then be tested by using the same test statistic of Freese's test I (Equation 2.12). This modified test will reject the null hypothesis and accept the model at a significance level $\alpha$ when the calculated statistic $\left( \chi^2_{(n)df} \right)$ is below $\chi^2_\alpha(n)$, the lower tail 100$\alpha$ percentile of the chi-square distribution with $n$ degrees of freedom. Reynolds indicated that this modified testing procedure was more conservative than the original procedure since stronger evidence is demanded to judge a model to be adequate. He stated that it was probably preferable to the original procedure for most model users who need to be reasonably sure that the model would meet their requirements.

Reynolds also attempted to relate the maximum anticipated error, $E^*$, to a confidence interval estimate. That is, under the assumptions that $d$ is a normally distributed variable with a zero mean and a variance, $\sigma^2$, a 100$(1 - \alpha)$ percent confidence interval for the $1 - \gamma$ quantile of the distribution of $|d|$ can be constructed as: $(\epsilon_L, \epsilon_U)$, and where

$$\epsilon_L = \left[ \frac{\chi^2_{1-\gamma}(1) \sum_{i=1}^{n} d_i^2}{\chi^2_{1-\alpha/2}(n)} \right]^{1/2}$$

and

$$\epsilon_U = \left[ \frac{\chi^2_{1-\gamma}(1) \sum_{i=1}^{n} d_i^2}{\chi^2_{\alpha/2}(n)} \right]^{1/2}$$

Instead of making a decision to accept or reject a model, this interval estimate provides some indication how large the errors are expected to be if the model is used for estimation. That is, the interpretation of $(\epsilon_L, \epsilon_U)$ is that one can be 100$(1 - \alpha)$ percent sure that 100$(1 - \gamma)$ percent of model errors will be within $(\epsilon_L, \epsilon_U)$, if the model is repeatedly used for estimation.

Gregoire and Reynolds (1988 p. 306) further discussed Freese's (1960) original procedure and Reynolds' (1984) modified procedure. They indicated that Freese's procedure may be too liberal for typically small values of $\gamma$ and $\alpha$ in the sense of not requiring
reasonable evidence of accuracy attainment, and Reynolds' modified procedure may be too conservative in the sense of requiring too stringent evidence for accepting a model. They then suggested a three-decision procedure for combining the advantages from two previous modifications of Freese's procedure (i.e., the modification of calculating the maximum anticipated error \(E^*\) and Reynolds' (1984) modified testing procedure). The decision rules of their suggested three-decision procedure are:

1. If \(e < \epsilon_L\), conclude that the actual accuracy of the model is less than that required (specified as \(e\) and \(\gamma\));

2. If \(\epsilon_L \leq e < \epsilon_U\), no confident conclusion can be made for the accuracy of the model based on the available data; and

3. If \(e \geq \epsilon_U\), conclude that the actual accuracy of the model exceeds the required level.

Among these three modifications of Freese's (1960) procedure, the idea of calculating the maximum anticipated error or percent error has been applied by many researchers for forestry model testing (e.g., Boehmer and Rennie 1976; Ek and Monseu 1979).

2.3.1.4 Previous applications of Freese's procedure

Evert (1981) applied Freese's (1960) test I to validate five mortality models, which were developed from thirty permanent sample plots provided by the Canadian Forestry Service at the Petawawa National Forestry Institute, Chalk River, Ontario. The evaluation data included five periods of measurements of three permanent sample plots belonging to the same series of plots that were used for model estimations. They were not used in the development of the models. Instead of specifying the error limit, \(e\), the standard error of estimate of the evaluated model was set to be the hypothesized variance \(\sigma_h^2\) as the
required accuracy. The result was that the calculated Freese’s chi-square statistics were all less than the tabulated values at a significance level of 0.05. All five models were considered to be acceptable.

Nevers and Barrett (1966) applied Freese’s (1960) test I to determine the accuracy of height accumulation volumes from penta prism caliper measurements. Based on the estimated volumes from 20 standing white pine (*Pinus albicaulis* Engelm.) trees, and the actual volumes from the measurements made by climbing the trees, they concluded that at least 95% of the estimated tree volumes from the penta prism were within 2 cubic feet (0.0566 m$^3$) of their actual volumes (no $\alpha$ level reported) for the test. In their application, the Wilcoxon’s signed rank test was used first to check the unbiasedness of $d$; it showed the presence of a highly significant bias in estimated volumes. Similarly, Cost (1971) also applied Freese’s test I to evaluate the accuracy of standing-tree volume estimates based on McClure Mirror caliper measurements. He first measured the upper diameters of 25 sample trees using the instrument to determine the estimated volumes, and then felled the sample trees to determine the true volumes. Wilcoxon’s signed rank test was also used to test the unbiasedness of $d$ in this application. The accuracy requirement, $e$, was set to 1.5 cubic feet (0.0425 m$^3$) with $\gamma$ set as 0.05. The conclusion was that at least 95% of errors of the individual tree volume estimates from McClure Mirror caliper measurements were expected to be within 1.5 cubic feet (no $\alpha$ level reported).

Moser and Hall (1969) used Freese’s test II to check the accuracy of a growth and yield function of basal area, which was obtained from the mathematical integration of a growth-rate equation. The measurements from 40 permanent sample plots were used as the actual values; the required accuracy was that the predicted values should be within 10 percent of the actual values. The conclusion was that for the 12-year period, the derived function provided the necessary accuracy for predicting average basal area.
Hazard and Berger (1972) studied the applicability of a ponderosa pine (*Pinus ponderosa* Laws.) volume table when applied to different species and geographical locations. A random sample of 50 trees was selected from the tested area, which had a uniform distribution of trees ranging in dbh from 5 to 40 inches (12.7 to 101.6 cm). The computed volumes based on the measurements using a Barr and Stroud optical dendrometer, model FP-15, were used as the true volumes for comparison. A *p* value of 10 percent was specified as the required accuracy. The result of Freese's (1960) test II showed that the volume table failed to meet the accuracy requirement. They then studied the differences between the estimated and true volumes, and found that the differences ranged from -62.3 cubic feet (−1.7643 m³) to +5.6 cubic feet (0.1586 m³). Based on this result, they concluded that Freese's test I is influenced considerably by bias and not just variation between two methods of volume estimation.

Thies and Harvey (1979) applied Freese's (1960) chi-square test II to determine the accuracy of a photographic method for measuring tree defect. The required accuracy was that the estimated defect areas by the photographic method were within ±5% (*i.e.*, *p*) of the true areas. The conclusion was that the photographic method met the required accuracy standard.

### 2.3.2 Procedures for validating forestry simulation models

Simulation models are widely used to help understand complicated systems or processes. In forestry, tree and stand simulation models have become an important tool used for decision-making in forest management. With the increased use of simulation models has come an awareness of the need to validate the models before a model can be used with confidence.

A bibliography on the validation of simulation models is given by Balci and Sargent (1984). The philosophical aspects of validating simulation models were well discussed
in the papers by Van Horn (1971), Naylor and Finger (1967), Mankin et al. (1977). In general, these authors agreed that simulation model validation was a multi-stage process of examination (Naylor and Finger 1967, p.95). The initial stages of this process involve the examination of the structure and operation of the model to make sure that it is working as intended (verification stages). The next stages of this process involve comparing model output with what is observed in the real system (validation stages).

Many statistical procedures have been proposed in computer simulation and other literature for validation. Examples include the two-sample Hotelling's $T^2$ test for the models with multivariate response variables (Balci and Sargent 1982); the decision-analytic approach for minimizing the expected losses based on the type I and II errors resulting from model acceptance or rejection (Greig 1979); and the spectral analysis approach or statistical hypothesis testing procedure for validating time-series simulation models (Fishman and Kiviat 1967; Naylor and Finger 1967; Feldman et al. 1984). However, these proposed procedures are only applicable for their specific designed condition. They are in general not appropriate for testing forestry prediction models, especially tree volume models.

In forestry, Reynolds et al. (1981) suggested that statistical hypothesis testing approaches, which combine $n$ independent tests for testing the same hypothesis into one overall test, may be used. They had several reasons for this suggestion. First, the input variables ($x$) of forestry simulation models usually involve a wide range of possible levels. For example, the initial stand and site characteristics (e.g., stand age, site index and basal area) input to a stand model may include a wide range of values for a given forest management unit. Therefore, the goal in developing and using a forestry simulation model is to be able to predict the conditional distribution of the variable of interest ($y$). In other words, conditional, not unconditional accuracy, should be emphasized in simulation model validation. The appropriate null hypothesis for validating forestry simulation models is then $H_0 : F(y \mid x) = G(y \mid x)$ for all $-\infty < y < \infty$ and $x \in A$, where
\[ F(y \mid x) \text{ and } G(y \mid x) \text{ are the true and predicted conditional distribution of } y \text{ given } x, \text{ respectively; and } A \text{ represents some specified set of } x \text{ values. Second, using a sample set of } n \text{ pairs of observations, } (y_i, x_i) \text{ selected from a population, the method to compare the simulated distribution, } G(y \mid x), \text{ with } F(y \mid x), \text{ is usually to make } m \text{ runs of the simulation model for each given } x_i. \text{ This will give } m \text{ independent and identically distributed values, say } z_i = (z_{i1}, z_{i2}, \ldots, z_{im}), \text{ where } z_i \text{ represents } m \text{ predicted values conditioned on } x_i. \text{ In this way, Reynolds et al. (1981) indicated that since each of the } n \text{ pairs of values } (y_i, z_i) \text{ for } i = 1, \ldots, n, \text{ was generated under different values of } x_i, \text{ they might represent different initial stand and site conditions. To compare the true and predicted conditional distributions, the data obtained in this way should not be grouped into one large set for comparing the conditional distribution. Instead, the } n \text{ independent pairs must be kept separate. Also, since there is only one actual observed value } (y_i) \text{ in every individual pair, any test applied to an individual pair would not have much power by itself.}

They then proposed six parametric testing procedures and three nonparametric procedures. One of these testing approaches assumes that } F(y \mid x_i) \text{ and } G(y \mid x_i) \text{ are normal distributions with the same mean and variance. Because of the independence of } y_i \text{ and } z_{ij}, \text{ the statistic for each pair of } (y_i, z_i) \text{ is:}

\[ t_i = \frac{y_i - z_i}{s_i \sqrt{1 + \frac{1}{m}}} \]

which has a } t \text{-distribution with } m - 1 \text{ degrees of freedom, where } z_i = \sum_j z_{ij} / m \text{ and } s_i^2 = \sum_j (z_{ij} - \bar{z}_i)^2 / (m - 1). \text{ The sum of these test statistics would be:}

\[ U = \sum_{i=1}^{n} t_i \]

Under the null hypothesis and for } m \geq 4, \text{ the statistic}

\[ U^* = U / \sqrt{\frac{n(m - 1)}{m - 3}} \]
has approximately a standard normal distribution. Thus, the test can be carried out by comparing the observed value of the statistic with the appropriate critical value from the standard normal distribution. The idea for formulating other proposed testing approaches is very similar. Reynolds et al. (1981) also indicated that the difference between this and Freese’s (1960) procedure was that Freese’s procedure is useful for determining whether the predicted value $z_i$ was close to the actual value $y_i$, but their procedures are useful for determining whether the distribution of $z_i$ is close to the distribution of $y_i$.

### 2.3.3 Reynolds’ estimation procedures

Freese’s (1960) procedure and its modifications were designed to compare the accuracy of a model against an established standard stated by users. However, in many cases of model validation, the question that the user needs to answer is not whether the model meets a particular standard, but rather what is the magnitude of the error that can be expected when the model is used for estimation or prediction. Reynolds (1984 p.461) stated that for the latter, statistical estimation approaches are more appropriate. Instead of simply accepting or rejecting a model, estimation approaches provide the user with some indication of how far predictions using the model will be from their actual values. In this way, they may be more meaningful for the user. Also, they are applicable even when bias is present.

Reynolds (1984) then suggested various interval estimates (confidence interval, prediction interval and tolerance interval) for the mean of $d$. In general, if the users wanted to estimate the true mean of $d$ based on the observed errors of a model, a $100(1-\gamma)$ percent confidence interval should be used. If the model was used for making future prediction, a $100(1-\gamma)$ percent prediction interval of a future value of $d$ would be appropriate. The tolerance interval might be used when the users want to know the upper and lower limit (tolerance limit) for the distribution of $d$, which contains at least a given percent of the
entire distribution of $d$. As with Freese’s (1960) procedure, all these suggested interval estimates require the normality assumption for $d$, except for the non-parameteric prediction interval for $k$ future values of $d$. For any continuous distribution of $d$, this prediction interval is defined as:

$$P(\text{Min}(d_i) \leq d_j \leq \text{Max}(d_i)) = 1 - \gamma_1$$

(2.31)

and

$$\gamma_1 = \frac{n(n-1)}{(n+k)(n+k-1)}$$

(2.32)

where $\text{Min}(d_i)$ and $\text{Max}(d_i)$, respectively are the smallest and largest errors of $d$ in the current sample; $d_j$ for $j = 1, \ldots, k$, are $k$ future values of $d$; and $\gamma_1$ is the probability that all $k$ future errors, $d_j$, are within $(\text{Min}(d_i), \text{Max}(d_i))$. This interval means that the smallest and largest values of $d$ from the current sample give a $100(1 - \gamma_1)$ percent prediction interval for all $k$ future values of $d$.

To assist with the application of Reynolds’ (1984) interval estimates procedures, Rauscher (1986) developed a PC BASIC program, ATEST, which accepts the input of errors ($d_i$ for $i = 1 \ldots n$) and outputs the calculated confidence interval, prediction interval and tolerance interval. The program also provides a routine to test the normality for the error, $d$, and relative errors, $d/y$. If the normality assumption is not met, the program calculates various intervals based on the sample 10-percent trimmed mean and the jackknifing variance, which are more robust to non-normality, instead of using the common mean and variance.

### 2.3.4 Other fixed sample size procedures

Besides the procedures discussed in previous subsections, other statistical procedures for comparing the estimated and actual values of a model were also found in forestry literature. A brief discussion of these procedures follows.
Chapter 2. Literature Review

The first procedure found was a paired t test\(^1\) for the null hypothesis \(H_0 : \mu_d = 0\), where \(\mu_d\) is the mean for the normally distributed errors, \(d\). It is well known that the test statistic of this test is \(t_{n-1} = \bar{d}/S_d\), where \(\bar{d}\) and \(S_d\) are the sample mean and standard error of \(\bar{d}\), respectively, and \(t_{n-1}\) has a Student's \(t\) distribution with \(n - 1\) degrees of freedom. Since \(\bar{d}\) and \(S_d\) are sample measures of bias and precision of a technique, respectively, Freese (1960 p.145) indicated that the paired \(t\) test was not suitable for the purpose of accuracy testing, because it compares one form of accuracy (bias) to the other form (lack of precision), frequently with anomalous results.

The analysis of variance (ANOVA) procedure appears to be a logical choice to compare the estimated and actual values of a technique (model) when the observations from a population can be grouped (e.g., tree dbh or height classes). Reynolds et al. (1981, p.352) pointed out that the ANOVA procedure was not strictly appropriate for model validation since it required the same variance within each group, and this was not likely to be true in many applications of forestry model testing.

West (1983, p.186) used the Kolmogorov-Smirnov (K-S) test of goodness of fit to evaluate the predictive ability of a forest simulation model. Naylor and Finger (1967) indicated that this distribution-free (nonparametric) test was concerned with the degree of agreement between a simulated (predicted) and an observed series of the model's response variable.

Finally, the regression approach can be used to regress the observed values on the estimated or predicted values. A test is then used to confirm whether the resulting regression equation has an intercept and a slope which are not significantly different from zero and unity, respectively. This procedure was used to assess the predictive abilities of three stand growth models by Daniels et al. (1979) and the accuracy of aluminum band

---

\(^1\)The pairs in model testing will be the predicted and observed values, \(\hat{y}_i\) and \(y_i\) for \(i = 1 \ldots n\), which are determined by different sets of the values for independent variables, \(x_i\).
dendrometers for measuring the diameter growth of trees by Auchmoody (1976).

In an interesting study conducted by Ek and Monserud (1979), all procedures presented in this subsection, as well as Freese’s (1960) procedure, were used to evaluate two tree growth models (e.g., FOREST and SHAF). However, no conclusion was made on which procedure was better for that application.

2.4 Variable Sample Size Procedures for Testing Forestry Populations

In the last section, procedures for forestry model validation were reviewed. These procedures can be used for detailed evaluation of the accuracy of a forestry model. However, to apply these procedures, a sufficiently large sample set with a preset size (e.g., a large sample of trees or plots) must be collected before any testing or estimating process can be carried out. These procedures are not efficient for applicability testing of forestry models, because the purpose of an applicability test is to classify an existing model into two wide classes (i.e., acceptable or unacceptable), and it is possible that a decision about the applicability of the model can be made with a small sample size. This is particularly true when the actual accuracy of the model is far below or above the accuracy requirement specified by the user. Therefore, alternatives to the fixed sample size procedures are needed. In sampling theory, a variable sample size procedure (sequential sampling) may be used as an alternative to the fixed sample size procedures.

In the statistics literature, sequential sampling procedures for testing or estimating the parameters of a population are generally called sequential analysis approaches. Conceptually, sequential analysis is a method of statistical inference with the characteristic feature that the number of observations required by the procedure is not determined in advance (Wald 1947, p.1). The basic theory of sequential analysis for statistical hypothesis testing was developed by Wald in America and Barnard in Britain at about the same
time during World War II (Wetherill and Glazebrook, 1986). Because of its value, the sequential approach was classified as restricted in America and was not made available for wider use until after World War II (Wald 1947, p.3). The advantage of a sequential analysis approach, as applied to testing statistical hypothesis, is that it requires, on average, a substantially smaller number of observations than equally reliable test procedures based on a fixed sample size procedure (Wald, 1947).

In discussing the problem of considering the use of a model fitted for one population for application to a different population, Marshall and LeMay (1990) indicated that three types of approaches could be applied to this problem. First, a new sample set could be selected, and a new model could be fitted (independent sampling approach). Second, the parameters of the existing model could be adjusted using new sample data (adjustment approach). Third, a sequential sampling plan could be used to sample until a decision regarding the applicability of the existing model is made, and a new model may be fitted only if necessary (sequential sampling approach). They concluded that among these three approaches, the sequential sampling approach is the only one which explicitly tests the applicability of the existing model. Also, for a given sampling design and precision requirement, a sequential sampling approach requires the least number of samples.

Brack and Marshall (1990) made an attempt to apply the sequential approach for checking model estimation. In their study, the sequential sampling approach was used to check the applicability of a mean dominant height (MDH) equation for radiata pine (Pinus radiata D. Don) plantation in specific cutting units in New South Wales, Australia. In their sequential testing plan, the checking of model estimations was considered to be a sequence of independent Bernoulli trials with two possible outcomes (acceptable and unacceptable) by comparing each observed relative error, \( d/y \), to a preset limit (5% or 10% of actual value). A binomial variable (the sum of the number of the acceptable predictions) was then used to measure the accuracy of the model. By using this procedure,
they showed that the sequential approach could reach a decision on whether the estimated MDH of the model was acceptable or unacceptable in as little as three plots (sample size). However, since a Bernoulli variable could take only the value of 0 or 1, much information about model errors was ignored. They suggested that other better measures of model accuracy should be used instead of a binomial variable in order to use the sequential approach for model testing.

These two initial discussions and attempts indicated that sequential sampling procedures may be potentially superior to fixed sample size procedures for the purpose of applicability testing of forestry models.

Many applications of sequential sampling plans to pest surveys have been found in the forestry and agriculture literature. In discussing the advantages of sequential sampling plans for forest pest surveys, Fowler and Lynch (1987a) stated that the sequential approach is intuitively appealing in that very few observations (little work, time, or money) are needed to make a terminating decision when insect populations are very sparse or very abundant. Such plans required, on the average, only 40 to 60% as many observations as an equally reliable fixed-sample procedure. Therefore, sequential procedures should find fairly wide applicability where the objective is to classify insect densities or damage into two or three broad classes (e.g., treatment and non-treatment) for management decisions and where observations are time-consuming, costly, or destructive. These properties of the sequential sampling procedures exactly meet the desired criteria for the alternative procedures of the applicability testing of forestry models.

Based on the similarity of the purposes between a forest pest survey and an applicability test of forestry prediction models, and also on the potential advantages of the sequential approach for achieving the objective of this research, a search of forestry and other literature for the theory and application of sequential analysis approaches was carried out.
2.4.1 Sequential probability ratio tests

2.4.1.1 Wald’s general decision rule for SPRT plans

The most widely used sequential sampling plans in forestry are the sequential probability ratio test (SPRT) plans. The theory of SPRT plans was developed by Wald (1945, 1947). A detailed review on its basic theory and application will be given first.

The SPRT, in general, is a statistical procedure mainly for hypothesis testing. It is designed to test a simple hypothesis, \( H_0 : \theta = \theta_0 \) against another simple hypothesis, \( H_1 : \theta = \theta_1 \), where \( \theta \) is the parameter of interest, and \( \theta_0, \theta_1 \) are two class limits of population classification for decision-making purposes (e.g., acceptable and unacceptable, or no need for treatment and need for treatment, etc.).\(^2\) To design an SPRT procedure, besides the two class limits, \( \theta_0 \) and \( \theta_1 \), the desired levels of probability for making type I and II errors, \( \alpha \) and \( \beta \), are all preset. However, the sample size, \( n \), is not preset, but is a random variable (variable sample size), which depends on the values of the observations taken in each random experiment.

For any given sample, the SPRT divides the \( n \)-dimensional sample space \( (S_n = \{x_1, x_2, x_3, \ldots, x_n\}) \) into three mutually exclusive sets (e.g., \( S_n = \{A_{n0} \cup A_{n1} \cup A_n\} \)) for each \( n \). The observations, \( x_1, \ldots, x_n \), are sequentially sampled from the given distribution (population). After the first random observation \( x_1 \) is taken, \( H_0 \) is accepted if \( x_1 \) lies in \( A_{10} \); \( H_0 \) is rejected if \( x_1 \) lies in \( A_{11} \), or a second observation \( x_2 \) is taken if \( x_1 \) lies in \( A_1 \). If \( x_1 \) is in \( A_1 \) and a second observation \( x_2 \) taken, \( H_0 \) is accepted, \( H_0 \) is rejected, or a third observation \( x_3 \) is needed to taken according to whether the point \( (x_1, x_2) \) lies in \( A_{20}, A_{21} \) or \( A_2 \). This process stops only when \( H_0 \) has been either accepted or rejected.

In practical applications, the basic theory is realized through the following procedures. At each stage of the test, an observation is taken at random from the given distribution.

\(^2\)In a statistical context, \( \theta_0 \) and \( \theta_1 \) are also called the null and alternative parameters, respectively.
and the likelihood ratio:

\[ R_n = \prod_{i=1}^{n} \left[ \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} \right] \]  

(2.33)

is computed from the total number of observations taken up to that point \( n \), where \( f(x_i, \theta_0) \) and \( f(x_i, \theta_1) \) are the probability density functions under the null and alternative hypotheses, respectively. Based on the computed probability ratio, \( R_n \), one of the following decisions will be made for each stage:

\[ R_n \leq B, \text{ stop sampling and accept } H_0; \]
\[ R_n \geq A, \text{ stop sampling and reject } H_0, \text{ or} \]
\[ B < R_n < A, \text{ continue sampling.} \]

This is called the decision rule of Wald’s SPRT procedure. \( A \) and \( B \) are two constants that are chosen so that \( 0 < B < A \), and so that the test has the desired preset values of \( \alpha \) and \( \beta \). Wald (1947) gave the approximate formulae for determining \( A \) and \( B \) as:

\[ A \approx \frac{1 - \beta}{\alpha} \]
\[ B \approx \frac{\beta}{1 - \alpha} \]

The approximate equalities are obtained by ignoring the possible overshooting (exceeding) of the decision boundaries. In other words, \( R_n \geq A \) is approximated as \( R_n = A \), and \( R_n \leq B \) is approximated as \( R_n = B \). \( R_n \) can be simplified by taking the natural logarithm \( \ln \) of each density function ratio in the product, which yields:

\[ \ln(R_n) = Z_n = \sum_{i=1}^{n} z_i, \text{ and } z_i = \ln \left[ \frac{f_1(x_i, \theta_1)}{f_0(x_i, \theta_0)} \right] \]

The decision procedure is then (1) if \( Z_n \geq a \), stop sampling and reject \( H_0 \); (2) if \( Z_n \leq b \), stop sampling and accept \( H_0 \); and (3) if \( b < Z_n < a \), continue sampling, where \( a = \ln(A) \) and \( b = \ln(B) \).
For some probability distributions (e.g., normal or binomial distribution), the cumulative sum of the observed log-likelihood ratio, $Z_n$, is equivalent to using the cumulative sum of the observations, $\sum_{i=1}^{n} x_i$, which is simple to calculate by setting $Z_n = a$, $Z_n = b$ and solving for $\sum_{i=1}^{n} x_i$ to determine the upper rejection and lower acceptance boundaries, respectively. These decision boundaries are represented by two parallel lines ($L_0$ and $L_1$). The decision process is then: (1) if $\sum_{i=1}^{n} x_i \geq h_2 + sn$, stop sampling and reject $H_0$; (2) if $\sum_{i=1}^{n} x_i \leq h_1 + sn$, stop sampling and accept $H_0$; or (3) $h_1 + sn < \sum_{i=1}^{n} x_i < h_2 + sn$, continue sampling, where the intercepts, $h_1$ and $h_2$, and the slope $s$ are calculated from the knowledge of the underlying distribution and the specified values of $\theta_0, \theta_1, \alpha$, and $\beta$. The decision boundaries, $h_1 + sn$ and $h_2 + sn$, are also called the acceptance and rejection lines, respectively.

Fowler and Lynch (1987a) indicated that in the derivation of SPRT, the assumptions made by Wald (1947) were:

1. Only one observation (sampling unit) is taken at each stage of sequential process;

2. There is no predetermined upper limit to the number of observations taken before a terminating decision is made; and

3. A terminating decision is made as soon as a decision boundary is crossed.

In applying the SPRT procedure for hypothesis testing, the probability density function of the random variable of interest must be known (e.g., normal, binomial, etc.). Also, two class limits must be set for the null ($\theta_0$) and alternative ($\theta_1$) hypothesis parameter values of the underlying distribution, and the associated error probabilities, $\alpha$ and $\beta$, must also be specified. Having met these conditions, an SPRT can be established and its decision boundaries given.

Before an SPRT procedure is actually carried out, it is desirable to know the properties of the test in order to determine its potential feasibility (e.g., expected sampling cost).
Wald (1947) described the properties of an SPRT over all possible values of the random variable of interest with Operating Characteristic (OC) and Average Sample Number (ASN) functions, and he also derived the general formulae of the OC and ASN functions for an SPRT procedure.

2.4.1.2 Wald’s general OC function for SPRT plans

Wald’s (1947) Operating Characteristic (OC) of an SPRT procedure was defined as:

\[ L(\theta) = P(\text{accept } H_0 \mid \theta), \]

which is the probability of accepting the null hypothesis \((H_0)\), assuming that \(\theta\) is the true parameter. \(L(\theta)\) is called the OC value for given \(\theta\). Wald derived the general OC function of any SPRT plan (see, Wald 1947, p.50) as:

\[ L(\theta) \equiv \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}} \]  \hspace{1cm} (2.34)

where \(h(\theta)\) is the nonzero solution of:

\[ \int_{-\infty}^{\infty} \left[ \frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^{h(\theta)} f(x, \theta) \, dx = 1 \]

if the distribution of \(x\) is continuous, or:

\[ \sum_x \left[ \frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^{h(\theta)} f(x, \theta) = 1 \]

if the distribution of \(x\) is discrete.

In order to explain how to derive the OC equation for a specific SPRT procedure based on Wald’s general OC function, an example for testing the standard deviation \((\sigma)\) of a normal variable with known mean \((\mu)\) is given. First, two class limits of the standard deviation \((\sigma_0 \text{ and } \sigma_1)\), and the desired probability levels of the type I and II error \((\alpha \text{ and } \beta)\), must be specified by the user. Based on these specified values, the hypothesis
$H_0 : \sigma = \sigma_0 \text{ vs. } H_1 : \sigma = \sigma_1$, can be established for classifying the true standard deviation of a normal variable $x$ with known mean $\mu$ into two classes. By replacing $A$ and $B$ in Wald's general function (Equation 2.34) using the specified error probabilities, the OC equation for this specific SPRT is:

$$L(\sigma) \approx \frac{\left(\frac{1 - \beta}{\alpha}\right)^h - 1}{\left(\frac{1 - \beta}{\alpha}\right)^h - \left(\frac{\beta}{1 - \alpha}\right)^h}$$

(2.35)

where $h = h(\sigma)$, which must be determined in order to calculate OC values using this equation. By Wald's (1947) definition, $h$ is the nonzero solution of:

$$\frac{1}{\sqrt{2\pi}\sigma} \left(\frac{\sigma_0}{\sigma_1}\right)^h \int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2}\right)^h e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = 1$$

(2.36)

Wald (1947) showed that the above expression could be simplified as

$$\sigma \left(\frac{\sigma_1}{\sigma_0}\right)^h = \sqrt{1/\left(\frac{h}{\sigma_1^2} - \frac{h}{\sigma_2^2} + \frac{1}{\sigma^2}\right)}$$

(2.37)

Instead of solving Equation 2.37 with respect to $h$, Wald (1947) suggested that it could be solved with respect to $\sigma$ to obtain:

$$\sigma = \sqrt{\frac{\left(\frac{\sigma_2}{\sigma_1}\right)^{2h} - 1}{\frac{h}{\sigma_1^2} - \frac{h}{\sigma_2^2}}}$$

(2.38)

With the use of Equations 2.35 and 2.38, the OC curve for this specific SPRT can be plotted. First, for any given value of $h$, we compute $L(\sigma)$ and $\sigma$ from Equations 2.35 and 2.38. The pair, $[\sigma, L(\sigma)]$, obtained in this way gives us a point on the OC curve with $\sigma$ and $L(\sigma)$ as the horizontal and vertical axis, respectively. The computation of the pair, $[\sigma, L(\sigma)]$, is then repeated for a sufficiently large number of values of $h$. Fowler (1978) indicated that $h$ from -4.0 to 4.0 with an interval 0.5 ($h \neq 0$) would usually be sufficient for constructing the OC curve of an SPRT.
2.4.1.3 Wald’s general ASN function for SPRT plans

The Average Sample Number (ASN) function is the expected (average) number of observations ($E_\theta(n)$) needed to make a terminating decision, assuming that $\theta$ is the true parameter. The general ASN function derived by Wald (1947) for any SPRT is:

$$E_\theta(n) \approx \frac{L(\theta) \ln(B) + [1 - L(\theta)] \ln(A)}{E_\theta(z)}$$  \hspace{1cm} (2.39)

and

$$E_\theta(z) = E(Z) = \mathbb{E}
\left[ \ln \left( \frac{f(x, \theta_1)}{f(x, \theta_0)} \right) \right]$$  \hspace{1cm} (2.40)

where $E_\theta(n)$ is the ASN value for given $\theta$; $E_\theta(z)$ is the expected value of log-likelihood ratio of $Z_i$; and $L(\theta)$ is the OC value for given $\theta$ obtained from Equation 2.34. To calculate $E_\theta(n)$, the OC value and $E_\theta(z)$ must first be determined for a given $\theta$.

In order to explain how to derive the ASN equation for a specific SPRT procedure based on Wald’s general OC function, the example for testing the standard deviation ($\sigma$) of a normal variable with known mean ($\mu$) is also given. By replacing $A$ and $B$ in Wald’s general ASN equation (Equation 2.39) with the specified values of $\alpha$ and $\beta$, and assuming the OC value ($L(\sigma)$) for given $\sigma$ has been calculated using the previous procedures, the ASN equation for this specific SPRT plan is:

$$E_{\sigma}(n) \approx \frac{L(\sigma) \ln(\frac{\beta}{1-\alpha}) + [1 - L(\sigma)] \ln(\frac{1-\beta}{\sigma})}{E_{\sigma}(z)}$$  \hspace{1cm} (2.41)

where $E_{\sigma}(z)$, the expected value of the log likelihood ratio for this test is:

$$E_{\sigma}(z) = \mathbb{E}
\left[ \ln \left( \frac{\frac{1}{\sigma_1} e^{-\frac{(x-\mu)^2}{2\sigma_1^2}}}{\frac{1}{\sigma_0} e^{-\frac{(x-\mu)^2}{2\sigma_0^2}}} \right) \right] = \mathbb{E}
\left[ \ln \left( \frac{\sigma_0}{\sigma_1} \right) + \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) (x - \mu)^2 \right]$$  \hspace{1cm} (2.42)

$$E_{\sigma}(z) = \ln \left( \frac{\sigma_0}{\sigma_1} \right) + \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sigma^2$$  \hspace{1cm} (2.43)

By replacing $E_{\sigma}(z)$ in Equation 2.41 with Equation 2.43, the ASN equation of an SPRT
for testing the standard deviation of normal variable with known mean becomes:

$$E_\sigma(n) \equiv \frac{L(\sigma) \ln\left(1 - \frac{\beta}{\alpha}\right) + [1 - L(\sigma)] \ln\left(\frac{1 - \beta}{\alpha}\right)}{\ln\left(\frac{\sigma^2}{\sigma_1^2}\right) + \frac{1}{2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) \sigma^2}$$  \hspace{1cm} (2.44)

where $\sigma^2$ is the true variance of the variable of interest. To construct the ASN curve for this test, the procedures are (1) $h$ values of the OC equation (Equation 2.35) are chosen from -4.0 to 4.0 with an interval of 0.5 ($h \neq 0$); (2) for each $h$ selected in this interval, Equations 2.35 and 2.38 are used to determine $\sigma$ and $L(\sigma)$; (3) for each pair $[\sigma, L(\sigma)]$ obtained in this way, Equation 2.44 is used to calculate an ASN value; and (4) the pairs $[\sigma, E_\sigma(n)]$ are plotted to form an ASN curve.

2.4.1.4 Modifications of Wald’s decision boundaries

As described in Section 2.4.1.1, the decision boundaries of a Wald’s (1947) SPRT may be represented as two parallel lines (i.e., $L_0$ and $L_1$). The decision process then involves comparing the cumulative sum of observations, $\sum_{i=1}^{n} x_i$, to these two decision lines to decide whether to stop or continue the testing at each stage of sampling. Wald (1947, p.158) proved that the probability was one that an SPRT plan would eventually terminate. However, for the situation when the true parameter, $\theta$, of the population is close to the mean of two class limits (i.e., $(\theta_0 + \theta_1)/2$), it is possible that the number of observations required by an SPRT plan will tend to be very large. This may result in a sample size which is even larger than that required by an equally reliable fixed sample size procedure.

To prevent this situation, Wald (1947) suggested modifying the decision boundaries of an SPRT plan to force a terminating decision to be made at some maximum number of observations, and called it the truncation of an SPRT plan. He also described a possible method for using the truncation in practice. The method was that an upper limit ($n_0$) of the number of observations was first preset. Then, if $n_0$ was reached and
no terminal decision could be made, a final decision might be made using the rule: (1) if $T_{n_0} \geq (a_{n_0} + r_{n_0})/2$, $H_0$ is rejected; and (2) if $T_{n_0} < (a_{n_0} + r_{n_0})/2$, $H_0$ is accepted, where $T_{n_0}$ is the cumulative observed values of the test statistic up to the sampling stage of $n_0$ (i.e., $T_{n_0} = \sum_{i=1}^{n_0} d_i^2$ for the SATP I); $a_{n_0} = h_1 + sn_0$ and $r_{n_0} = h_2 + sn_0$. Because $a_{n_0}$ and $r_{n_0}$ are the values of the acceptance ($L_0$) and rejection ($L_1$) lines at the $n_0^{th}$ stage of sampling, respectively, they are also called the acceptance and rejection values at $n_0$. This rule will be denoted as Wald’s (1947) rule of truncation for an SPRT plan. The modification will certainly change the resulting probabilities of the type I and II errors and the expected sample number (ASN) of an SPRT. However, Wald stated that if $n_0$ was chosen to be large enough (i.e., two or three times the calculated ASN value), the effect caused by the truncation would be small.

Originally, Wald (1947) derived the decision boundaries, and the general OC and ASN functions of SPRT plans based on the assumption that the random observations were taken singly in every stage of sequential sampling. Wald realized that this assumption was not practical in many situations, because taking random observations singly would increase the travel time between the sampling units, and degrade the advantage of an SPRT plan. Wald stated that in these situations, group selection (i.e., more than one observation is taken at each stage of sampling) might be used instead of single selection. The only possible effect of doing this would be an increase in the number of observations required by the test, and the resulting $\alpha$ and $\beta$ might be substantially smaller than their desired values. In other words, this modification may make an SPRT plan more conservative.

Wald (1947) did not provide any numerical evidence to confirm the real effects caused by these two modifications. Some studies in the forestry literature on the effects of these modifications will be presented in the next subsection.
2.4.1.5 Previous forestry applications and studies of Wald's SPRT plans

Numerous applications of sequential procedures can be found in the forestry as well as the agricultural literature. However, almost all of these applications used Wald's (1947) SPRT plans for pest surveys. Pieters (1978) tabulated 60 SPRT applications for agricultural and forest pest surveys. Fowler and Lynch (1987b) tabulated 70 SPRT plans (from 65 articles), 28 were for forest and 42 for agricultural pest surveys. The underlying distributions of the variables of interest in these applications included binomial, negative binomial, normal, and Poisson distributions. These applications of Wald's SPRT plans were designed to test the percent infestation \( q \) using a binomial or negative binomial distribution, or to test the mean density \( \mu \) (e.g., the average number of live larvae per plant) using a normal distribution with known variance or using a Poisson distribution.

The formulae for determining the decision boundaries and the OC and ASN functions of Wald's (1947) SPRT plans for forest pest surveys were tabulated by Walters (1955), and Fowler and Lynch (1987a). Talerico and Chapman (1970) developed a FORTRAN IV computer program (SEQUAN) to calculate the decision boundaries, and OC and ASN equations, and plot the decision boundaries for binomial, negative binomial, normal and Poisson distributions.

In applying Wald's (1947) SPRT procedure, it is important to choose appropriate values for the two class limits and the levels of error probabilities. In pest surveys, the class limits \( (i.e., \theta_0 \text{ and } \theta_1) \) represent economic thresholds or pest density levels. Choice of these values and the resulting gap between them \( (i.e., \text{the interval between } \theta_0 \text{ and } \theta_1) \) depends on the biology and behavior of the insect and its damage, the relationship between pest densities and resulting damage, economic constraints, time or labor constraints, and other criteria. The choice of the levels of error probabilities \( (i.e., \alpha \text{ and } \beta) \) is mainly based on the perceived seriousness of each error (Fowler and Lynch 1987a,
Wald's (1947) OC and ASN functions are helpful in assisting the users to design an optimal SPRT for a given problem. Since the problem of exceeding (overshooting) was ignored in their derivation, Wald's OC and ASN equations are only approximate. Wald (1947) stated that the errors inherent in his equations because of exceeding the decision boundary are small if $\alpha$ and $\beta$ are small (i.e., < 0.05), and $\theta_0$ and $\theta_1$ (class limits) are sufficiently close together. Because $\alpha$ and $\beta$ are usually 0.10, and the class limits are usually relatively wide for most sequential plans in forest pest management, several studies (i.e., Fowler 1978; Fowler 1983; Fowler and Lynch 1987a; and Lynch et al. 1990) have shown that the errors inherent in Wald's equations as a result of "overshooting" the decision boundaries of the SPRT can be large. The relative errors increased for the OC and ASN equations as the difference between the null ($\theta_0$) and alternative ($\theta_1$) test parameter values increased. Relative errors also increased for the ASN equation as the probabilities of type I and II errors increased. Wald's equations, in general, overestimated the true error probabilities and underestimated the true ASN. The practical consequences of these errors to the samplers of pest surveys are: (1) the actual error probabilities can be considerably smaller than the nominal error probabilities used to build the sampling plan; and (2) considerably more observations are taken in the field, on average, than necessary (Fowler 1983). Therefore, Wald's SPRT plans appear to be conservative testing procedures, because they actually give lower risk values (error probabilities) for a given amount of sampling. This could be considered a safety measure to assure reliable classification during decision making (Lynch et al. 1990).

In some situations, if the user considers Wald's (1947) equations not acceptable, two Monte Carlo alternative OC and ASN functions suggested by Fowler (1983) may be used. These are:
1. Nominal values of $\theta_0, \theta_1, \alpha$ and $\beta$ are first used to determine Wald's SPRT decision boundaries, and then based on these decision boundaries, the Monte Carlo technique is used to estimate the actual OC and ASN curves associated with the determined SPRT procedure. If the user considers the Monte Carlo OC and ASN curves acceptable, then these Monte Carlo OC and ASN curves, and the actual error probabilities, $\hat{\alpha}$ and $\hat{\beta}$, should be used instead of Wald's equations.

2. If the Monte Carlo OC and ASN obtained in (1) are not considered acceptable, the actual error probabilities, $\hat{\alpha}$ and $\hat{\beta}$, will then be used to calculate the new nominal level of error probabilities, $\alpha'$ and $\beta'$. These are $\alpha' = \alpha^2/\hat{\alpha}$ and $\beta' = \beta^2/\hat{\beta}$. The new nominal error probabilities, $\alpha'$ and $\beta'$, are then used to determine new decision boundaries. The new SPRT procedure based on $\alpha'$ and $\beta'$, will have the desired level of error probabilities (i.e., $\alpha$ and $\beta$).

For some entomology applications, the decision process of sequential sampling plans based on Wald's (1947) SPRT plans have been modified by taking more than one observation at each stage of the plan, not making a terminating decision until some minimum number of observations has been taken, or forcing a terminating decision to be made after some maximum number of observations. In practice, these modifications are made for biological and economic reasons after the sampling plan has been developed, and do not in any way affect the sequential decision lines. Thus, the modification is in applying the sampling plan and not in developing it. These deviations from the assumptions of Wald's SPRT will, of course, increase the inaccuracy of Wald's equations when used to describe the properties of the modified sampling plan. The errors caused by modifying Wald's assumptions were denoted as decision boundary modification errors by Fowler (1978). Fowler investigated the effects of these modifications on the accuracy of Wald's equations for a normal distribution using Monte Carlo simulation techniques. Based on
the simulation results, he concluded that taking more than one observation at each stage of the plan decreases the actual $\alpha$ and $\beta$ of the test and increases the actual ASN values. Truncating the decision process at some maximum number of observations increases the actual $\alpha$ and $\beta$ of the test and decreases the actual ASN values. Postponing a terminating decision until some minimum number of observations is taken decreases the actual $\alpha$ and $\beta$ of the test and increases the actual ASN values. He also concluded that the increase in the inaccuracy of Wald's equations due to these modifications depends on how far they deviate from the assumptions of Wald's SPRT and what combination of modifications is used operationally. The seriousness of the inaccuracies depends on the magnitude of the inaccuracy and on the costs of making a decision.

For forest management purposes, application of sequential sampling plans have been limited to surveying the adequacy of forest stocking. Smith and Ker (1958) applied a sequential sampling plan to assess the adequacy of the reproduction in the University of British Columbia's (B.C.) research forest at Haney. The sampling unit they defined was a cluster of four 1-milacre (0.1 acres or 0.0405 ha) quadrats along the sampling lines, and they used a test to confirm that the Poisson distribution was most appropriate for representing the mean stocking of four milacre quadrats of their data. Based on the results, they concluded that sequential sampling could result in considerable savings in the field work required in a reproduction survey. Fairweather (1985) gave a detailed discussion of applying a sequential sampling plan for assessing the adequacy of stocking. He indicated that the objective of stocking assessment was to classify forest lands into two broad classes, "stocked" and "non-stocked". Sequential sampling plans should be very efficient in this application, especially when the actual stocking of the forest was either very high or very low. However, the disadvantages in using sequential sampling were: (1) the sample size could be very large when the true stocking percentage was between the acceptable and unacceptable limits; (2) selection of an inappropriate spatial distribution
could result in an incorrect decision; and (3) theoretically, sample plots (either singly or in clusters) have to be located randomly on the site. This may reduce the efficiency of sequential sampling by increasing the travel time between plots. Fairweather conducted a Monte Carlo simulation study using stocking data obtained from five logging sites in northern Maine’s (United States) spruce-fir region. Sequential sampling plans based on Poisson and binomial distributions were tested. In his simulation, a maximum of 150 plots was arbitrarily set to stop the sampling for the “no decision” case. Fairweather also indicated that systematic selection, instead of random selection, appeared to be a valid method for assessing stocking adequacy.

2.4.2 Other sequential sampling plans used in forestry

Besides Wald’s (1947) SPRT plans, two other types of sequential sampling plans have been also used in forest pest management.

Iwao’s (1975) sequential sampling plans were designed to classify the mean number of forest pests per sampling unit (mean density, \( \mu \)) relative to a critical value of the mean density, \( \mu_c \). In testing the null hypothesis \( H_0 : \mu = \mu_c \) against all other alternatives (i.e., \( \mu < \mu_c \) or \( \mu > \mu_c \)), the basis used by Iwao for deriving the decision boundaries was different from Wald’s SPRT plans. Iwao’s sequential plans depended on a principal assumption, that the relationship between the variance (\( \sigma^2 \)) and the mean density of a population can be represented as:

\[
\sigma^2 = f(\mu) = (\lambda + 1)\mu + (\nu - 1)\mu^2
\]

(2.45)

where \( \lambda \) and \( \nu \) are two constants determined by Iwao’s (1968) linear relationship between Lloyd’s (1967) mean-crowding index (\( \bar{m} \)) and the mean density (i.e., \( \bar{m} = \lambda + \nu \mu \)), where \( \bar{m} \) is defined as \( \mu + [(\sigma^2/\mu) - 1] \). Based on this assumption, the acceptance and rejection boundaries (lines) of Iwao’s (1975) sequential plans are the upper (\( T_u \)) and lower (\( T_l \))
limits of the 100(1 – α) percent confidence interval for the total number \((T_n)\) of individuals of interest in \(n\) observations when \(H_0\) is true, where:

\[
T_u = n\mu_c + z_{1-\alpha/2}\sqrt{nf(\mu_c)}
\]

\[
T_l = n\mu_c - z_{1-\alpha/2}\sqrt{nf(\mu_c)}
\]

and \(z_{1-\alpha/2}\) is the 100(1 – α/2) percentile of a standard normal distribution.

By taking random observations from a population and recording the cumulative number of the individuals of interest in \(n\) observations \((T_n)\), the null hypothesis is tested after each stage of sampling using the following rule: (1) if \(T_n > T_u\), stop sampling and conclude \(\mu > \mu_c\); (2) if \(T_n < T_l\), stop sampling and conclude \(\mu < \mu_c\); and (3) if \(T_l < T_n < T_u\), then continue sampling until a terminal decision can be made, or a predetermined maximum number of samples \((n_{\text{max}})\) is reached. A terminal decision can then be made by comparing the observed mean density, \(\mu' = T_{n_{\text{max}}}/n_{\text{max}}\) to the critical mean density, \(\mu_c\).

Iwao (1975) defined \(n_{\text{max}}\) as:

\[
n_{\text{max}} = \left[\frac{z_{1-\alpha/2}\sqrt{f(\mu_c)}}{D_0\mu_c}\right]^2
\]

where \(D_0\) is the required level of precision specified as the ratio of the standard error of mean to the population mean density.

A discussion of the statistical aspects of Iwao's (1975) sequential plans was given by Nyrop and Simmons (1984). In general, the use of Iwao's plans relaxes the requirement of known distribution of the variable of interest and the difficulty of specifying the class limits, \(\theta_0\) and \(\theta_1\), for applying Wald's (1947) SPRT plans. However, its precision depends upon the estimated relationship of the variance and the mean density of a population \(i.e., \text{Equation 2.45}\). Also, it is only appropriate for testing the mean densities, for which a relationship between the variance and the mean has been defined. Some applications are found in Coggin and Dively (1982), and Mukerji et al. (1988).
Instead of classifying a population parameter (mean density) into two or three wide classes, Kuno's (1969) sequential plans were designed to estimate the mean density ($\mu$) of a population at a specified level of precision ($D_0$) with a minimum expected sample size. Kuno's plans also require the assumption that the variance ($\sigma^2$) could be represented as a known function of the mean density for given population (i.e., $\sigma^2 = f(\mu)$). Based on this assumption, Kuno expressed the standard error of sample mean as $S_\bar{x} = \left[ f(\bar{x})/n \right]^{1/2}$, and sampling precision as:

$$D = \frac{S_\bar{x}}{\bar{x}} = \frac{\sqrt{f(\bar{x})/n}}{\bar{x}} = \frac{n}{T_n} \sqrt{f(\frac{T_n}{n})/n} \tag{2.49}$$

where $T_n$ is the cumulative number of individuals observed in $n$ random observations; and $T_n/n$ (i.e., $\bar{x}$) is an estimate of the mean density, $\mu$. By replacing $D$ by the required level of precision, and rearranging the relationship with respect to $T_n$, Equation 2.49 becomes:

$$T_n = \sqrt{\frac{n}{D_0^2} f\left(\frac{T_n}{n}\right)} \tag{2.50}$$

Equation 2.50 represents the total number of individuals of interest required in $n$ random observations to give the desired precision $D_0$. Kuno (1969) called this equation the "stop line" since the sampler can stop once the observed cumulative number of individuals in any $n$ random observations ($T_n'$) equals or exceeds $T_n$. By using Kuno's (1969) sequential plans, an estimate of population mean density with the desired level of precision will be ($T_n/n$), and a $100(1 - \alpha)$ percent confidence interval for the mean density can also be obtained. Some applications of Kuno's (1969) sequential plans are found in Allen et al. (1972), Newton (1989), and Newton and LeMay (1992).

### 2.4.3 Development of sequential analysis theory in statistics

In its early stages, sequential analysis was heavily dominated by Wald's (1947) SPRT. In the statistics literature, bibliographies were given by Johnson (1961) and Jackson (1960).
Chapter 2. Literature Review

The scientific fields in which Wald's SPRT plans were most widely used are clinical research (Armitage 1975) and industrial quality control (Davies 1954; Burr 1953). In the former, patients usually enter a study serially and ethical considerations require that any unnecessary use of inferior treatment should be avoided. For the latter, the sampling inspection for batches of products is usually time-consuming or destructive. In these applications, sequential analysis is appealing.

Because of its practical values, Wald's (1947) original theory of SPRT has been extended to sequential testing of composite hypotheses, sequential estimation, sequential decision-making, and many other statistical applications. The theory and development of sequential analysis are well documented by Wald (1947), Wetherill and Glazebrook (1986) and Ghosh and Sen (1991). Some special testing procedures found in the literature were: (i) the sequential t-test for population means (Rushton, 1950); (ii) the truncated sequential t-test for a population mean (Fowler 1969); (iii) the sequential F-test for analysis of variance (Ghosh 1967); and (iv) sequentially testing the correlation coefficient for bivariate normal variables (Choi 1971).

2.5 Summary

To meet the first objective of this research, a literature search was carried out in forestry and other literature. Since applicability testing of forestry models is an extension of model validation, the available procedures for forestry model validation were reviewed, focusing on the statistical procedures for comparing estimated and observed values of a model.

In general, model validation is a process of building the confidence level for using a model. If a model is developed for estimation or prediction purposes, the predictive
ability (accuracy) of the model should be determined before it can be used with confidence. To determine the accuracy of a model (the closeness of the estimated values to their target values), statistical hypothesis testing or estimation methods may be used. Hypothesis testing methods are appropriate if the user of a model wants to determine whether the accuracy of the model meets an established requirement of accuracy. In forestry, Freese's (1960) procedure and its modifications were designed for this purpose, especially when the test model is a deterministic function (e.g., regression equation) and the average accuracy of the model is mainly of concern (the unconditional accuracy for a set of input variables that is randomly selected from all possible values). If the test model is a stochastic simulation model or conditional accuracy of the model is of concern (the accuracy of estimated values for a specified set of input variables), then Reynolds et al.'s (1981) procedures for validating computer simulation models are appropriate. Statistical estimation methods are appropriate when the user of a model wants to know the magnitude of the errors if the model is repeatedly used for estimation in a given population. Reynolds' (1984) estimation procedures and Reynolds and Chung's (1986) regression procedures are designed for this purpose.

However, almost all procedures found for forestry model validation are fixed sample size procedures; the decision about the accuracy of a model when applying these procedures can only be made when a data set with a predetermined, sufficiently large sample size is available. The purpose of applicability testing is not to provide accurate estimates for the error parameters of forestry estimation models, but to classify a model as acceptable or unacceptable for a given application. It may be possible that a decision to accept or reject the model can be made with a small sample size, especially when the actual accuracy of a model is far below or above the requirement of the user. A fixed sample size procedure is then not considered to be efficient for applicability testing of forestry models. This is particularly of concern when data collection is expensive, time-consuming
or destructive.

Sequential sampling plans have been widely used in forest pest management. In these applications, three types of sequential sampling plans were involved. Wald's (1947) and Iwao's (1975) sequential sampling plans were designed for hypothesis testing (classifying a population parameter in two or three wide classes). Kuno's (1969) sequential sampling was designed for estimating the population mean at a specified level of precision with minimum expected sample size. The applications of the sequential sampling plans to forest pest surveys were appealing, because these procedures required, on average, only 40 to 60% as many observations as an equally reliable fixed sample size procedure (Fowler and Lynch 1987a).

Although sequential approaches have been widely used in forest and agricultural pest management, and other scientific fields, few studies were found that addressed the problem of sequentially testing the predictive ability (accuracy) of any type of mathematical model. Also, no sequential testing plans that met the desired criteria outlined for the alternative procedures for applicability testing of forestry models, were found except for two initial studies in forestry (Marshall and LeMay 1990; Brack and Marshall 1990).
Chapter 3

Extension of Freese's Procedure to Sequential Accuracy Testing Plans

Information from the literature suggested that the variable sample size (sequential sampling) procedures could be superior to the fixed sample size procedures in terms of sampling cost savings. However, in order to apply sequential analysis approaches to model testing, some appropriate sequential testing plans, which have a better measure of the accuracy of model than a binomial variable, must be developed (Marshall and LeMay 1990).

Among the fixed sample size procedures found in the literature, Freese's (1960) procedure seems to be most appropriate for the purpose of applicability testing of a model, because this procedure was designed to determine whether the accuracy of an estimation technique meets a specified accuracy requirement of the user. Also, Freese's procedure has been widely used for evaluating the accuracy of forestry estimation or prediction models (e.g., Evert 1981; Ek and Monserud 1979), and it is quite natural to consider whether this procedure can be extended to a sequential testing plan.

The distributional assumptions and the derivation of this procedure have been well researched (i.e., Reynolds 1984; Gregoire and Reynolds 1988). The only requirements for the development and use of a sequential testing plan such as Wald's (1947) SPRT, are that the distribution of the variable of interest must be known, and that the two class limits and the probabilities of the type I and II errors must be specified. If the specified class limits represent the accuracy requirements for model acceptance and rejection, a sequential sampling plan may be used to classify the applicability of a model into one of
two pre-defined classes (i.e., acceptable or unacceptable).

This led to the decision to extend Freese's (1960) procedure of accuracy testing into a Wald's SPRT plan to meet the first objective of this research. The extension was made for each of Freese's three chi-square test statistics. The derivations are presented in this chapter.

3.1 Sequential Accuracy Testing Plan I

3.1.1 Modification for Freese's original formulation

In order to extend Freese's test I into an SPRT plan, the following modifications were made. First, instead of a predetermined value, the sample size required was allowed to be a random variable. Observations of the dependent variable (y) and independent variables (x) of the tested model would be sequentially taken at random from the application population.

Second, since Wald's SPRT is only applicable for testing a simple hypothesis\(^1\) (\(H_0\)) against another simple hypothesis (\(H_1\)), the composite hypotheses\(^2\) of Freese's test I (i.e., \(H_0 : \sigma^2 \leq \sigma_h^2\) vs. \(H_1 : \sigma^2 > \sigma_h^2\), where \(\sigma_h^2 = e^2/z^2_{1-\gamma/2}\)) were modified into two simple hypotheses. To solve this problem, instead of specifying a single allowable error, \(e\), two error limits, \(e_0\) and \(e_1\), were used. The choice of \(e_0\) and \(e_1\) is somewhat arbitrary, and \(e_0 < e < e_1\). The \(e_0\) can be interpreted as the upper limit of errors for classifying a model as acceptable, and \(e_1\) as the lower limit of errors for classifying a model as unacceptable.

Based on these two specified error limits and the desired probability level (\(\gamma\)) in the accuracy requirement (Equation 2.17), the modified test hypotheses of the extended test

---

\(^1\)A simple hypothesis is defined as a hypothesis under which the probability distribution is completely known.

\(^2\)A composite hypothesis is defined as a hypothesis under which the probability distribution is not completely known.
Chapter 3. Sequential Accuracy Testing Plans

I are:

\[
H_0: \sigma^2 = \sigma_0^2 \\
H_1: \sigma^2 = \sigma_1^2, \sigma_0^2 < \sigma_1^2
\]  

(3.51)

where \( \sigma^2 \) is the true variance of \( d \); \( \sigma_0^2 \) and \( \sigma_1^2 \) are the hypothesized variances of \( d \) for model acceptance and rejection, respectively; \( \sigma_0^2 = e_0^2/z_{1-\gamma/2}^2 \) and \( \sigma_1^2 = e_1^2/z_{1-\gamma/2}^2 \). The interpretations of the modified test hypotheses are as follows. Since Freese's test statistic was derived based on the assumption that the model error, \( d \), is normally distributed variable with a zero mean and variance \( \sigma^2 \), an accurate model will have:

\[
P(|d| \leq e) = 1 - \gamma
\]

and \( \sigma^2 = \sigma_h^2 \) (i.e., the true error variance is equal to the hypothesized variance), where \( \sigma_h^2 \) is determined from the specified values of \( e \) and \( \gamma \). Because \( \sigma_0^2 \) and \( \sigma_1^2 \) are chosen as \( \sigma_0^2 < \sigma^2 < \sigma_1^2 \) for the modified hypotheses, it implies that:

\[
P(|d| \leq e) > 1 - \gamma
\]

if the modified null hypothesis is true, and:

\[
P(|d| \leq e) < 1 - \gamma
\]

if the modified alternative hypothesis is true. Therefore, the modified hypotheses may be used to substitute for the original hypotheses of Freese's test I. Also, the null and alternative parameters represent the class limits for model acceptance and rejection, respectively. In a sequential analysis context, the equivalence between the modified test hypotheses and those of the original composite hypothesis was discussed by Wald (1947, p.125). He stated that for any value of the tested parameter, \( \sigma^2 \), we always could find two values (i.e., \( \sigma_0^2 \) and \( \sigma_1^2 \)), such that the classification of the true variance to be greater
than $\sigma_0^2$ was considered an error of practical importance whenever the true variance was less than $\sigma_0^2$. Similarly, the classification of the true variance to be less than $\sigma_0^2$ was regarded as an error of practical importance whenever the true variance was greater than $\sigma_1^2$. For a true variance value between $\sigma_0^2$ and $\sigma_1^2$, there was no particular concern as to which decision was made.

Finally, in order to derive the decision boundaries of the extended test I, Wald’s (1947) general stopping rule, $B < R_n < A$, for an SPRT plan was used. This requires that the probability distribution of $d$ is known, and that the probability levels of the type I and II errors, $\alpha$ and $\beta$, must be preset for the test. The distributional assumptions of Freese’s test I, which were outlined by Reynolds (1984) and Gregoire and Reynolds (1988) were retained. That is, the $n$ differences between the estimated and actual values from a sequence of model estimation, $d_1, d_2, \ldots d_n$, are independent and identically distributed (iid) normal variables with a zero mean and a variance, $\sigma^2$. Also, the unconditional distribution of model errors, $d$, was assumed to be used in testing.

### 3.1.2 Sequential decision boundaries

Using the distributional assumptions stated for $d$, the decision boundaries of the extended test I were derived as follows. The joint probability density function of the observed errors, $d_i$ for $i = 1, \ldots, n$, under the null hypothesis $H_0$ (Equation 3.51) is:

$$g_0(d_1, \ldots, d_n) = \prod_{i=1}^{n} f_0(d_i) = \frac{1}{(2\pi)^{n/2} \sigma_0^n} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} d_i^2\right)$$

(3.52)

where $\sigma_0^2 = e_0^2/z_{1-\gamma/2}^2$. Similarly, the joint probability density function of $d$ under the modified alternative hypothesis $H_1$ (Equation 3.51) is:

$$g_1(d_1, \ldots, d_n) = \prod_{i=1}^{n} f_1(d_i) = \frac{1}{(2\pi)^{n/2} \sigma_1^n} \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{n} d_i^2\right),$$

(3.53)

where $\sigma_1^2 = e_1^2/z_{1-\gamma/2}^2$. With these known density functions, Wald’s (1947) general decision rules for an SPRT plan described in section 2.4.1.1 could be applied to derive the
Chapter 3. Sequential Accuracy Testing Plans

59

decision boundaries of the extended test I by first determining the sample likelihood ratio 
\( R_n \) of the alternative to the null hypotheses:

\[
R_n = \frac{g_1(d_1, \ldots, d_n)}{g_0(d_1, \ldots, d_n)} = \prod_{i=1}^{n} \frac{f_1(d_i)}{f_0(d_i)}
\]

\[
R_n = \frac{\frac{1}{(2\pi)^{n/2}(\sigma_1^2)^{n/2}} \exp \left( -\frac{1}{2\sigma_1^2} \sum_{i=1}^{n} d_i^2 \right)}{\frac{1}{(2\pi)^{n/2}(\sigma_0^2)^{n/2}} \exp \left( -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} d_i^2 \right)}
\]

\[
R_n = \left[ \frac{\sigma_0^2}{\sigma_1^2} \right]^{n/2} \exp \left[ \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^{n} d_i^2 \right]
\]

Wald’s rule, \( B \leq R_n \leq A \), was then applied to divide the \( n \)-dimensional sample space into three mutually exclusive subspaces, where \( A \equiv (1 - \beta)/\alpha \) and \( B \equiv \beta/(1 - \alpha) \); and \( \alpha \) and \( \beta \) are the specified probability levels of the type I and II errors for the test, respectively. This was obtained by replacing \( R_n \) in Wald’s rule using the above result.

\[
B \leq \left[ \frac{\sigma_0^2}{\sigma_1^2} \right]^{n/2} \exp \left[ \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^{n} d_i^2 \right] \leq A
\]

or

\[
B \left[ \frac{\sigma_1^2}{\sigma_0^2} \right]^{n/2} \leq \exp \left[ \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^{n} d_i^2 \right] \leq A \left[ \frac{\sigma_1^2}{\sigma_0^2} \right]^{n/2}
\]

This expression was simplified by taking the natural logarithms (\( \ln \)) for each term, and replacing \( \sigma_0^2 \) by \( (e_0^2/z_{1-\gamma/2}^2) \) and \( \sigma_1^2 \) by \( (e_1^2/z_{1-\gamma/2}^2) \).

\[
\ln(B) + \frac{n}{2} \ln \left[ \frac{e_1^2}{e_0^2} \right] \leq \frac{1}{2} \frac{z_{1-\gamma/2}^2}{z_{1-\gamma/2}^2(1 - \frac{1}{e_0^2} - \frac{1}{e_1^2})} \sum_{i=1}^{n} d_i^2 \leq \ln(A) + \frac{n}{2} \ln \left[ \frac{e_1^2}{e_0^2} \right]
\]

This result was further simplified to:

\[
\frac{2 \ln(B) + n \ln(e_1^2/e_0^2)}{z_{1-\gamma/2}^2(1 - \frac{1}{e_0^2} - \frac{1}{e_1^2})} \leq \sum_{i=1}^{n} d_i^2 \leq \frac{2 \ln(A) + n \ln(e_1^2/e_0^2)}{z_{1-\gamma/2}^2(1 - \frac{1}{e_0^2} - \frac{1}{e_1^2})}
\] (3.54)

Equation 3.54 gives the derived decision boundaries of the extended test I, and it can be represented as two parallel decision lines by letting:

\[
h_1 = \frac{2 \ln(\frac{\beta}{1-\alpha})}{z_{1-\gamma/2}^2(1 - \frac{1}{e_0^2} - \frac{1}{e_1^2})}
\] (3.55)
Chapter 3. Sequential Accuracy Testing Plans

\[ h_2 = \frac{2 \ln(\frac{1-\beta}{\alpha})}{[\chi^2_{1-\gamma/2}(\frac{1}{e_0} - \frac{1}{e_1})]} \]  
\[ s = \frac{\ln(e/e)}{[\chi^2_{1-\gamma/2}(\frac{1}{e_0} - \frac{1}{e_1})]} \]  
\[ T_1 = \sum_{i=1}^{n} d_i^2 \]  

(3.56)  
(3.57)  
(3.58)

where \( T_1 \) is the cumulative squared errors of model estimation, the test statistic of the extended test I; \( h_1 \) and \( h_2 \) are the intercepts; and \( s \) is the common slope of the decision lines. Equation 3.54 becomes:

\[ h_1 + sn \leq T_1 \leq h_2 + sn \]

At each stage of sequential sampling, one of the following decisions would be made:

1. \( T_1 \geq h_2 + sn \), stop sampling and conclude that the model is not acceptable (i.e., reject \( H_0 \));

2. \( T_1 \leq h_1 + sn \), stop sampling and conclude that the model is acceptable (i.e., accept \( H_1 \)); or

3. \( h_1 + sn < T_1 < h_2 + sn \), continue sampling. No terminal decision can be made at this stage.

Cases 1 and 2 are denoted as the terminal decisions of the extended test I; and \( L_0 = h_1 + sn \) and \( L_1 = h_2 + sn \) are called the acceptance and rejection lines, respectively. This sequential testing plan was designed to perform the same function as Freese’s chi-square test I. It will be denoted as the sequential accuracy testing plan I throughout this dissertation, and labelled as SATP I. Based on the theory of Wald’s (1947) SPRT outlined in Chapter 2, such an extension of Freese’s test I appears to be reasonable. However, some problems in applying this extended procedure exist.
The key for this extension is the normality assumption of the model errors, \( d \). Gregoire and Reynolds (1988) showed that Freese’s original procedure was sensitive to non-normality. The extended procedure is not robust to any departure from normality since a smaller sample size is expected using a sequential sampling procedure. Therefore, in applying this extended test, care must be exercised to insure that the distribution of \( d \) is normal. To do this, a post-testing procedure of normality is proposed and will be presented in Chapter 5 of this dissertation.

In order to classify a model as either acceptable or unacceptable, two class limits for the errors (\( e_0 \) and \( e_1 \)) must be specified by the potential users. These class limits, of course, must be realistic and based on a fundamental knowledge of model evaluation and timber management purposes. These two specified values and their associated probability level, \( \gamma \), are required to determine the hypothesized variances for model acceptance and rejection, \( \sigma^2_0 \) and \( \sigma^2_1 \). If the mean squared error (\( MSE \)) from the model estimation is known for the original population, it may be possible to specify the hypothesized variances directly, instead of calculating them from the specified values of \( e_0 \), \( e_1 \) and \( \gamma \). Also, the smaller the difference between the class limits, the wider the two decision boundaries will be. The effect of the interval between two class limits will be addressed in Chapter 4.

### 3.2 Sequential Accuracy Testing Plan II

#### 3.2.1 Modifications of Freese’s original formulation

As indicated in Chapter 2, Freese’s test II is designed for testing the relative error \( (d/y) \) instead of \( d \). The basic assumptions and idea for accuracy testing are exactly the same as those used for Freese’s test I. According to Reynolds (1984), the derivation of Freese’s test II assumes that \( d_1/y_1, d_2/y_2, \ldots, d_n/y_n \) are \( iid \) normal variables with a zero mean and a variance \( \sigma^2_{d/y} \). The test hypotheses of Freese’s original test are: \( H_0 : \sigma^2_{d/y} \leq \sigma^2_p \) vs.
Chapter 3. Sequential Accuracy Testing Plans

62

\( H_1 : \sigma_{d/y}^2 > \sigma_p^2 \), where \( \sigma_p^2 = p^2/(100z_{1-\gamma/2})^2 \).

Similar to SATP I, to extend Freese’s test II into a Wald’s SPRT, the test hypotheses of the original test must be modified into two simple hypotheses as:

\[
H_0 : \sigma_{d/y}^2 = \sigma_{p0}^2 \text{ vs. } \ H_1 : \sigma_{d/y}^2 = \sigma_{p1}^2 \tag{3.59}
\]

where \( \sigma_{p0}^2 = p_0^2/(100z_{1-\gamma/2})^2 \), \( \sigma_{p1}^2 = p_1^2/(100z_{1-\gamma/2})^2 \); \( p_0 \) and \( p_1 \) are the percent error limits for model acceptance and rejection, respectively; \( p_0 < p < p_1 \); and \( p \) and \( \gamma \) are the same as in Freese’s test II. Also, both the probability levels of the type I and II errors of the extended test II, \( \alpha \) and \( \beta \), must be specified. The interpretations of these modified hypotheses are similar to those of SATP I. That is, if the null hypothesis is true, it implies:

\[
P(100|d/y| \leq p) > 1 - \gamma
\]

Similarly, if the alternative hypothesis is true, it implies:

\[
P(100|d/y| \leq p) < 1 - \gamma.
\]

As with SATP I, the modified hypotheses were then used to substitute for the original hypotheses of Freese’s (1960) test II.

3.2.2 Sequential decision boundaries

To derive decision boundaries for the extended test II, the methods are similar to those for SATP I. First, based on the assumption, \( d/y \sim N(0, \sigma_{d/y}^2) \), the joint probability density functions of \( d_1/y_1, d_2/y_2, \ldots, d_n/y_n \) under the modified null and alternative hypotheses (Equation 3.59) are:

\[
g_0(d_1/y_1, \ldots, d_n/y_n) = \prod_{i=1}^{n} f_0(d_i/y_i) = \frac{1}{(2\pi)^{n/2}\sigma_{p0}^n} \exp \left[ -\frac{1}{2\sigma_{p0}^2} \sum_{i=1}^{n} (d_i/y_i)^2 \right] \tag{3.60}
\]
\[ g_1(d_1/y_1, \ldots, d_n/y_n) = \prod_{i=1}^{n} f_1(d_i/y_i) = \frac{1}{(2\pi)^{n/2} \sigma_{p1}^{n}} \exp\left[ -\frac{1}{2 \sigma_{p1}^{2}} \sum_{i=1}^{n} (d_i/y_i)^2 \right] \] (3.61)

The sample likelihood ratio \((R_n)\) of the alternative to null hypothesis is:

\[ R_n = \frac{g_1(d_1/y_1, \ldots, d_n/y_n)}{g_0(d_1/y_1, \ldots, d_n/y_n)} = \prod_{i=1}^{n} \frac{f_1(d_i/y_i)}{f_0(d_i/y_i)} \]

\[ R_n = \frac{1}{(2\pi)^{n/2} (\sigma_{p0}^{2})^{n/2}} \exp\left[ -\frac{1}{2 \sigma_{p0}^{2}} \sum_{i=1}^{n} (d_i/y_i)^2 \right] \]

\[ R_n = \left( \frac{\sigma_{p0}^{2}}{\sigma_{p1}^{2}} \right)^{n/2} \exp \left[ \frac{1}{2} \left( \frac{1}{\sigma_{p0}^{2}} - \frac{1}{\sigma_{p1}^{2}} \right) \sum_{i=1}^{n} (d_i/y_i)^2 \right]. \]

Relating the variance limits, \(\sigma_0^2\) and \(\sigma_1^2\), to the user-specified values, \(p_0\) and \(p_1\), (i.e., \(\sigma_{p0}^{2} = \left[ p_0/(100z_{1-\gamma/2}) \right]^2\), and \(\sigma_{p1}^{2} = \left[ p_1/(100z_{1-\gamma/2}) \right]^2\)), and applying Wald's rule, \(B \leq R_n \leq A\) for sample space division, the following expression is obtained.

\[ B \leq \left[ \frac{p_0^2}{p_1^2} \right]^{n/2} \exp \left[ \frac{1}{2} \times 100^2 z_{1-\gamma/2}^2 \left( \frac{1}{p_0^2} - \frac{1}{p_1^2} \right) \sum_{i=1}^{n} (d_i/y_i)^2 \right] \leq A \]

Taking natural logarithms and doing the necessary algebraic simplification, the expression above becomes:

\[ \frac{2 \ln(B) + n \ln(p_0^2/p_0^2)}{100^2 \times z_{1-\gamma/2}^2 (1/p_0^2 - 1/p_1^2)} \leq \sum_{i=1}^{n} (d_i/y_i)^2 \leq \frac{2 \ln(A) + n \ln(p_0^2/p_0^2)}{100^2 \times z_{1-\gamma/2}^2 (1/p_0^2 - 1/p_1^2)} \]

Similar to SATP I, this resulting expression can be represented as two parallel lines by letting:

\[ h_1 = \frac{2 \ln \left( \frac{1-\alpha}{1-\alpha} \right)}{\left( 100z_{1-\gamma/2}^2 \left( \frac{1}{p_0^2} - \frac{1}{p_1^2} \right) \right)} \] (3.62)

\[ h_2 = \frac{2 \ln \left( \frac{1-\beta}{\alpha} \right)}{\left( 100z_{1-\gamma/2}^2 \left( \frac{1}{p_0^2} - \frac{1}{p_1^2} \right) \right)} \] (3.63)

\[ s = \frac{\ln \left( \frac{\beta^2}{\alpha^2} \right)}{\left( 100z_{1-\gamma/2}^2 \left( \frac{1}{p_0^2} - \frac{1}{p_1^2} \right) \right)} \] (3.64)
\[ T_2 = \sum_{i=1}^{n} (d_i/y_i)^2 \]  

(3.65)

where \( h_1, h_2 \) are the two intercepts for the upper and lower decision boundaries; \( s \) is the common slope of the decision boundaries; and \( T_2 \) is the cumulative squared relative errors and it is the test statistic of the extended test II. The simplified decision boundary is then:

\[ h_1 + sn \leq T_2 \leq h_2 + sn \]

With \( h_1, h_2 \) and \( s \) defined as above, the decision rules of the extended test II at each stage of sampling are exactly the same as those outlined for SATP I after replacing the test statistic with \( T_2 \).

This testing procedure is designed to perform the function of Freese's test II. It will be denoted as the sequential accuracy testing plan II, or SATP II. Similar to SATP I, SATP II requires the assumptions that the relative errors, \( d/y \), have a normal distribution with zero mean, and that random observations are selected singly. Therefore, in applying this test, care must be given to confirm the normality of \( d/y \). The appropriate post-test procedures are the same as those for the SATP I, and are explained in Chapter 5.

### 3.3 Sequential Accuracy Testing Plan III

SATPs I and II require a normality assumption for \( d \) or \( d/y \) with a zero mean. A zero mean for the error \( d \), or relative error \( d/y \), implies that the model estimation is unbiased. However, it may not be practical to assume the mean of \( d \) or \( d/y \) to be zero in many applications. The sum of squared errors is used as the test statistic in each of these two SATP procedures, and this is a measure of the total error of model estimation. Therefore, in applying these two sequential testing plans, an inaccurate model will be rejected regardless of the source of inaccuracy (larger bias, lower precision, or both).
Chapter 3. Sequential Accuracy Testing Plans

For the situation when the bias can be assumed to be the same for all estimated values, Freese (1960) derived test III to give an approximate testing of accuracy after eliminating bias. Gregoire and Reynolds (1988, p.309) indicated that this approximate test after eliminating the bias for the absolute error $d$ cannot be used for the relative error $d/y$. They stated that there is no apparent way to correct for bias that preserves normality when the error is represented as a relative unit.

Similar to SATP I, effort was made to extend Freese's (1960) test III into a Wald's (1947) SPRT plan. However, compared to the two previous extensions, the difficulty with this extension is that the population mean $\mu_d$ of $d$ (the non-tested parameter) is unknown when bias is present. As indicated previously, Wald's (1947) SPRT is only applicable when both the null and alternative hypothesis are simple hypotheses (all parameters in the null and alternative distributions are known). When $\mu_d$ is unknown, the modified hypotheses of the SATP I (Equation 3.51) are no longer simple, but composite. Wald's SPRT plan is then not applicable.

In order to formulate Freese's (1960) test III into a Wald's (1947) SPRT, some suggestions by Wald were adopted. These were that when testing the standard deviation of a normal variable ($\sigma$) in the case when the population mean ($\mu$) is unknown, an approximate SPRT plan may be obtained using the following modifications:

1. Replace the test statistic $\sum_{i=1}^{n}(x_i - \mu)^2$ by $\sum_{i=1}^{n}(x_i - \bar{x})^2$, where $x_i$ is the observed value of the variable, and $\bar{x} = \sum_{i=1}^{n}x_i/n$; and

2. Change the number of observations in the decision rule to $n - 1$ instead of $n$. That is, if the mean is unknown, the decision boundaries at the $n^{th}$ sampling stage are equal to the decision boundaries corresponding to the $(n-1)^{th}$ sampling stage when the mean is known.

Since the essence of Freese's (1960) procedures is to test whether the variance of a normal
variable \( d \) exceeds a given value \( \sigma^2_h \), Wald’s suggestions are applicable to extend Freese’s test III into a Wald’s SPRT plan. For testing the same test hypotheses (Equation 3.51) as SATP I, the test statistic \( (T_3) \) for the extended test III is:

\[
T_3 = \sum_{i=1}^{n} (d_i - \bar{d})^2
\]

(3.66)

At each stage of sequential sampling, one of three decisions will be made:

1. If \( T_3 \geq h_2 + s(n - 1) \), stop sampling and reject the model;
2. If \( T_3 \leq h_1 + s(n - 1) \), stop sampling and accept the model; or
3. If \( h_1 + s(n - 1) < T_3 < h_2 + s(n - 1) \), continue sampling. No terminal decision can be made at this stage.

where the values of \( h_1, h_2, \) and \( s \) are defined as for SATP I (Equations 3.55, 3.56 and 3.57). This sequential testing plan is designed to perform the same function as Freese’s test III, and it will be denoted as sequential accuracy testing plan III, or SATP III.

It should be indicated that SATP III was not obtained directly through computing the sample likelihood ratio \( R_n \) for the situation when the mean is unknown, but was a modification of the sequential testing plan when the mean of \( d \) is known (SATP I). Therefore, it is only an approximate SPRT plan. The validity of SATP III must be examined through simulation experiments, which will be addressed in Chapter 4.

3.4 Approximate OC and ASN Equations for the SATP Procedure

In previous sections, three Freese’s (1960) chi-square test statistics were formulated as Wald’s (1947) SPRT plans for determining the applicability of forestry estimation models. The rationale of the extensions was that Freese’s idea behind accuracy testing (i.e., to confirm the probability statement given by Equation 2.17) can be translated into testing
the hypotheses $H_0: \sigma^2 \leq \sigma_h^2$ vs. $H_1: \sigma^2 > \sigma_h^2$ under the normality assumption and a zero mean for model error $d$ or $d/y$.

As sequential testing plans, SATP procedures should have the advantage that, on average, the sample sizes required to reach a decision about the applicability of a model should be smaller than those required by an equally reliable fixed sample size procedure. However, since the sample size becomes a random variable, the sampling cost is unknown in advance of sampling. This may produce administrative difficulties in practical application. To solve this problem, Wald's approximate OC and ASN equations may be applied to describe the probability of accepting the null hypothesis (i.e., OC value) and the expected sample size (i.e., ASN) required for a given SATP procedure. These equations will assist potential users of the SATP procedures in designing an appropriate testing plan for a given problem.

Because the essence of Freese's (1960) test I and II is to determine whether the variance ($\sigma^2$) of a normal variable with a zero mean (i.e., model errors, $d$ or $d/y$) exceeds a specified value ($\sigma_h^2$), and the SATPs I and II are formulated to perform the same functions as the original procedure, this suggests that it may be possible to extend the OC and ASN equations of Wald's (1947) SPRT plan for testing the standard deviation of a normal variable for use with SATPs I and II. Based on this analysis, the appropriate OC equations of SATPs I and II are suggested based on the equations given in Section 2.4.1. These are:

\[
L(\sigma^2) = \frac{(1-\beta)^h - 1}{(1-\alpha)^h - (\frac{\beta}{1-\alpha})^h}
\]

(3.67)

\[
\sigma^2 = \frac{(\frac{\sigma_0}{\sigma_1})^{2h} - 1}{\frac{h}{\sigma_1^2} - \frac{h}{\sigma_0^2}}
\]

(3.68)

where $L(\sigma^2)$ is the OC value for given $\sigma^2$ (i.e., a possible value of the true variance of model error, $d$ or $d/y$); and $h$ can be any non-zero value ($h$ from -4.0 to 4.0 with an
interval of 0.5 usually may be used for constructing an OC curve). Equations 3.67 and 3.68 can be used to determine the approximate OC curve for any SATP I and II with a specified set of values, \( \sigma_0^2, \sigma_1^2, \alpha \) and \( \beta \). The procedures used are as follows. For any given value of \( h \), \( \sigma^2 \) and \( L(\sigma^2) \) are calculated using these equations. Using \( \sigma^2 \) as the horizontal axis and \( L(\sigma^2) \) as the vertical axis, the pair \( (\sigma^2, L(\sigma^2)) \) is plotted. When the pairs, \( (\sigma^2, L(\sigma^2)) \), are calculated for a sufficiently large number of values of \( h \) (e.g., from -4.0 to 4.0 with an interval 0.5), the OC curve for a given SATP procedure will be obtained as in Figure 3.1. The OC curve constructed in this way has the property that the OC values \( (L(\sigma^2)) \) are \( L(0) = 1 \), \( L(\sigma_0^2) = 1 - \alpha \), \( L(\sigma_1^2) = \beta \), and \( L(\infty) = 0 \). These properties may be used for examining the accuracy of an approximate OC equation through simulation experiments.

The approximated ASN equations suggested for SATP I and II are:

\[
E_{\sigma^2}(n) = \frac{L(\sigma^2) \left[ \ln(\frac{\beta}{\alpha - \beta}) - \ln(\frac{1-\beta}{1-\alpha}) \right] + \ln(\frac{1-\beta}{\alpha})}{\frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sigma^2 + \ln(\frac{\sigma_0}{\sigma_1})}
\]

or:

\[
E_{\sigma^2}(n) = \frac{L(\sigma^2)(h_0 - h_1) + h_1}{\sigma^2 - s}
\]

where \( E_{\sigma^2}(n) \) is the ASN value for a given \( \sigma^2 \); and \( h_0, h_1 \) and \( s \) are the determined intercepts and slope of the decision boundaries of SATP procedure.

To construct an ASN curve for a specified set of \( \sigma_0^2, \sigma_1^2, \alpha \) and \( \beta \), procedures are similar to those used to construct an OC curve. For a given \( h \) value, the \((\sigma^2, E_{\sigma^2}(n))\) pair can be calculated using Equation 3.68, and Equation 3.69 or 3.70. Repeating the calculation for a sufficiently large number of \( h \), and plotting all calculated pairs as in Figure 3.2, an approximate ASN curve will be obtained.

It should be noted that these OC and ASN equations were originally derived by ignoring any possible overshooting of the decision boundaries of an SPRT plan. The accuracy of these suggested equations in approximating the actual unknown OC and
Figure 3.1: Example of an approximate Operating Characteristic (OC) curve of SATP I for testing $H_0 : \sigma^2 = 1.50$ vs. $H_1 : \sigma^2 = 2.25$ with $\alpha = 0.05$ and $\beta = 0.10$. 
Figure 3.2: Example of an approximate Average Sample Number (ASN) curve of SATP I for testing $H_0 : \sigma^2 = 1.50$ vs. $H_1 : \sigma^2 = 2.25$ with $\alpha = 0.05$ and $\beta = 0.10$.

ASN functions of the SATP procedures must be examined through simulation. Also, since SATP III is a modification of SATP I for the situation that the mean of $d$ is not zero, the OC and ASN equations suggested in this section may be appropriate for determining the OC and ASN values for SATP III. However, this needs to be confirmed through simulation.
Chapter 4

Monte Carlo Simulations for Examining Sequential Accuracy Testing Plans

In Chapter 3, Freese's (1960) procedure of accuracy testing was extended to Wald's (1947) SPRT plans to meet the first objective of this research. It was indicated that the rationale of such an extension is that the accuracy requirement (Equation 2.17 or 2.22) can be translated into a variance bound (Equation 2.20 or 2.25) when the errors of a model are iid normal variables with a zero mean and a constant variance. However, the reliability and behavior of the developed SATP procedures must be confirmed through sampling experiments. The reliability of the SATP procedures means that the extended procedures can reach a correct decision of model acceptance or rejection with desired probability levels for the type I and II errors ($\alpha$ and $\beta$).

This chapter is designed to meet the second objective of the research, that is to validate the SATP procedures by constructing model errors of tree volume estimates using Monte Carlo techniques. The problem addressed by the simulation studies in this chapter was the applicability testing of tree volume models. The objectives of the simulation studies were:

1. To examine the reliability of the SATP procedures in terms of the probability of reaching a correct decision under different assumptions for the errors of tree volume models;

2. To investigate the expected sample size of the SATP procedures required to make a terminal decision for different specified parameters (i.e., $\sigma_0^2$, $\sigma_1^2$, $\alpha$, and $\beta$);
3. To examine the robustness of the SATP procedures to departure from the underlying assumptions (i.e., the mean of the model errors is not zero, or the variance of errors is not constant);

4. To examine the accuracy of the approximate OC and ASN equations suggested for the SATP procedures; and

5. To study the effect of modifications of Wald’s assumptions for SPRT on the performance of SATP procedures. The modifications considered were: (1) at each stage of sampling, instead of selecting one random observation only, a group of random observations was taken; and (2) a maximum sample size ($n_0$) was set to terminate the testing process for the situation when no terminal decision can be reached.

In general, by using Monte Carlo techniques, some insights into how the SATP procedures behaved under the simulated application conditions of model testing were provided. Since the study of Gregoire and Reynolds (1988) confirmed that Freese’s original procedure is not robust to the departure from normality, the simulations were limited to normally distributed errors, or mixtures of normally distributed errors. The normal generator proposed by Forsythe et al. (1977) was used, and the unconditional distributions of the errors ($d$) of tree volume models were simulated. The mean ($\mu$) and the variance ($\sigma^2$) of the normal distributional errors generated represent the bias and precision of model estimates, respectively. Three combinations for the mean and variance of the generated errors were considered in this chapter. These were:

1. Zero mean and constant variance. These normally distributed errors were used to simulate the errors of a tree volume model when the model estimates are unbiased, and constant error variance is present across all levels of the model independent variables;
2. A non-zero mean and constant variance. These generated errors were used to simulate the errors of a volume model when the model estimates are biased, but the variance of errors is constant; and

3. Zero mean and various variances. A mixture of five or ten single normal distribution generators with zero means and different variances was used to simulate heterogeneous variances of model errors, which are commonly found with Spurr’s (1952), or Schumacher and Hall’s (1933) volume models.

Because the underlying distributional assumptions for SATP I and II are exactly the same, only SATP I and III were examined in this chapter.

The chapter is divided into seven sections. The first section presents the estimates of the error parameters from three sets of sectional tree data and three selected tree volume models. The sections following that present the simulation methods and results of examining the developed SATP procedures using the different generated error populations. The discussions and conclusion are found in the last section. It should be noted that although the problem of applicability testing of tree volume models was addressed in the simulation studies, the methods and results obtained in this chapter are applicable to other forestry models with similar error parameters.

4.1 Estimates of Error Parameters of Tree Volume Estimations

In order to address the problem of applicability testing of tree volume models, the magnitude of the errors of tree volume estimation should be known. Realistic simulations can then be designed based on the known parameters of errors. For this purpose, three sets of tree sectional data were obtained. These data were collected by Inventory Branch, B.C. Ministry of Forests across the province in the 1970’s. Each data set included one
species: Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), lodgepole pine (*Pinus contorta* Dougl.), and aspen (*Populus tremuloides* Michx.). The available variables in these data sets included: Forest Inventory Zone (FIZ), site quality (denoted as good, medium, and poor), total tree age (age), total height (height), diameter outside bark at breast height (1.3 metres above the ground) (dbh), total stem volume (volume), and the sectional measurements (i.e., diameters inside bark at intervals along the stem, cumulative height, and sectional volumes). After deleting the trees that were forked or had broken tops, descriptive statistics for the remaining trees in each data set were computed for age, dbh, height and volume (Table 4.1).

Using these data, and Spurr's (1952) (Equation 2.4), Schumacher and Hall's (1933) (Equation 2.2) and the logarithm of Schumacher and Hall's (1933) (Equation 2.3) tree volume models, estimates of the error parameters (i.e., the mean and variance) of tree volume models were obtained. For each data set, each of the three volume models was fitted using linear least-squares or nonlinear least-squares methods. These estimated models were assumed to be the population models for each data set. Using these fitted models, the error parameters of volume estimation were calculated by applying the population models to a local area (subpopulation). A subset was selected from each data set based on the variable, FIZ (Forestry Inventory Zone), and the population models were used to estimate the volume ($V_i$) for each tree in the subset. The errors of volume estimation were then calculated as the differences between the actual and estimated volumes ($d_i = V_i - \hat{V}_i$). Errors were also calculated by applying a population model to a different population (different species). Each population model was used to estimate the tree volume for the other two species. For example, the Douglas-fir model was used to estimate tree volume for lodgepole pine and aspen data. The average difference, $\bar{d}$, the average absolute difference, $|\bar{d}|$, and the mean squared difference ($MSD$) between the actual and estimated volumes were computed for each application. The $\bar{d}$'s found in
Table 4.1: Descriptive statistics of B. C. tree sectional data used for determining the error parameters of volume estimations.

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of Trees</th>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Douglas-fir</td>
<td>990</td>
<td>Age (yrs.)</td>
<td>204.1</td>
<td>132.9</td>
<td>29.0</td>
<td>578.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dbh (cm)</td>
<td>55.3</td>
<td>36.8</td>
<td>5.4</td>
<td>216.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Height (m)</td>
<td>33.2</td>
<td>14.4</td>
<td>6.3</td>
<td>76.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Volume (m^3)</td>
<td>4.9210</td>
<td>8.652</td>
<td>0.0074</td>
<td>77.3126</td>
</tr>
<tr>
<td>lodgepole pine</td>
<td>2880</td>
<td>Age (yrs.)</td>
<td>111.7</td>
<td>46.3</td>
<td>15.0</td>
<td>302.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dbh (cm)</td>
<td>24.4</td>
<td>9.5</td>
<td>3.8</td>
<td>60.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Height (m)</td>
<td>21.7</td>
<td>6.8</td>
<td>3.1</td>
<td>39.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Volume (m^3)</td>
<td>0.6022</td>
<td>0.5346</td>
<td>0.0024</td>
<td>3.38786</td>
</tr>
<tr>
<td>aspen</td>
<td>1077</td>
<td>Age (yrs.)</td>
<td>95.4</td>
<td>26.7</td>
<td>27.0</td>
<td>192.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dbh (cm)</td>
<td>26.9</td>
<td>9.1</td>
<td>9.4</td>
<td>72.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Height (m)</td>
<td>22.3</td>
<td>4.6</td>
<td>10.3</td>
<td>34.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Volume (m^3)</td>
<td>0.6223</td>
<td>0.5405</td>
<td>0.0404</td>
<td>3.6916</td>
</tr>
</tbody>
</table>
this way ranged from -2.0630 to 0.1505 m\(^3\) (Table 4.2), and MSD's found ranged from 0.0049 to 19.9330 for Spurr's (1955) and Schumacher and Hall's (1933) volume models. The \(\bar{d}\)'s ranged from -0.2215 to 0.1646, and the MSD's ranged from 0.00536 to 0.0247 for the logarithm of Schumacher and Hall's (1933) volume model. The \(\bar{d}\) and MSD were used as the estimates of the bias and variance of the errors in volume estimates for B.C. trees. In the simulations, the ranges of the observed bias (i.e., -2.1 to 0.15) and the variance (i.e., 0.005 to 20.0) were covered by the means and variances of the generated normally distributed errors. In this way, the simulation results are applicable for testing the volume models, and should also be applicable for any other forestry models, which have similar ranges of error parameters.

4.2 Reliability of the SATP When Errors Are iid Normal Variables With a Zero Mean and a Constant Variance

In the first simulation to validate the SATP procedures, the errors of a model were assumed to be iid normal variables with a zero mean and a constant variance. Since the mean of the unconditional distribution of errors, \(d\), measures the "average" bias of a model, a mean of zero indicates that the model estimates are the unbiased estimates of the true values. To assume that a model is unbiased may not be practical. However, a mean of zero was one of the assumptions required to derive the SATPs I and II in Chapter 3 corresponding to Freese's (1960) test I and II. Also, since SATP III is simply a modification of SATP I to allow for a non-zero mean of \(d\), the results by using SATP I and III should be very similar if the mean of \(d\) is zero. Therefore, this simulation was used to validate the SATP I, and to compare SATP I and III under the ideal distributional condition.

Under the assumption of \(\mu_d = 0\), the accuracy of a model will depend only on the
Table 4.2: Calculated statistics of the errors for three volume models and B.C. tree sectional data.

| Population Data | Application Data | Number of Trees | Model | $\bar{d}$ | $|\bar{d}|$ | $MSD$ |
|-----------------|------------------|-----------------|-------|-----------|-----------|-------|
| Douglas-fir (DF) | FIZ B (DF)       | 575             | [1]   | -0.05387 | 0.54590   | 1.40340 |
|                 |                  |                 | [2]   | -0.02250 | 0.08860   | 0.01186 |
|                 |                  |                 | [3]   | -0.00003 | 0.49640   | 1.44599 |
|                 |                  |                 | [1]   | -0.09664 | 0.17354   | 0.03284 |
|                 |                  |                 | [2]   | 0.09343  | 0.11107   | 0.00883 |
|                 | AC               | 1077            | [3]   | 0.12508  | 0.12549   | 0.01845 |
|                 |                  |                 | [1]   | -0.07586 | 0.17634   | 0.03720 |
|                 |                  |                 | [2]   | 0.16458  | 0.17170   | 0.00920 |
|                 |                  |                 | [3]   | 0.15048  | 0.15068   | 0.01955 |
| Aspen (AC)      | FIZ L (AC)       | 662             | [1]   | 0.01028  | 0.05135   | 0.00935 |
|                 |                  |                 | [2]   | 0.01128  | 0.05824   | 0.00536 |
|                 |                  |                 | [3]   | 0.00166  | 0.04326   | 0.00591 |
|                 |                  |                 | [1]   | 0.03839  | 0.05834   | 0.00643 |
|                 | PL               | 2880            | [2]   | 0.09170  | 0.11100   | 0.01005 |
|                 |                  |                 | [3]   | 0.02343  | 0.04792   | 0.00516 |
|                 |                  |                 | [1]   | -1.71060 | 1.71900   | 14.84238 |
|                 | DF               | 990             | [2]   | -0.18089 | 0.19973   | 0.02468 |
| Lodgepole pine (PL) | FIZ I (PL)    | 953             | [1]   | 0.00330  | 0.06997   | 0.01250 |
|                 |                  |                 | [2]   | -0.01080 | 0.06134   | 0.00663 |
|                 |                  |                 | [3]   | -0.00117 | 0.06156   | 0.00965 |
|                 | DF               | 990             | [2]   | -0.22150 | 0.22275   | 0.01554 |
|                 |                  |                 | [3]   | -0.86878 | 0.92135   | 2.74764 |
|                 | AC               | 1077            | [3]   | -0.04143 | 0.05756   | 0.00782 |
|                 |                  |                 | [2]   | -0.08247 | 0.09360   | 0.00689 |
|                 |                  |                 | [1]   | -0.03726 | 0.05570   | 0.00490 |

Note: The values of Model [1] and [3] are in cubic metre units, and the values of Model [2] are in the logarithmic units (base 10). Model [1], [2] and [3] represents Spurr's (1952), Logarithmic Schumacher and Hall's and Schumacher and Hall's (1933) volume models; $\bar{d}$ and $|\bar{d}|$ are mean difference and mean absolute difference between the estimated and actual tree volumes, respectively; $MSD$ is mean squared difference between the estimated and actual tree volumes; and FIZ is Forestry Inventory Zone.
precision (variance) of the model estimates. Since the purpose in applying the SATP procedures is to classify a model as either acceptable or unacceptable, the validity of the SATP procedures depends on whether the SATP procedures can correctly classify the error variance of a model into two wide classes based on two preset class limits (i.e., $\sigma_0^2$ and $\sigma_1^2$). However, the derivation of the SATP procedures were based on Wald’s stopping rule, $B < R_n < A$, where $A$ and $B$ had been chosen by Wald to approximately obtain the specified probabilities of the type I and II errors, $\alpha$ and $\beta$ (Wetherill and Glazebrook 1986, p.16). These approximations used for the actual error probabilities of the SATP procedures must be examined for various levels of $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$. Also, in order to determine the usefulness of the SATP procedures in terms of sampling cost savings, the expected sample size required to reach a decision by the SATP procedures should be investigated. Therefore, the concerns addressed by this simulation were:

1. Whether the SATP procedures work reliably for classifying the error variance, $\alpha^2$, of a volume model as either acceptable or unacceptable for various specified values of $\sigma_0^2$ and $\sigma_1^2$, and for different values of the true error variances;

2. Whether the specified error probabilities, $\alpha$ and $\beta$, are equal to the actual error probabilities found in applying the SATP procedures.

3. What is the expected sample size required by the SATP procedures for reaching a terminal decision for different values of $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$; and

To address these problems, three simulations were performed by using normal generators to simulate taking random observations from the errors of tree volume models. The means of these normal generators were all set to be zero, and the variance was chosen to include five levels, 0.003, 0.03, 0.3, 3.0, and 20.0. The values were chosen in order to cover the range of the $MSD$’s found from B.C. tree data and the three volume models (i.e., 0.005 to 20.0 from Table 4.2).
4.2.1 Simulation methods

Based on Wald's (1947) formulations, for a given SPRT plan, the OC value gives the probability of making a correct decision and the ASN value is the expected sample size required for making a terminal decision. The relationship between the OC value and the error probabilities of an SPRT are $OC_{H_0} = 1 - \alpha$ and $OC_{H_1} = \beta$, if the possible overshooting of the decision boundaries can be ignored, where $OC_{H_0}$ and $OC_{H_1}$ are the OC values when the null or alternative hypothesis is true, respectively. Therefore, in order to examine the reliability and usefulness of the SATP procedures, or to investigate the actual error probabilities of these extended procedures, the OC and ASN values for different values of $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$, which may be used in the practical application of volume model testing, must be estimated. The general simulation procedures used to estimate these desired OC and ASN values were:

1. The test parameters, $\sigma_0^2$, $\sigma_1^2$, and the error probabilities, $\alpha$ and $\beta$ (i.e., the nominal error probabilities) were first specified. In deriving the SATP, $\sigma_0^2$ and $\sigma_1^2$ were assumed to be determined based on the error limits, $e_0$ and $e_1$ (or $p_0$ and $p_1$), and the accuracy level, $\gamma$. Each set of $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$ represented a simulated SATP procedure. The OC and ASN values were then estimated for each simulated SATP procedure;

2. Each set of $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$ was then used to determine the decision boundaries of the SATP I or III using the formulae presented in Chapter 3 (e.g., Equations 3.55, 3.56 and 3.57 for SATP I). Since $\sigma_0^2$ and $\sigma_1^2$ were specified instead of $e_0$, $e_1$ and $\gamma$, the denominators of Equations 3.55, 3.56 and 3.57 were rewritten as $(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2})$ when these formulae were used;
3. For each simulated SATP procedure, the normal generator with a zero mean and variance, \( \sigma_m^2 \), denoted as \( N(0, \sigma_m^2) \) was used to simulate taking random observations from the unconditional distribution of errors, \( d \). The simulated SATP procedure was carried out to test the hypotheses \( H_0 : \sigma^2 = \sigma_0^2 \) vs. \( H_1 : \sigma^2 = \sigma_1^2 \). The variance of the normal generator that was used represented the true error variance of a model;

4. After each observation was taken, the computed test statistic (e.g., Equation 3.58 for SATP I) was compared to the determined boundaries (i.e., \( h_1 + sn \) and \( h_2 + sn \)). A terminal decision to accept \( H_0 \) or \( H_1 \) was made, or an additional observation was taken based on the rules corresponding to SATP I or III as outlined in Chapter 3. This step was repeated until a terminal decision was finally reached (no truncation point was set in the simulations); and

5. Steps 3 and 4 were repeated for a preset large number of trials (iterations) for each simulated SATP procedure. The percentage of times that \( H_0 \) was accepted was then calculated as the estimate of the OC value when the true error variance is \( \sigma_m^2 \), denoted as \( OC_{\sigma_m^2} \). The associated average number of the observations sampled was also calculated as the estimates of the ASN value, denoted as \( ASN_{\sigma_m^2} \).

Before the simulations were conducted, a sensitivity analysis indicated that in most cases, 500 and 1,000 iterations would result in adequate precisions for the Monte Carlo estimates of the ASN and OC values, respectively. However, for estimating the actual error probabilities, 2,000 iterations were used to obtain more precise estimates of the OC values.
4.2.2 Reliability of the SATP procedures for classifying the error variance of volume models

Under normality of $d$ with a zero mean and a constant variance, it was shown in Chapter 2 that the accuracy requirement (Equation 2.17) can be translated into an error variance bound (Equation 2.20). Freese's (1960) procedure then becomes a test to determine whether the error variance exceeds a given value (i.e., the user-supplied hypothesized variance, $\sigma^2_h$). In deriving SATPs I and III, the error variance bound, $\sigma^2_h$, was modified into two class limits of the error variance, $\sigma^2_0$ and $\sigma^2_1$, for model acceptance and rejection, respectively. For a given application, the relationship between the tested parameters, $\sigma^2_0$ and $\sigma^2_1$, and the true error variance, $\sigma^2$, may be described by one of the three following situations. These are: (1) $\sigma^2 \leq \sigma^2_0$ (i.e., the true error variance is in the acceptance region); (2) $\sigma^2 \geq \sigma^2_1$ (i.e., the true error variance is in the rejection region); and (3) $\sigma^2_0 < \sigma^2 < \sigma^2_1$ (i.e., the true error variance is between the acceptable and unacceptable variance limits). To determine the reliability of the SATP procedures, the probability of making a correct decision under each of these situations should be determined.

As a standard, Wald's (1947) requirements imposed for the OC function of an SPRT plan may be used for evaluating the reliability of the SATP procedures in terms of the probability of making a correct decision. For a given SPRT plan, the space of the test parameter ($\theta$) can be divided into three mutually exclusive regions. These are: (1) $\theta \leq \theta_0$; (2) $\theta_0 < \theta < \theta_1$; and (3) $\theta \geq \theta_1$. Regions (1) and (3) are called the acceptance and rejection regions, respectively. Region (2) is called the indifference region. It represents a situation for which thesis is no practical importance as to which decision is made if the true parameter falls in this region. In order to test two simple hypotheses, $H_0$ against $H_1$, Wald (1947, p.32) imposed some requirements on the OC function of an SPRT plan for given values of $\alpha$ and $\beta$. These were: (1) $OC$ must be equal to or greater than $1 - \alpha$
for the region of acceptance; (2) \( OC \) must be equal to or less than \( \beta \) for the region of rejection; and (3) \( OC = 1 - \alpha \) when \( \theta = \theta_0 \) and \( OC = \beta \) when \( \theta = \theta_1 \). These requirements are met only if the possible overshooting of the decision boundaries can be ignored.

To conduct this simulation, five values were selected for the true error variance (i.e., the variance of normal generator, \( \sigma^2_m \)) for each simulated SATP I or III plan. These were:

1. \( \sigma^2_{m1} = 0.9 \sigma^2_0 \);
2. \( \sigma^2_{m2} = \sigma^2_0 \);
3. \( \sigma^2_{m3} = \frac{1}{2}(\sigma^2_0 + \sigma^2_1) \);
4. \( \sigma^2_{m4} = \sigma^2_1 \); and
5. \( \sigma^2_{m5} = 1.1 \sigma^2_1 \).

The multipliers, 0.9 and 1.1, were chosen arbitrarily to give values less than the acceptable variance, \( \sigma^2_0 \), and greater than the unacceptable variance, \( \sigma^2_1 \), respectively. To cover a wider range of possible values of error variance, the following combinations were selected for \( \sigma^2_0, \sigma^2_1, \alpha \) and \( \beta \):

1. Five levels were set as the acceptable variance, \( \sigma^2_0 \) (i.e., 0.003, 0.03, 0.3, 3.0, 20.0).
   These values were selected to cover the range of MSD found in the tree data and volume models;

2. For each \( \sigma^2_0 \) level, three levels were arbitrarily chosen for the unacceptable variance, \( \sigma^2_1 \). These values were specified as a ratio of \( \sigma^2_1 \) to \( \sigma^2_0 \) (e.g., \( \sigma^2_1/\sigma^2_0 = 1.10, 1.20, 2.00 \)).
   The first two ratios were varied for different values of \( \sigma^2_0 \) in order to cover a wide range of simulated conditions. For the purpose of comparison, the last ratio was the same (i.e., 2.00 was used). These values do not represent the optimal choice for the interval of \( \sigma^2_0 \) to \( \sigma^2_1 \), but only provide some indications of how to set this interval for practical application; and

3. Two levels of nominal error probabilities were chosen. These were \( \alpha = \beta = 0.05 \) and 0.10. These values were commonly used for Wald’s SPRT plans in forest pest surveys.

These selected combinations resulted in a total of 30 simulated SATP I or III plans (i.e., \( 5 \times 3 \times 2 = 30 \)).
Chapter 4. Monte Carlo Simulation

For each simulated SATP I plan (a combination of \( \sigma_0^2 \), \( \sigma_1^2 \), \( \alpha \) and \( \beta \)), the normal generators with a zero mean and one of five variances (i.e., \( \sigma_{m1}^2 \), \( \sigma_{m2}^2 \), \( \sigma_{m3}^2 \), \( \sigma_{m4}^2 \) and \( \sigma_{m5}^2 \)) were used to simulate taking random observations from the errors of a volume model. For each normal generator used, the procedures described in section 4.2.1 were used to estimate the OC value of SATP I for a given true variance. For the five generators used, the estimated OC values were obtained and denoted as \( \hat{OC}_{0.9\sigma_0^2} \), \( \hat{OC}_{\sigma_1^2} \), \( \hat{OC}_{\sigma_2^2} \), \( \hat{OC}_{\sigma_4^2} \) and \( \hat{OC}_{1.1\sigma_1^2} \) representing the estimated OC value when the true error variance was less than \( \sigma_0^2 \), equal to \( \sigma_0^2 \) and \( \sigma_1^2 \) (i.e., \( \sigma_{m3}^2 \)), equal to \( \sigma_1^2 \) and greater than \( \sigma_1^2 \). Since the estimate of interest in this simulation was the OC value, 1,000 iterations were used based on the results of a sensitivity analysis.

For a given distributional generator, and level of \( \alpha = \beta \), the estimated OC values were the same when the ratio of \( \sigma_1^2 \) to \( \sigma_0^2 \) was the same (i.e., 2.00). Therefore, the results for \( \sigma_1^2/\sigma_0^2 = 2.00 \) are only given for \( \sigma_0^2 = 20.0 \) (Tables 4.3 and 4.4). This result is explainable.

From the derived decision boundaries of SATP I (Equation 3.54), the OC value of an SATP I depends only on the relationship between \( \sigma_1^2/\sigma_0^2 \) and \( \sigma^2/\sigma_1^2 \) for given \( \alpha \) and \( \beta \), where \( \sigma^2 \) is the true variance. Because the choice of the true variance (i.e., the variance of normal generator) was the same for each pair of \( \sigma_0^2 \) and \( \sigma_1^2 \) (i.e., \( \sigma^2 = \sigma_0^2 \), \( \sigma_1^2 \) etc.), simulated SATP I plans using the same ratio of \( \sigma_1^2 \) to \( \sigma_0^2 \) would result in the same OC values.

Based on the Monte Carlo OC estimates obtained, the reliability of SATP I for classifying the error variance of a model into two pre-defined classes was examined by comparing the resulting OC values with Wald's (1947) stated requirements for the OC function of an SPRT. These were:

1. To confirm the requirement, \( OC \geq 1 - \alpha \), for the acceptance region (i.e., \( \sigma^2 \leq \sigma_0^2 \)), the values of \( \hat{OC}_{0.9\sigma_0^2} \) and \( \hat{OC}_{\sigma_0^2} \) were examined. It was found that the resulting OC
values satisfied the requirement for all levels of the specified variances, and \( \alpha \) and \( \beta \). Since the OC values in this region directly represent the probability of making a correct decision, the resulting OC values were all greater than their nominal values (i.e., \( 100(1 - \alpha) = 95\% \) or \( 90\% \)). This shows that under the normality assumption with a zero mean and constant variance of \( d \), the developed SATP I worked reliably for classifying the variance (precision) of a volume model into the acceptance class when the true variance of errors is indeed within the specified acceptance level.

2. To confirm the requirement, \( OC \leq \beta \), for the rejection region (i.e., \( \sigma^2 \geq \sigma_1^2 \)), the values of \( OC_{\sigma_1^2} \) and \( OC_{1,1\sigma_1^2} \) were examined. It was also found that the resulting OC values satisfied the requirement. The values of \( 1 - OC \) in the region of rejection represent the probability of making a correct decision (i.e., rejecting \( H_0 \) and concluding \( H_1 \)). Since all resulting OC values were less than \( \beta \), the probabilities of making a correct decision were greater than their nominal values \( 100(1 - \beta) = 95\% \) or \( 90\% \) when the true variance of error fell in the rejection region.

3. For the region of indifference (i.e., \( \sigma_0^2 < \sigma^2 < \sigma_1^2 \)), Wald (1947) did not indicate any requirement for the OC values. However, the values should be less than \( 1 - \alpha \) and greater than \( \beta \). The resulting OC values (i.e., \( OC_{\sigma_0^2} \)) ranged from 0.359 to 0.474 for \( \alpha = \beta = 0.05 \), and ranged from 0.397 to 0.495 for \( \alpha = \beta = 0.10 \).

The OC values of SATP III were also obtained using the same simulation procedures. The results were almost identical to those obtained for the SATP I. Based on the results obtained, it can be concluded that the SATP I and III are reliable for classifying the error variance of a model into two predetermined classes (acceptable and unacceptable) when the model errors are iid normal variables with a zero mean and a constant variance.
Table 4.3: Monte Carlo OC estimates of SATP I for testing $H_0 : \sigma^2 = \sigma^2_0$ vs. $H_1 : \sigma^2 = \sigma^2_1$ with $\alpha = \beta = 0.05$ when five normal generators with zero means and different variances were used to simulate the errors of volume models.

<table>
<thead>
<tr>
<th>$\sigma^2_0$</th>
<th>$\sigma^2_1/\sigma^2_0$</th>
<th>$N(0, 0.9\sigma^2_0)$</th>
<th>$\overline{OC}_{0.9\sigma^2_0}$</th>
<th>$N(0, \sigma^2_0)$</th>
<th>$\overline{OC}_{\sigma^2_0}$</th>
<th>$N(0, \sigma^2_1)$</th>
<th>$\overline{OC}_{\sigma^2_1}$</th>
<th>$N(0, 1.1\sigma^2_1)$</th>
<th>$\overline{OC}_{1.1\sigma^2_1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.003</td>
<td>1.15</td>
<td>1.000</td>
<td>0.999</td>
<td>0.981</td>
<td>0.435</td>
<td>0.049</td>
<td>0.049</td>
<td>0.000</td>
<td>0.006</td>
</tr>
<tr>
<td>0.030</td>
<td>1.25</td>
<td>0.998</td>
<td>0.974</td>
<td>0.472</td>
<td>0.048</td>
<td>0.048</td>
<td>0.048</td>
<td>0.009</td>
<td>0.012</td>
</tr>
<tr>
<td>0.300</td>
<td>1.35</td>
<td>0.996</td>
<td>0.970</td>
<td>0.452</td>
<td>0.048</td>
<td>0.048</td>
<td>0.048</td>
<td>0.014</td>
<td>0.015</td>
</tr>
<tr>
<td>3.000</td>
<td>1.45</td>
<td>0.993</td>
<td>0.965</td>
<td>0.443</td>
<td>0.047</td>
<td>0.047</td>
<td>0.047</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>20.000</td>
<td>1.55</td>
<td>0.991</td>
<td>0.965</td>
<td>0.434</td>
<td>0.043</td>
<td>0.043</td>
<td>0.043</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>20.000</td>
<td>1.60</td>
<td>0.989</td>
<td>0.970</td>
<td>0.359</td>
<td>0.048</td>
<td>0.048</td>
<td>0.048</td>
<td>0.024</td>
<td></td>
</tr>
</tbody>
</table>

Note: The results are based on 1,000 iterations. $\sigma^2_0$ and $\sigma^2_1$ are the specified acceptable and unacceptable variance limits of model errors, respectively; $\alpha$ and $\beta$ are the specified probabilities of the type I and II errors for the test, respectively; $N(0, \sigma^2)$ represents the normal generator with zero and variance $\sigma^2$, which is used for simulating the error of a tree volume model; $\sigma^2_0 = \frac{1}{2}(\sigma^2_0^2 + \sigma^2_1^2)$; $\overline{OC}_{\sigma^2}$ is the Monte Carlo OC value when the true variance is $\sigma^2_0$; $0.9\sigma^2_0$ and $1.1\sigma^2_0$ were arbitrarily chosen to represent the cases when the true variance is less than $\sigma^2_0$ and greater than $\sigma^2_0$, respectively.
Table 4.4: Monte Carlo OC estimates of SATP I for testing $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with $\alpha = \beta = 0.10$ when five normal generators with zero means and constant variances were used to simulate the errors of volume models.

<table>
<thead>
<tr>
<th>$\sigma_0^2$</th>
<th>$\sigma_1^2 / \sigma_0^2$</th>
<th>$N(0, 0.9\sigma_0^2)$</th>
<th>$N(0, \sigma_0^2)$</th>
<th>$N(0, \sigma_1^2)$</th>
<th>$N(0, 1.0\sigma_1^2)$</th>
<th>$N(0, 1.1\sigma_1^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.003</td>
<td>1.20</td>
<td>0.995</td>
<td>0.942</td>
<td>0.495</td>
<td>0.094</td>
<td>0.016</td>
</tr>
<tr>
<td>0.030</td>
<td>1.30</td>
<td>0.991</td>
<td>0.931</td>
<td>0.477</td>
<td>0.094</td>
<td>0.021</td>
</tr>
<tr>
<td>0.300</td>
<td>1.40</td>
<td>0.985</td>
<td>0.929</td>
<td>0.478</td>
<td>0.093</td>
<td>0.026</td>
</tr>
<tr>
<td>1.500</td>
<td>1.50</td>
<td>0.980</td>
<td>0.938</td>
<td>0.472</td>
<td>0.095</td>
<td>0.029</td>
</tr>
<tr>
<td>2.000</td>
<td>1.55</td>
<td>0.977</td>
<td>0.936</td>
<td>0.464</td>
<td>0.096</td>
<td>0.031</td>
</tr>
<tr>
<td>2.000</td>
<td>1.60</td>
<td>0.974</td>
<td>0.941</td>
<td>0.445</td>
<td>0.099</td>
<td>0.038</td>
</tr>
<tr>
<td>2.000</td>
<td>2.00</td>
<td>0.973</td>
<td>0.945</td>
<td>0.397</td>
<td>0.096</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Note: The results are based on 1,000 iterations. $\sigma_0^2$ and $\sigma_1^2$ are the specified acceptable and unacceptable variance limits of model errors, respectively; $\alpha$ and $\beta$ are the specified probabilities of the type I and II errors for the test, respectively; $N(0, \sigma^2)$ represents the normal generator with zero and variance $\sigma^2$, which is used for simulating the errors of a tree volume model; $\sigma_0^2 = \frac{1}{3}(\sigma_0^2 + \sigma_1^2)$; $O\bar{C}_{\sigma_0^2}$ is the Monte Carlo OC value when the true variance is $\sigma_0^2$; $0.9\sigma_0^2$ and $1.1\sigma_1^2$ were arbitrarily chosen to represent the cases when the true variance is less than $\sigma_0^2$ and greater than $\sigma_1^2$, respectively.
4.2.3 Accuracy of the nominal error probabilities of the SATP

Simulations were performed to examine whether the specified error probabilities, $\alpha$ and $\beta$, were equal to the actual error probabilities of the SATP procedures. This examination was considered necessary, because Wald's (1947) stopping rule, $B < R_m < A$, was used to derive the decision boundaries of the SATP procedures. Wald (1947, p.46) stated that since $\alpha$ and $\beta$ were usually small (commonly less than 0.05), the approximate boundaries of $A (A \cong (1 - \beta)/\alpha)$ and $B (B \cong \beta/(1 - \alpha))$ would result in actual error probabilities of an SPRT that were very nearly equal to the specified values, $\alpha$ and $\beta$. However, after examining some SPRT plans used in forest pest surveys, Fowler (1983) indicated since $\alpha$ and $\beta$ were usually 0.10 for most sequential sampling plans for pest surveys, the difference between the nominal and actual error probabilities of Wald's SPRT as a result of overshooting the decision boundaries could be large.

In order to evaluate the accuracy of the nominal error probabilities of the SATP procedures, Monte Carlo estimates of the actual probabilities of the type I ($\alpha_m$) and II ($\beta_m$) errors were obtained, and compared to the nominal values. In this simulation, $\alpha_m$ and $\beta_m$ were estimated based on Wald's (1947) requirements imposed on the OC function of an SPRT. These are that the OC value of an SPRT when the null hypothesis is true should be $1 - \alpha$ (i.e., $OC_{H_0} = 1 - \alpha$), and the OC value when the alternative hypothesis is true should be $\beta$ ($OC_{H_1} = \beta$). Therefore, the actual probabilities of $\alpha$ and $\beta$ of an SATP can be estimated through Monte Carlo estimates of the OC values when the true error variance is equal to $\sigma_0^2$ or $\sigma_1^2$, respectively (i.e., $OC_{\sigma_0^2}$ and $OC_{\sigma_1^2}$). The estimated actual error probabilities of the type I and II errors for an SATP were:

$$\alpha_m = 1 - OC_{\sigma_0^2}$$

$$\beta_m = OC_{\sigma_1^2}$$
The procedures described in Section 4.2.1 were used again to estimate \( \hat{OC}_{\sigma_0} \) and \( \hat{OC}_{\sigma_1} \) for the same sets of the simulated SATP I used in Section 4.2.2. Since the estimates of interest were the actual error probabilities, 2,000 iterations were used in this simulation.

To compare the estimated actual error probabilities to the nominal error probabilities, the following values were calculated: (1) the difference between nominal and actual probabilities (i.e., \( DE_\alpha = \alpha - \alpha_m \) and \( DE_\beta = \beta - \beta_m \)); and (2) the relative error of the nominal error probabilities (i.e., \( RE_\alpha = (\alpha - \alpha_m)/\alpha_m \) and \( RE_\beta = (\beta - \beta_m)/\beta_m \)).

The nominal error probabilities were consistently higher than the estimated actual error probabilities (Tables 4.5 and 4.6). In general, the differences were larger for \( \alpha \) than for \( \beta \). For example, when the nominal error probabilities were set to \( \alpha = \beta = 0.05 \), the differences between the nominal and estimated actual \( \alpha \) ranged from 0.0155 to 0.0395, and the relative errors of the nominal \( \alpha \) ranged from 0.45 to 3.75. However, the differences between the nominal and estimated actual \( \beta \) only ranged from 0.0050 to 0.0110, and the relative errors of the nominal \( \beta \) only ranged from 0.11 to 0.30.

Also, the smallest value of \( \sigma_0^2 \) and \( \sigma_1^2 \) chosen in this simulation (i.e., \( \sigma_0^2 = 0.003 \) and \( \sigma_1^2/\sigma_0^2 = 1.15 \)) had the narrowest distance between the acceptance and the rejection lines (i.e., \( h_2 - h_1 = 0.271 \) when \( \alpha = \beta = \gamma = 0.05 \)), it resulted in the largest errors (\( DE_\alpha = 0.0395 \) and \( DE_\beta = 0.0110 \)). The largest value of \( \sigma_0^2 \) and \( \sigma_1^2 \) (i.e., \( \sigma_0^2 = 20.0 \) and \( \sigma_1^2/\sigma_0^2 = 2.00 \)) had the widest distance between the acceptance and the rejection lines (i.e., \( h_2 - h_1 = 471.112 \) when \( \alpha = \beta = \gamma = 0.05 \)), and resulted in the smaller errors (\( DE_\alpha = 0.0195 \) and \( DE_\beta = 0.0050 \)).

These results suggest that the difference between the nominal and actual error probabilities is a result of overshoting the decision boundaries of the SATP. As the interval between two decision lines (\( L_0 \) and \( L_1 \)) decreases, overshoting of the decision boundaries becomes more common. Therefore, the SATP procedures may be viewed as conservative.
testing procedures since these procedures always resulted in a smaller probability of making a wrong decision (a high precision testing procedure). On the other hand, increasing the precision of a testing procedure usually results in a larger sample size. The results suggest that the sample size required by the SATP I procedure is likely larger than that required to meet the specified probabilities of errors.

The accuracies of the nominal error probabilities of the SATP procedures found in this simulation were very similar to those obtained by Fowler (1978; 1983) for SPRT plans used in forest pest surveys. Therefore, in some situations, if the potential users of the SATP procedures want to obtain correct nominal error probabilities, the Monte Carlo OC and ASN functions presented in Section 2.4.1.5 may be used with the SATP procedures.

The same simulation procedures were also carried out for evaluating the accuracy of the nominal error probabilities of SATP III. The results were almost the same as those obtained for SATP I.

### 4.2.4 Expected sample size of the SATP procedures

Using normal generators with zero means and constant variance, simulations were designed to provide some indication of sampling cost savings in using the SATP procedures. Since the sample size in a sequential plan is a random variable, the expected sample size (ASN) required for reaching a terminal decision was examined.

To design this simulation, a wider range of the four specified parameters, $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$ was examined. Two situations were considered. These were: (1) fixed $\sigma_0^2$, and increasing $\sigma_1^2$ (e.g., $\sigma_0^2 = 0.003$, $\sigma_1^2 = 1.25\sigma_0^2$, $1.5\sigma_0^2$, $2\sigma_0^2$); and (2) fixed $\sigma_1^2$, and increasing $\sigma_0^2$ (e.g., $\sigma_1^2 = 0.03$, $\sigma_0^2 = 0.65\sigma_1^2$, $0.75\sigma_1^2$, $0.85\sigma_1^2$). The choice of these combinations were arbitrary. Also, three levels were set for $\alpha$ and $\beta$ (i.e., 0.05, 0.10, and 0.20). Similarly to the OC values, for $\sigma^2 = \sigma_0^2$, the ASN values of an SATP procedure will depend only on
Chapter 4. Monte Carlo Simulation

Table 4.5: Monte Carlo estimates of the actual error probabilities when the SATP I is applied to test $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with various $\sigma_0^2$, $\sigma_1^2$ and $\alpha = \beta = 0.05$.

<table>
<thead>
<tr>
<th>$\sigma_0^2$</th>
<th>$\sigma_1^2 / \sigma_0^2$</th>
<th>$\alpha_m$</th>
<th>$DE_{\alpha}$</th>
<th>$RE_{\alpha}$</th>
<th>$\beta_m$</th>
<th>$DE_{\beta}$</th>
<th>$RE_{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.15</td>
<td>1.20</td>
<td>0.0105</td>
<td>0.0395</td>
<td>3.75</td>
<td>0.0390</td>
<td>0.0110</td>
<td>0.28</td>
</tr>
<tr>
<td>0.003</td>
<td>1.25</td>
<td>0.0200</td>
<td>0.0300</td>
<td>1.50</td>
<td>0.0435</td>
<td>0.0065</td>
<td>0.15</td>
</tr>
<tr>
<td>0.030</td>
<td>1.30</td>
<td>0.0255</td>
<td>0.0245</td>
<td>0.96</td>
<td>0.0440</td>
<td>0.0060</td>
<td>0.14</td>
</tr>
<tr>
<td>0.035</td>
<td>1.35</td>
<td>0.0310</td>
<td>0.0190</td>
<td>0.61</td>
<td>0.0435</td>
<td>0.0065</td>
<td>0.15</td>
</tr>
<tr>
<td>0.300</td>
<td>1.40</td>
<td>0.0345</td>
<td>0.0155</td>
<td>0.45</td>
<td>0.0435</td>
<td>0.0065</td>
<td>0.15</td>
</tr>
<tr>
<td>3.000</td>
<td>1.45</td>
<td>0.0330</td>
<td>0.0170</td>
<td>0.52</td>
<td>0.0425</td>
<td>0.0075</td>
<td>0.18</td>
</tr>
<tr>
<td>1.55</td>
<td>1.50</td>
<td>0.0330</td>
<td>0.0170</td>
<td>0.52</td>
<td>0.0400</td>
<td>0.0100</td>
<td>0.25</td>
</tr>
<tr>
<td>20.000</td>
<td>1.60</td>
<td>0.0340</td>
<td>0.0160</td>
<td>0.47</td>
<td>0.0390</td>
<td>0.0110</td>
<td>0.28</td>
</tr>
<tr>
<td>2.00</td>
<td>2.00</td>
<td>0.0305</td>
<td>0.0195</td>
<td>0.64</td>
<td>0.0450</td>
<td>0.0050</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Note: The estimates were based on 2,000 iterations. $\alpha_m$ and $\beta_m$ are the estimated actual probabilities of type I and II errors, respectively; $DE_{\alpha}$ and $DE_{\beta}$ are the differences between the nominal and actual values of type I and II error, respectively; $RE_{\alpha}$ and $RE_{\beta}$ are the relative differences of the nominal type I and II errors to their actual estimates, respectively.
Chapter 4. Monte Carlo Simulation

Table 4.6: Monte Carlo estimates of the actual error probabilities when the SATP I is applied to test \( H_0 : \sigma^2 = \sigma_0^2 \) vs. \( H_1 : \sigma^2 = \sigma_1^2 \) with various \( \sigma_0^2, \sigma_1^2 \) and \( \alpha = \beta = 0.10 \).

<table>
<thead>
<tr>
<th>( \sigma_0^2 )</th>
<th>( \sigma_1^2 / \sigma_0^2 )</th>
<th>( \alpha_m )</th>
<th>( DE_\alpha )</th>
<th>( RE_\alpha )</th>
<th>( \beta_m )</th>
<th>( DE_\beta )</th>
<th>( RE_\beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.003</td>
<td>1.15</td>
<td>0.0470</td>
<td>0.0530</td>
<td>1.13</td>
<td>0.0810</td>
<td>0.0190</td>
<td>0.23</td>
</tr>
<tr>
<td>0.030</td>
<td>1.20</td>
<td>0.0605</td>
<td>0.0395</td>
<td>0.66</td>
<td>0.0870</td>
<td>0.0130</td>
<td>0.15</td>
</tr>
<tr>
<td>0.300</td>
<td>1.25</td>
<td>0.0680</td>
<td>0.0320</td>
<td>0.47</td>
<td>0.0875</td>
<td>0.0125</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.300</td>
<td>1.30</td>
<td>0.0700</td>
<td>0.0300</td>
<td>0.43</td>
<td>0.0875</td>
<td>0.0125</td>
<td>0.14</td>
</tr>
<tr>
<td>3.000</td>
<td>1.35</td>
<td>0.0670</td>
<td>0.0330</td>
<td>0.49</td>
<td>0.0860</td>
<td>0.0140</td>
<td>0.16</td>
</tr>
<tr>
<td>20.000</td>
<td>1.40</td>
<td>0.0645</td>
<td>0.0355</td>
<td>0.55</td>
<td>0.0890</td>
<td>0.0110</td>
<td>0.12</td>
</tr>
<tr>
<td>20.000</td>
<td>1.45</td>
<td>0.0660</td>
<td>0.0320</td>
<td>0.52</td>
<td>0.0920</td>
<td>0.0125</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.000</td>
<td>1.50</td>
<td>0.0665</td>
<td>0.0335</td>
<td>0.50</td>
<td>0.0940</td>
<td>0.0060</td>
<td>0.06</td>
</tr>
<tr>
<td>20.000</td>
<td>1.55</td>
<td>0.0640</td>
<td>0.0360</td>
<td>0.56</td>
<td>0.0965</td>
<td>0.0035</td>
<td>0.04</td>
</tr>
<tr>
<td>20.000</td>
<td>1.60</td>
<td>0.0595</td>
<td>0.0405</td>
<td>0.68</td>
<td>0.0960</td>
<td>0.0040</td>
<td>0.04</td>
</tr>
<tr>
<td>20.000</td>
<td>2.00</td>
<td>0.0550</td>
<td>0.0450</td>
<td>0.82</td>
<td>0.0935</td>
<td>0.0065</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Note: The estimates were based on 2,000 iterations. \( \alpha_m \) and \( \beta_m \) are the estimated actual probabilities of type I and II errors, respectively; \( DE_\alpha \) and \( DE_\beta \) are the differences between the nominal and actual values of type I and II error, respectively; \( RE_\alpha \) and \( RE_\beta \) are the relative differences of the nominal type I and II errors to their actual estimates, respectively.
Chapter 4. Monte Carlo Simulation

\[ \alpha, \beta \text{ and } \sigma_2^2/\sigma_1^2. \] That is, for given \( \alpha \) and \( \beta \), ASN values can be expected to be the same if the ratio of \( \sigma_2^2 \) to \( \sigma_1^2 \) is same. Finally, since \( \sigma_0^2 \) and \( \sigma_1^2 \) were specified instead of \( \epsilon_0 \) and \( \epsilon_1 \), the specification of \( \gamma \) was not required.

For each combination of \( \sigma_0^2 \), \( \sigma_1^2 \), \( \alpha \) and \( \beta \), the normal generator, \( N(0, \sigma_0^2) \) was used to simulate taking random observations from the distribution of \( d \). The simulation procedures described in section 4.2.1 were then used to estimate the ASN values when \( \sigma_0^2 \) is the true error variance. Since the estimate of interest is the ASN value, 500 iterations was used based on the results of the sensitivity analysis reported previously.

In general, it was found that the expected sample size (ASN) required to reach a terminal decision decreased rapidly for fixed levels of error probabilities (Table 4.7) as the interval between two class limits (\( \sigma_2^2 \) and \( \sigma_1^2 \)) increased. By fixing the level of \( \alpha \) (or \( \beta \)), and increasing the level of \( \beta \) (or \( \alpha \)), ASN also decreased. These results can be explained by examining the equations for the decision boundaries of SATP I. The decision boundaries of SATP I (i.e., \( h_1 + sn \) and \( h_2 + sn \)) are always two parallel lines. The intercepts (\( h_1 \) and \( h_2 \)) of these lines are determined by Equations 3.55 and 3.56, and the common slope (\( s \)) is determined by Equation 3.57. Intuitively, any change that causes an increase in the distance between these two decision lines (the absolute difference of \( h_1 \) and \( h_2 \)), or an increase in the value of \( s \) will result in a higher ASN. For example, for given values of \( \alpha \) and \( \beta \), as the difference between the two class limits increases, the denominators of the three equations will increase, thus resulting in a rapid decrease in the values of \( h_1 \) and \( h_2 \). The distance between the two decision lines is therefore closer, and ASN values will decrease. Similarly, for given values of the two class limits, as the value of \( \alpha \) or \( \beta \) increases, the distance between the two decision lines decreases, and the resulting ASN then decreases.

Based on the results of this simulation, it was found that for any \( \alpha \) and \( \beta \) examined, a resulting ASN value of less than 30 observations could be obtained by using a ratio of \( \sigma_1^2 \)
Table 4.7: Monte Carlo ASN estimates and the associated standard error error when $\sigma^2 = \sigma_0^2$ for applying SATP I to test $H_0: \sigma^2 = \sigma_0^2$ vs. $H_1: \sigma^2 = \sigma_1^2$ with various combinations of $\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$.

| $\sigma_0^2$ | $\sigma_1^2$ | $\alpha = 0.05$ | | $\alpha = 0.10$ | | $\alpha = 0.20$ |
|---|---|---|---|---|---|
| $\beta = \beta = \beta = 0.05$ | $\beta = \beta = \beta = 0.10$ | $\beta = \beta = \beta = 0.20$ | $\beta = \beta = \beta = 0.20$ | $\beta = \beta = \beta = 0.20$ |
| 1.25$\sigma_0^2$ | 240.8 | 177.8 | 117.6 | 222.2 | 159.8 | 105.9 | 195.2 | 135.3 | 79.6 |
| 0.60 | 0.03  | (7.080) | (5.999) | (4.319) | (6.449) | (5.120) | (3.888) | (5.867) | (4.339) | (2.531) |
| 1.50$\sigma_0^2$ | 75.4 | 57.8 | 39.7 | 71.0 | 53.6 | 35.8 | 62.4 | 46.0 | 28.5 |
| 0.30 | 0.03 | (2.111) | (1.755) | (1.443) | (1.860) | (1.529) | (1.232) | (1.511) | (1.264) | (0.873) |
| 2.00$\sigma_0^2$ | 29.8 | 22.8 | 15.9 | 28.3 | 21.7 | 14.7 | 26.1 | 19.4 | 12.2 |
| 0.65$\sigma_1^2$ | (0.798) | (0.651) | (0.536) | (0.727) | (0.622) | (0.465) | (0.658) | (0.520) | (0.324) |
| 0.75$\sigma_1^2$ | 67.5 | 51.5 | 35.2 | 64.2 | 47.9 | 32.0 | 56.6 | 41.6 | 26.0 |
| 0.30 | 0.03 | (1.815) | (1.556) | (1.244) | (1.661) | (1.381) | (1.077) | (1.368) | (1.142) | (0.795) |
| 0.85$\sigma_1^2$ | 144.1 | 108.5 | 74.6 | 136.5 | 101.0 | 65.6 | 117.4 | 83.3 | 51.1 |
| 0.20 | 0.03 | (4.134) | (3.589) | (2.961) | (3.853) | (3.218) | (2.384) | (2.944) | (2.379) | (1.660) |
| 1.30$\sigma_0^2$ | 428.0 | 321.8 | 216.6 | 406.5 | 301.4 | 197.1 | 359.1 | 254.1 | 142.8 |
| 0.30 | 0.03 | (13.237) | (10.861) | (8.508) | (12.618) | (10.207) | (7.867) | (11.372) | (7.978) | (4.855) |
| 1.60$\sigma_0^2$ | 174.8 | 129.1 | 86.9 | 164.1 | 117.3 | 77.8 | 141.2 | 99.3 | 61.0 |
| 0.20 | 0.03 | (5.236) | (4.239) | (3.415) | (4.856) | (3.608) | (2.771) | (3.986) | (2.888) | (2.052) |
| 1.90$\sigma_0^2$ | 57.9 | 44.6 | 30.4 | 55.3 | 42.2 | 27.9 | 49.8 | 36.3 | 22.6 |
| 0.20 | 0.03 | (1.531) | (1.338) | (1.028) | (1.435) | (1.269) | (0.939) | (1.254) | (0.996) | (0.676) |
| 0.60$\sigma_1^2$ | 34.2 | 26.2 | 18.5 | 32.4 | 24.6 | 16.8 | 29.7 | 22.0 | 13.9 |
| 0.70$\sigma_1^2$ | (0.898) | (0.769) | (0.635) | (0.814) | (0.705) | (0.545) | (0.737) | (0.626) | (0.394) |
| 0.30 | 0.03 | (1.355) | (1.114) | (0.907) | (1.240) | (1.076) | (0.850) | (1.113) | (0.894) | (0.580) |
| 0.80$\sigma_1^2$ | 96.8 | 73.0 | 50.3 | 90.8 | 67.4 | 44.9 | 77.9 | 57.1 | 35.7 |
| 0.20 | 0.03 | (2.964) | (2.407) | (1.818) | (2.682) | (2.089) | (1.507) | (1.962) | (1.563) | (1.125) |
| 3.00 | 0.03 | (240.8) | (177.8) | (117.6) | (222.2) | (159.8) | (105.9) | (195.2) | (135.3) | (79.6) |
| 0.87 | (7.080) | (5.999) | (4.519) | (6.448) | (5.120) | (3.888) | (5.877) | (4.339) | (2.531) |

Note: the results are based on 500 iterations; and the numbers in the brackets are the standard error error of ASN. When $\sigma_0^2$ is specified as 0.60$\sigma_1^2$, it means that the value of $\sigma_1^2$ is first specified, then $\sigma_0^2$ is determined as 0.60$\sigma_1^2$; similarly, when $\sigma_1^2$ is specified as 1.25$\sigma_0^2$, it means that the value of $\sigma_0^2$ is first specified, then $\sigma_1^2$ is determined as 1.25$\sigma_0^2$. 
to $\sigma_0^2$ that is larger than 2.0. To test the hypothesis $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$, the sample size ($n$) required by a fixed sample size procedure to have the desired values of $\alpha$ and $\beta$ will be the value, $n$, which satisfies the following relationship:

$$\frac{\chi_{1-\alpha}^2(n-1)}{\chi_\beta^2(n-1)} = \frac{\sigma_1^2}{\sigma_0^2}$$

(4.73)

where $\chi_{1-\alpha}^2(n-1)$ and $\chi_\beta^2(n-1)$ are the 100$(1-\alpha)$ and 100$\beta$ percentiles of a chi-square distribution with $(n-1)$ degrees of freedom. Since it has been previously indicated that the OC and ASN values of an SATP procedure when $\sigma_0^2$ or $\sigma_1^2$ is true will depend only on the ratio $\sigma_1^2/\sigma_0^2$ for given $\alpha$ and $\beta$, the resulting ASN values of the simulated SATP I may be compared to the sample sizes required by a fixed sample size procedure with the same $\alpha$ and $\beta$ levels, and same value of the $\sigma_1^2/\sigma_0^2$ ratio. However, based on last section, the SATP procedures are always resulted in smaller error probabilities than those specified. For a simulated SATP I with a ratio, $\sigma_1^2/\sigma_0^2$, of 2.0, the actual probabilities of the type I and II errors ($\alpha_m$ and $\beta_m$) were 0.0305 and 0.0450, respectively, when the specified values were $\alpha = \beta = 0.05$ (Table 4.5). The actual probabilities of the type I and II errors were 0.0550 and 0.0935, respectively, when the specified values were $\alpha = \beta = 0.10$ (Table 4.6). Therefore, when making a comparison to fixed sample size procedures, these actual error probabilities should be used instead of those specified.

Based on Equation 4.73, using a fixed sample size procedure for testing the hypotheses, $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$, the sample size for $\alpha = 0.0305$, $\beta = 0.0450$, and a ratio, $\sigma_1^2/\sigma_0^2$, of 2.0 is 55. Similarly, the sample size of a fixed sample size procedure required for $\alpha = 0.0550$, $\beta = 0.0935$, and a ratio, $\sigma_1^2/\sigma_0^2$, of 2.0 is 35. The resulting ASN values of the simulated SATP I that correspond to these two fixed sample size tests are 29.8 and 21.7 (Table 4.7). This means that in doing the same tests, SATP procedures use only 54.2% and 62% of sample sizes, respectively, when compared to the fixed sample size procedures with the same error probabilities. This result is similar to those found
by Fowler (1978). That is, the use of SATP procedures will also be expected to result in a 40 to 60% of sampling cost-saving compared to an equally reliable fixed sample size procedure.

The ASN values of SATP III for all combinations of the four parameters were also estimated and examined. The results were similar to those obtained for SATP I. However, ASN values of SATP III were one to five observations higher than those of SATP I. This result may be explained by the fact that SATP III requires that the sample mean of d be estimated, and therefore, a minimum of two observations is needed to make a terminal decision. SATP I does not have this requirement.

4.3 Reliability of the SATP When Errors Are iid Normal Variables With a Non-Zero Mean and a Constant Variance

Simulations were also performed to validate the SATP procedures, assuming that the unconditional distribution of the errors of a volume model are iid normal variable with a non-zero mean (i.e., estimation bias is present). Acknowledging that the mean of model errors is not zero may be more realistic for applicability testing of forestry models, because changes in growth type, site quality, or genotype between forest or species likely cause systematic estimation errors.

From Chapter 2, the derivation of Freese's (1960) test I and II requires an assumption that d or d/y is normally distributed with a zero mean. This assumption is critical for translating the probability statement of accuracy requirement (Equation 2.17) into the variance bound (Equation 2.20). If the mean of d or d/y is not zero, this translation is not possible, because the distribution of d² or (d/y)² will not be the standard chi-square, and will instead be the noncentral chi-square. If the errors, d, is the variable of interest and a non-zero mean of d is found, Freese (1960) suggested that \( \sum_{i=1}^{n} (d_i - \bar{d})^2 \) should be used
for measuring the actual accuracy of a model (i.e., Freese’s test III) instead of $\sum_{i=1}^{n} d_i^2$. Therefore, SATP III is simply a modification of SATP I to allow for a non-zero mean of $d$. However, as indicated in Chapter 3, SATP III was not derived by directly computing the likelihood ratio, $R_n$, for the situation when $\mu_d \neq 0$. Instead it was obtained by modifying the test statistic and the decision boundaries of SATP I. Therefore, the validity of the SATP procedures may be examined by comparing the performance of SATP I and III using normal distribution generators with non-zero means and constant variance. The problems of interest for this simulation were: (1) what are the behaviours of SATP I under different magnitudes of estimation bias; and (2) how reliable is SATP III when applied to classify the applicability of a model after bias-correction.

To carry out this simulation, five levels were set for the means of the normal generators (i.e., 0.1, 0.3, 0.5, 1.0 and 2.0). These values were chosen to cover the absolute values of the mean of volume estimation errors found in B.C. tree data and selected models (i.e., from -2.1 to 0.15). The variance of each generator was set to be equal to the null parameter, $\sigma_0^2$, for each simulated SATP. Using these selected normal generators, the simulation procedures in Section 4.2.1 were performed to estimate the OC values. However, to limit the work of this simulation, only one level of error probabilities was used (i.e., $\alpha = \beta = 0.05$), and the OC values when $\sigma_0^2$ was the true parameter (i.e., $OC_{\sigma_0^2}$) were estimated only. For each simulation, 1,000 iterations were used based on the results of the sensitivity analysis described earlier.

The effect of the bias of model estimation on the testing decision of SATP I was substantial, especially when the two specified limits of error variance were close together (Table 4.8). For these closer variance limits, the estimates of the OC values approached zeros, which means that the probability of rejecting a model is nearly one even when the hypothesized acceptance variance $\sigma_0^2$ is known to be equal to the true variance of the generated errors. As the introduced bias (the mean of the normal generators) increased,
Table 4.8: Comparison of Monte Carlo OC values when $\sigma^2 = \sigma_0^2$ between SATP I and III for testing $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ with $\alpha = \beta = 0.05$ and various $\sigma_0^2$ and $\sigma_1^2$ using normal generators with a non-zero mean.

<table>
<thead>
<tr>
<th>$\sigma_0^2$</th>
<th>$\sigma_1^2/\sigma_0^2$</th>
<th>$N(0.1,\sigma_0^2)$</th>
<th>$N(0.3,\sigma_0^2)$</th>
<th>SATP I $N(0.5,\sigma_0^2)$</th>
<th>SATP I $N(1.0,\sigma_0^2)$</th>
<th>SATP I $N(2.0,\sigma_0^2)$</th>
<th>SATP III All five generators</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.003</td>
<td>1.15</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.992</td>
</tr>
<tr>
<td>0.030</td>
<td>1.25</td>
<td>0.012</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.978</td>
</tr>
<tr>
<td>0.300</td>
<td>1.35</td>
<td>0.946</td>
<td>0.107</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.969</td>
</tr>
<tr>
<td>3.000</td>
<td>1.45</td>
<td>0.964</td>
<td>0.955</td>
<td>0.878</td>
<td>0.165</td>
<td>0.000</td>
<td>0.962</td>
</tr>
<tr>
<td>20.000</td>
<td>1.55</td>
<td>0.968</td>
<td>0.967</td>
<td>0.964</td>
<td>0.933</td>
<td>0.601</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>1.60</td>
<td>0.967</td>
<td>0.969</td>
<td>0.965</td>
<td>0.938</td>
<td>0.657</td>
<td>0.967</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>0.969</td>
<td>0.966</td>
<td>0.962</td>
<td>0.952</td>
<td>0.819</td>
<td>0.967</td>
</tr>
</tbody>
</table>

Note: The values obtained were based on 1,000 iterations. $\sigma_0^2$ and $\sigma_1^2$ are the specified acceptable and unacceptable variance limits of model errors, respectively; $N(0.1,\sigma_0^2)$ represents the normal generator with mean 0.1 and variance, $\sigma_0^2$.

the effect of bias on the performance of SATP I was greater. This result confirms that the testing decision of SATP I will depend on both the actual variance (precision) and mean (bias) of model errors. The results discussed above seem to support the statement made by Reynolds (1984, p.457) when he discussed Freese's (1960) test I and II for model testing. That is, if the tested model is biased, then the error variance would need to be even smaller than the variance bound ($i.e., \sigma^2 \leq \sigma_1^2/\gamma$) specified for model acceptance. If the bias is large enough, it may not be possible to meet the accuracy requirement no matter how small the error variance.
4.4 Reliability of the SATP When Errors Are Normally Distributed With a Zero Mean and Heterogenous Variances

For the two previous simulations, normally distributed errors with a constant variance were considered, which is consistent with the distributional assumption required to extend Freese's (1960) procedure into a Wald's (1947) SPRT plan. However, models with non-constant (heterogeneous) variances of errors are commonly used in forestry. This situation is particularly true in tree volume estimation since the variation of tree volume increases as tree size (e.g., dbh and height) increases. Although weighted least-squares, generalized least-squares approaches, or appropriate transformations may be applied to solve the problem of heterogeneous variance (Cunia 1964, Greene 1990), the principal purpose in applying the volume models is to estimate tree volume. Models obtained from an ordinary least-squares approach are still unbiased even with heterogenous variance. Therefore, the validation of the SATP procedure should be extended for the model errors with heterogenous variance.

4.4.1 Simulation Methods

To simulate the model errors with heterogenous variances, a simple practical situation for volume model testing was considered. This was that the variances of model errors increase with tree dbh class or dbh class groups. Also, the errors from each dbh class or group can be assumed to follow a normal distribution with a zero mean and a constant variance, \( \sigma^2_{(i)} \) for \( i = 1 \ldots k \), where \( k \) is the number of the dbh classes or groups in the population. In order to carry out this simulation, three problems should be addressed. First, the variance pattern of the model errors (the mathematical relationship between the error variances and tree dbh) must be specified. Second, the mathematical model (distribution) used for generating the error population must be selected. Finally, a method to evaluate
the reliability of the SATP procedures for this simulated condition must be selected.

To address the first problem, assuming that the error variances of a volume model increases as tree dbh increases, the variance estimation model suggested by Cunia (1964) was used. This is:

\[ \sigma^2_{(i)} = \alpha_0 \text{dbh}_i^{\alpha_1} \]  

where \(\sigma^2_{(i)}\) is the variance for the model errors of the \(i^{th}\) dbh class; and \(\alpha_0\) is an unknown constant for given volume model. In the simulation, \(\alpha_0\) was arbitrarily set to be 0.005 since the minimum value of MSD found for the B.C. data was 0.0049; and \(\alpha_1\) is the power coefficient. In this simulation, two patterns of variances were considered, \(\sigma^2_{(i)} \propto \text{dbh}_i\) and \(\sigma^2_{(i)} \propto \text{dbh}_i^2\) (i.e., \(\alpha_1 = 1\) and 2, respectively). Also, since the range of dbh of the tree sectional data was found to be from 3.8 to 216.4 cm (most trees were between 3.8 and 100 cm, especially in the lodgepole pine and aspen sets), to reduce the range of simulation, dbh was transformed by dividing by 10. This gave two variance models:

1. Model 1: \(\sigma^2_{(i)} = 0.05x_i\), and \(x_i = 2, 4, 6, 8, 10\).

2. Model 2: \(\sigma^2_{(i)} = 0.5x_i^2\), and \(x_i = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\).

where \(x_i = \text{dbh}_i/10\), and the numbers, 1, 2, \ldots 10, cover a dbh range from 10 to 100 cm with a 10 or 20 cm class interval.

Second, to develop the simulated error populations with a variance pattern following Model 1 or 2 above, a combination (mixture) of normal generators was used. These were:

1. To simulate Model 1, five normal generators with zero means and variances of 0.1, 0.2, 0.3, 0.4 or 0.5 were linked using a uniformly distributed variable with values from 1 to 5. At each stage of sampling, the uniform generator was first used to generate a number. If the generated number was 1, the normal generator with zero mean and variance 0.1 was used to simulate a random observation for the current
stage of sequential sampling. If the value of the uniform generator was 2, the normal generator with zero mean and variance 0.2 was used for sampling, etc. This simulated population will be denoted as \( d_{m1} \sim N(0, 0.05x_i) \) for \( x_i = 2, 4, 6, 8, 10 \).

2. To simulate Model 2, ten normal generators with zero means and variances of 0.5, 2.0, 4.5, 8.0, 12.5, 18.0, 24.5, 32.0, 40.5 or 50.0 were linked using a uniformly distributed variable with values from 1 to 10 using the same method used to generate \( d_{m1} \). This simulated population will be denoted as \( d_{m2} \sim N(0, 0.5x_i^2) \) for \( x_i = 1, 2, \ldots, 10 \).

The probability distribution of these two simulated error populations of tree volume models with heterogeneous variances may be defined statistically as a mixture of a finite number \( (k) \) of normal distributions (Johnson and Kotz 1970, p.87). The general probability density function of a mixture of normal distributions is:

\[
\sum_{i=1}^{k} w_i (\sqrt{2\pi}\sigma_{(i)})^{-1} \exp\left[-\frac{1}{2}(x - \mu_i)^2/\sigma_{(i)}^2\right]
\]  \hspace{1cm} (4.75)

where \( w_i \) are the weights for \( i = 1 \ldots k \) (0 < \( w_i \), \( \sum_{i=1}^{k} w_i = 1 \)); \( \mu_i \) and \( \sigma_{(i)}^2 \) are the mean and variance of the \( i^{th} \) normal distributions, respectively. In this simulation, the \( \mu_i \) were zeros for all \( i \), and the \( w_i \) were equal for all \( i \) (i.e., the \( k \) normal distributions were mixed by a weight coefficient of \( \frac{1}{6} \) or \( \frac{1}{10} \)). Eisenberger (1964) studied the statistical properties of the mixture of two normal distributions, \( N(\mu_1, \sigma_{(1)}^2) \) and \( N(\mu_2, \sigma_{(2)}^2) \) with a weight coefficient \( p \) (0 < \( p < 1 \)) from the first, and \( (1-p) \) from the second normal distribution. From the general form given, this mixed normal variable \( x \) has a density function:

\[
f(x, p) = pN(\mu_1, \sigma_{(1)}^2) + (1-p)N(\mu_2, \sigma_{(2)}^2)
\]

Based on his study, Eisenberger concluded for fixed values of variances \( \sigma_{(1)}^2 \) and \( \sigma_{(2)}^2 \), that the distribution of the mixture will be unimodal, and independent of the weights \( p \) and
1 - \( p \) if the difference between the means is sufficiently small. If the difference between two means exceeds a critical value given by:

\[
(\mu_2 - \mu_1)^2 > \frac{8\sigma^2_{(1)}\sigma^2_{(2)}}{\sigma^2_{(1)} + \sigma^2_{(2)}}
\]

the distribution of \( z \) may be unimodal or bimodal, depending on the value of \( p \). However, if \( p \) is sufficiently close to zero or one, the distribution will be always unimodal. Johnson and Kotz (1970) gave the formula of the 1st to 5th moment about zero of the mixture of two normal distributions. For a random variable, \( x \), with the probability density function of Equation 4.75 and zero means (i.e., \( \mu_i = 0 \) for all \( i \)), the mathematical expectation and variance of \( x \) are:

\[
E(x) = 0; \text{ and }\]

\[
Var(x) = \sum_{i=1}^{k} w_i \sigma_i^2
\]

For a model with heterogeneous error variances as generated using the mixed normal distributions described above, the validity of applying the SATP procedures is not clear, because the translation between the accuracy requirement specified as \( e \) and \( \gamma \) and the variance bound, \( \sigma^2_h = e^2/z_{1-\gamma/2}^2 \), is not valid. However, the problem of interest is whether the SATP procedures can correctly classify a model as acceptable with a probability of \( 1 - \alpha \) when at least \( 100(1 - \gamma) \) percent of the model errors are within \( e \) units. In other words, knowing that the accuracy requirement, \( P(|d| \leq e) \geq 1 - \gamma \), is true for a model, will the SATP procedures reach a decision to accept the model with a probability of \( 1 - \alpha \) when model errors follow a mixture of the normal distributions with zero means and different variances? To test this question, the true values \( e \) for different \( \gamma \) levels had to be determined for each of the two generated mixed normal populations. For this purpose, 10 samples with a large sample size of 10,000 were taken from \( d_{m1} \) or \( d_{m2} \). For each of the 10 samples, the \( e \) value such that \( 100(1 - \gamma) \) percent of the observations were
within \( e \) units was calculated. The average of the \( e \) values over the 10 samples was then calculated for each \( \gamma \) level and population. In this way, the error limits, \( e \), corresponding to the \( \gamma \) levels of 0.30, 0.20, 0.10 and 0.05 were determined. These values were:

1. For \( d_{m1} \), \( e = 0.5674, 0.7022, 0.9010 \) and 1.0735; and

2. For \( d_{m2} \), \( e = 4.5454, 5.6248, 7.2174 \) and 8.5995.

The calculated values \( e \) were then used to set the \( e_0 \) values, and the acceptable variance limit \( \sigma_0^2 \) was determined as \( e_0^2/z_{1-\gamma/2}^2 \) for calibrating a simulated SATP I or III. In this way, the error population generated should be accepted with a probability of \( 1 - \alpha \), if the approximation (i.e., \( \sigma_0^2 = e_0^2/z_{1-\gamma/2}^2 \)) can be used. Simulations were then conducted to estimate the probability of accepting \( H_0 \) as a standard to evaluate the reliability of the SATP procedures.

The procedures for examining SATP I using the generated mixed normal distributed populations were:

1. For each of four levels of \( e_0 \) values, three levels were chosen for the unacceptable error limit, \( e_1 \), specified as a ratio of \( e_1 \) to \( e_0 \) (i.e., \( e_1/e_0 = 1.1, 1.2 \) etc.);

2. Three levels of error probabilities were chosen. These were: \( \alpha = \beta = 0.05, 0.10, \) and 0.20;

3. The values for each combination of \( e_0, e_1, \alpha = \beta \) and the corresponding value of \( \gamma \) were first used to determine the decision boundaries of SATP I using Equations 3.55, 3.56 and 3.57. Each combination of \( e_0, e_1, \gamma, \) and \( \alpha = \beta \) represented a simulated SATP I for testing the hypothesis: \( H_0 : \sigma^2 = e_0^2/z_{1-\gamma/2}^2 \) vs. \( H_1 : \sigma^2 = e_1^2/z_{1-\gamma/2}^2 \);

4. Using the generated population, \( d_{m1} \), random observations were taken, with one observation at each stage of sampling. The observed test statistic, \( T_1 = \sum_{i=1}^{n} d_i^2 \),
was then compared to the decision boundaries, \( (h_1 + sn) \) and \( (h_2 + sn) \), where \( d_i \) is the sampled value at the \( i^{th} \) stage of sampling and \( n \) is the cumulative number of sampled values up to the current stage of sampling. Based on the decision rules outlined in Section 3.1.2, the decision of accepting or rejecting \( H_0 \) was made, or an additional observation was taken;

5. Step 4 was repeated until a terminal decision was finally reached;

6. For a simulated test determined in Step 3, Steps 4 and 5 were repeated 1,000 times. The percent of times that \( H_0 \) was accepted (i.e., the estimate of the \( OC \) value), and associated \( ASN \) value were calculated; and

7. Steps 3 to 6 were repeated for all combinations of \( e_0, e_1, \gamma \) and \( \alpha = \beta \).

The same procedures were repeated using the simulated population, \( d_{m2} \).

### 4.4.2 Simulation Results and Discussions

For the generated error population, \( d_{m1} \), it was found that all of the resulting probabilities of accepting \( H_0 \) (i.e., \( OC \) values) were equal to or larger than \( 1 - \alpha \) for all levels of \( \alpha = \beta \) (Table 4.9). For \( \alpha = \beta = 0.05 \), the resulting \( OC \) values ranged from 0.951 to 0.975. For \( \alpha = \beta = 0.10 \), the resulting \( OC \) values ranged from 0.900 to 0.964, and for \( \alpha = \beta = 0.20 \), the resulting \( OC \) values ranged from 0.820 to 0.927. Similar results were also found for \( d_{m2} \) (Table 4.10).

Since the values, \( e \) and \( \gamma \), used to set up the null hypothesis were known to exactly meet the accuracy requirement represented as \( P(|d_{mi}| \leq e) = 1 - \gamma \), the results of this simulation confirmed that the use of the SATP I procedure resulted in the acceptance of the generated error populations (models) with a probability of at least \( 1 - \alpha \). In other words, the error populations with the heterogeneous variances addressed in this
simulation did not appear to adversely affect the performance of the SATP procedures. Also, the resulting ASN values were similar to those found in Table 4.7. For an expected sample size of less than 30 observations, in most cases, the ratio of $e_1$ to $e_0$ should be larger than 2.00 for $\alpha = \beta = 0.05$ or 0.10. For $\alpha = \beta = 0.20$, the $ASN$ was less than 30 for $e_1/e_0$ ratios greater than 1.25.

It should be indicated that although only two patterns (models) of error variances were considered in this simulation study (i.e., $\sigma^2 \propto \text{dbh}_i$ and $\sigma^2 \propto \text{dbh}_i^2$), they represent the most common models of error variances in tree volume estimation (Cunia 1964). Therefore, the results provide a useful indication for application of the SATP procedures to testing tree volume models with heterogeneous variances of errors. However, the results obtained in this simulation depend on two critical assumptions. These are: (1) the errors of a model can be divided into some mutually exclusive subgroups. The errors in each of these subgroups can be assumed to be iid normal variables, and (2) the observations are taken uniformly from every subgroup (i.e., $k$ normal distributions of errors are mixed with equal weights.).

4.5 Accuracy of the Approximate OC and ASN Equations of the SATP

As indicated in Chapter 2, the OC and ASN functions are the two most important properties of Wald's (1947) SPRT plans, because these functions can be applied to describe the probabilities of making correct testing decisions and the expected sample size of an SPRT over all possible values of the parameter of interest. In practical applications of Wald's SPRT plans, these functions will assist the users to obtain an appropriate SPRT plan for a given problem. In Chapter 3, OC and ASN equations based on Wald's general functions were suggested as approximations of the unknown OC and ASN functions of the SATP procedures (i.e., Equations 3.67, 3.68 and 3.69). These equations were originally
Table 4.9: Monte Carlo estimated probabilities and associated average sample size for accepting $H_0$ when the SATP I is applied to test $H_0 : \sigma^2 = e_0^2/\gamma^{2}z_{1-\gamma/2}$ vs. $H_1 : \sigma^2 = e_1^2/\gamma^{2}z_{1-\gamma/2}$ with $e_0$ and $\gamma$ determined from the mixed normal generated population, $N(0, 0.05x_i)$.

<table>
<thead>
<tr>
<th>$e_0$</th>
<th>$e_1/e_0$</th>
<th>$\gamma$</th>
<th>$\alpha = \beta = 0.05$</th>
<th>$\alpha = \beta = 0.10$</th>
<th>$\alpha = \beta = 0.20$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{OC}$</td>
<td>$\hat{ASN}$</td>
<td>$\hat{OC}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>($SE$)</td>
<td>($SE$)</td>
<td>($SE$)</td>
</tr>
<tr>
<td>1.15</td>
<td>1.35</td>
<td>0.30</td>
<td>0.954 (4.370)</td>
<td>0.910 (2.912)</td>
<td>0.820 (1.350)</td>
</tr>
<tr>
<td>0.5674</td>
<td>1.35</td>
<td>0.30</td>
<td>0.957 (0.842)</td>
<td>0.906 (0.552)</td>
<td>0.858 (0.30)</td>
</tr>
<tr>
<td>2.00</td>
<td>1.35</td>
<td>0.30</td>
<td>0.973 (0.160)</td>
<td>0.961 (0.125)</td>
<td>0.923 (0.072)</td>
</tr>
<tr>
<td>1.20</td>
<td>1.40</td>
<td>0.20</td>
<td>0.956 (1.918)</td>
<td>0.902 (1.389)</td>
<td>0.825 (0.758)</td>
</tr>
<tr>
<td>0.7022</td>
<td>1.40</td>
<td>0.20</td>
<td>0.954 (0.597)</td>
<td>0.929 (0.474)</td>
<td>0.880 (0.249)</td>
</tr>
<tr>
<td>2.00</td>
<td>1.40</td>
<td>0.20</td>
<td>0.975 (0.158)</td>
<td>0.964 (0.120)</td>
<td>0.927 (0.070)</td>
</tr>
<tr>
<td>1.25</td>
<td>1.45</td>
<td>0.10</td>
<td>0.951 (1.261)</td>
<td>0.900 (0.997)</td>
<td>0.842 (0.528)</td>
</tr>
<tr>
<td>0.9010</td>
<td>1.45</td>
<td>0.10</td>
<td>0.954 (0.523)</td>
<td>0.929 (0.400)</td>
<td>0.887 (0.221)</td>
</tr>
<tr>
<td>2.00</td>
<td>1.45</td>
<td>0.10</td>
<td>0.975 (0.158)</td>
<td>0.964 (0.120)</td>
<td>0.924 (0.070)</td>
</tr>
<tr>
<td>1.30</td>
<td>1.50</td>
<td>0.05</td>
<td>0.952 (1.027)</td>
<td>0.909 (0.741)</td>
<td>0.851 (0.425)</td>
</tr>
<tr>
<td>1.0735</td>
<td>1.50</td>
<td>0.05</td>
<td>0.957 (0.445)</td>
<td>0.931 (0.327)</td>
<td>0.890 (0.183)</td>
</tr>
<tr>
<td>2.00</td>
<td>1.50</td>
<td>0.05</td>
<td>0.974 (0.159)</td>
<td>0.962 (0.121)</td>
<td>0.923 (0.071)</td>
</tr>
</tbody>
</table>

Note: The values are based on 1,000 iterations. $N(0, 0.05x_i)$ represents the simulated mixture of 5 normal distributions with all zero means and variances, 0.1, 0.2, 0.3, 0.4 and 0.5, and the five normal distributions were mixed in a equal weight, $\frac{1}{5}$; $e_0$ and $e_1$ are the specified acceptable and unacceptable error limits; $d_{m1}$ represents the mixed normal generator, $N(0, 0.05x_i)$; $\gamma$ is the actual accuracy level associated with $e_0$; both $e_0$ and $\gamma$ were determined from the generated population based on the 10 samples of size, 10,000; $\hat{OC}$ is the Monte Carlo estimated probability of accepting $H_0$ when the sequential test I is simulated; $\hat{ASN}$ is the average sample number associated with $\hat{OC}$; $SE$ is the standard error of the estimated $\hat{ASN}$.  

Table 4.10: Monte Carlo estimated probabilities and associated average sample size for accepting $H_0$ when the SATP I is applied to test $H_0: \sigma^2 = e_0^2/z_{1-\gamma/2}^2$ vs. $H_1: \sigma^2 = e_1^2/z_{1-\gamma/2}^2$ with $e_0$ and $\gamma$ determined from the mixed normal generated population, $N(0, 0.5z_2^2)$.

<table>
<thead>
<tr>
<th>$e_0$</th>
<th>$e_1/e_0$</th>
<th>$\gamma$</th>
<th>$\alpha = \beta =$</th>
<th>$\bar{OC}$</th>
<th>$\bar{ASN}$</th>
<th>$\bar{OC}$</th>
<th>$\bar{ASN}$</th>
<th>$\bar{OC}$</th>
<th>$\bar{ASN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td>(SE)</td>
<td>(SE)</td>
<td>0.10</td>
<td>(SE)</td>
<td>0.20</td>
<td>(SE)</td>
</tr>
<tr>
<td>1.15</td>
<td>4.5454</td>
<td>0.30</td>
<td>0.978 (3.262)</td>
<td>121.9 (2.426)</td>
<td>0.950 (1.590)</td>
<td>86.3 (1.590)</td>
<td>0.894 (1.371)</td>
<td>47.0 (1.371)</td>
<td>13.8 (1.371)</td>
</tr>
<tr>
<td>2.00</td>
<td>5.6248</td>
<td>0.20</td>
<td>0.984 (0.924)</td>
<td>9.4 (2.208)</td>
<td>0.972 (0.773)</td>
<td>9.0 (2.08)</td>
<td>0.940 (0.150)</td>
<td>4.5 (0.095)</td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>7.2174</td>
<td>0.10</td>
<td>0.984 (1.831)</td>
<td>72.5 (1.450)</td>
<td>0.962 (0.900)</td>
<td>52.4 (1.450)</td>
<td>0.926 (0.349)</td>
<td>28.8 (0.093)</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>8.5995</td>
<td>0.05</td>
<td>0.986 (0.704)</td>
<td>9.3 (0.201)</td>
<td>0.974 (0.585)</td>
<td>6.8 (0.138)</td>
<td>0.948 (0.093)</td>
<td>4.5 (0.093)</td>
<td></td>
</tr>
<tr>
<td>1.30</td>
<td>8.5995</td>
<td>0.05</td>
<td>0.986 (0.568)</td>
<td>9.2 (0.195)</td>
<td>0.974 (0.585)</td>
<td>6.7 (0.138)</td>
<td>0.952 (0.093)</td>
<td>4.5 (0.093)</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td></td>
<td></td>
<td>0.982 (0.525)</td>
<td>9.1 (0.525)</td>
<td>0.978 (0.374)</td>
<td>6.7 (0.138)</td>
<td>0.954 (0.093)</td>
<td>4.5 (0.093)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The values are based on 1,000 iterations. $N(0, 0.5z_2^2)$ represents the simulated mixture of 10 normal distributions with all zero means and variances, 0.5, 2.0, 4.5, 8.0, 12.5, 18.0, 24.5, 32.0, 40.5, 50.0 and the normal distributions were mixed in an equal weight, $\frac{1}{10}$; $e_0$ and $e_1$ are the specified acceptable and unacceptable error limits; $d_m$ represents the mixed normal generator, $N(0, 0.5z_2^2)$; $\gamma$ is the actual accuracy level associated with $e_0$; both $e_0$ and $\gamma$ were determined from the generated population based on the 10 samples of size, 10,000; $\bar{OC}$ is the Monte Carlo estimated probability of accepting $H_0$ when the sequential test I is simulated; $\bar{ASN}$ is the average sample number associated with $\bar{OC}$; SE is the standard error of the estimated ASN.
derived by ignoring possible overshooting of the decision boundaries; their accuracy in representing the unknown OC and ASN functions of the SATP procedure is unknown.

Fowler (1978, 1983) investigated the accuracy of Wald’s OC and ASN functions for SPRT plans of binomial, negative binomial, normal and Poisson distributions used in forest pest management. He concluded that Wald’s OC functions underestimated the actual OC function near the lower class limit and overestimated it near the upper limit. Also, Wald’s ASN functions underestimated the actual ASN function with the maximum relative error being near the maximum ASN value. The practical consequences of these errors are: (1) the actual error probabilities can be smaller than the nominal probabilities of errors used to build the sampling plan; and (2) more observations are usually taken in the field than required to meet the desired probabilities of errors (i.e., $\alpha$ and $\beta$) (Fowler 1978, p.12). However, all SPRT procedures studied by Fowler were for testing population means, not for testing population variances. Also, the parameters used in the simulation of his study were adopted from pest surveys; they may not be appropriate for model testing. Therefore, it was necessary to investigate the accuracy of the approximate OC and ASN equations suggested before applying them in practical situations.

### 4.5.1 Simulation methods

Since an OC and an ASN equation can be developed for any given SPRT specified by a combination of the values of four parameters ($\sigma_0^2$, $\sigma_1^2$, $\alpha$ and $\beta$), it is impossible to examine the accuracy of the suggested OC and ASN equations for all possible combinations of these parameters in the range found for the volume models selected. Only two sequential accuracy testing plans were selected to provide some indications of the OC and ASN equations suggested for SATP procedures. These two testing plans selected were:

1. Plan 1: $\sigma_0^2 = 0.3$, $\sigma_1^2 = 0.8$, $\gamma = \alpha = \beta = 0.05$; and
2. Plan 2: $\sigma_0^2 = 1.50, \sigma_1^2 = 2.25, \gamma = \alpha = 0.05, \beta = 0.10.$

The considerations in selecting these plans were to have the tested variance within the middle of the range of the actual error variance found in the B.C. tree data and models selected (i.e., $MSD$ from 0.003 to 19.9 in Table 4.2), and to have the levels for the type I and II errors be equal and unequal.

To evaluate the accuracy of the approximate OC and ASN equations of the SATP procedure, the actual unknown OC and ASN function for a given testing plan must be estimated using Monte Carlo techniques. The estimated equations are then used as a standard for comparing with the OC and ASN values determined by the approximate equations. The procedures used to construct the Monte Carlo OC and ASN equations of the Plan 1 for SATP I were:

1. The specified values of $\sigma_0^2 = 0.3, \sigma_1^2 = 0.8, \gamma = \alpha = \beta = 0.05$ were first used to determine decision boundaries for SATP I using Equations 3.55, 3.56 and 3.57;

2. A set of values which are uniformly distributed across, and extending beyond the acceptance and unacceptance limits ($\sigma_0^2$ and $\sigma_1^2$) was obtained by setting the values of $h$ in Equation 3.68 to be from -4.0 to 4.0 with an interval of 0.5 ($h \neq 0$). For each $h$ value, Equation 3.68 was then used to calculate $\sigma^2$. The calculated values represented various possible values of the true variance of errors around the tested variances. These were the data points for the horizontal axis when constructing the OC or ASN curves. These calculated values will be denoted as $\sigma_i^2$ for $i = 1 \ldots 18$;

3. The normal generator, $N(0, \sigma_{H_1}^2)$, was used to simulate taking random observations from the error population of a volume model, with one observation taken at each stage of sampling. SATP I was carried out to test the hypothesis $H_0 : \sigma^2 = 0.3$ vs. $H_1 : \sigma^2 = 0.8$ until a terminal decision was reached, where $\sigma_{H_1}^2$ was the first
calculated value in Step 2;

4. Step 3 was repeated 2,000 times. The percent of times that \( H_0 \) was accepted and the average of the number of observations taken were calculated, and used as the estimates of the actual OC and ASN values for a given \( \sigma^2 = \sigma^2_1 \);

5. By changing the variance to \( \sigma^2_2, \sigma^2_3, \text{ etc.} \), Steps 3 and 4 were repeated for all 18 values calculated in step 2. These 18 pairs of Monte Carlo OC and ASN values are a function of the calculated values, \( \sigma^2_1 \). Together, they were used to represent the actual OC and ASN equations of Plan 1 for the SATP I; and

6. By changing the specified values to those of Plan 2, all steps above were repeated. The Monte Carlo OC and ASN equations of Plan 2 for SATP I were obtained.

Monte Carlo OC and ASN values of these two selected plans were also obtained for SATP III using the same procedures. Finally, the approximate OC and ASN values were also obtained from the suggested OC and ASN equations (i.e., Equations 3.67 3.68 and 3.69) for the same set of the calculated \( \sigma^2_1 \) values by following the procedure described in Section 3.4. The OC and ASN values determined by the approximate equations are denoted as \( \overline{OC} \) and \( \overline{ASN} \), those obtained from simulations are denoted as \( OC_m \) and \( ASN_m \).

4.5.2 Comparisons between the approximate and Monte Carlo OC and ASN equations

The suggested OC equation approximates the actual OC function well for both SATP I and III, in general (Table 4.11 and Figure 4.3). However, near the acceptable variance limit (\( \sigma^2 \)) and between the two limits, the suggested OC equation underestimated the actual OC function of the SATP procedures. The maximum difference between the Monte
Chapter 4. Monte Carlo Simulation

Monte Carlo estimated OC values and those determined from the OC equation was -0.067 for SATP I and -0.068 for SATP III. The OC equation suggested appears to be appropriate for approximating the unknown OC function of the SATP procedure.

The suggested ASN equation consistently underestimated the actual ASN function for both SATP I and III (Table 4.12 and Figure 4.4). The differences of the ASN values between Monte Carlo estimates and those determined from the equation ranged from -0.6 to -19.0 for SATP I and from -1.6 to -17.0 for SATP III. The maximum difference occurred near $\sigma^2$. This result suggests that the actual sample size required by the SATP procedure will be larger than the values calculated from the suggested ASN equation. However, the suggested ASN equation could still be used to provide some indication of the expected sample size. Such an indication is very helpful in assisting potential users to obtain the appropriate SATP plan for a given problem.

The Monte Carlo ASN values for SATP I and III showed no large differences. Since the distributional assumptions of SATP I and II are exactly the same, the suggested OC and ASN equation can be used to approximate the actual OC and ASN functions for the three developed sequential testing plans.

4.6 Effects of Truncation and Group Sampling on the Performance of the SATP

As described in Section 2.4.1.1, the decision boundaries of Wald’s (1947) SPRT plans were derived based on two principal assumptions. One is that the random observations were taken from a population singly at each stage of sampling, and the other was that no upper limit to the number of observations was set. These assumptions were retained in extending Freese’s (1960) procedure to obtain the SATP procedures. However, these
Table 4.11: Comparisons of the Monte Carlo OC and ASN curves of the SATP I and III and the approximate equations with $\sigma^2 = 0.3, \sigma^2 = 0.8, \alpha = \beta = 0.05$.

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>Approximate Equations</th>
<th>SATP I</th>
<th>SATP III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$OC$ $ASN$</td>
<td>$OC_m$ $SE_{OC_m}$ $ASN_m$ $SE_{ASN_m}$</td>
<td>$OC_m$ $SE_{OC_m}$ $ASN_m$ $SE_{ASN_m}$</td>
</tr>
<tr>
<td>0.1176</td>
<td>1.000 8.0</td>
<td>1.000 0.000 8.6 0.029</td>
<td>1.000 0.000 9.6 0.031</td>
</tr>
<tr>
<td>0.1327</td>
<td>1.000 8.4</td>
<td>1.000 0.000 9.0 0.039</td>
<td>1.000 0.000 9.9 0.038</td>
</tr>
<tr>
<td>0.1516</td>
<td>1.000 8.9</td>
<td>1.000 0.000 9.5 0.045</td>
<td>1.000 0.000 10.5 0.047</td>
</tr>
<tr>
<td>0.1755</td>
<td>1.000 9.6</td>
<td>1.000 0.000 10.2 0.067</td>
<td>1.000 0.000 11.2 0.062</td>
</tr>
<tr>
<td>0.2062</td>
<td>1.000 10.6</td>
<td>0.999 0.001 11.4 0.085</td>
<td>0.999 0.001 12.3 0.083</td>
</tr>
<tr>
<td>0.2465</td>
<td>0.990 12.3</td>
<td>0.994 0.002 13.3 0.119</td>
<td>0.995 0.002 14.2 0.116</td>
</tr>
<tr>
<td>0.3000</td>
<td>0.950 14.9</td>
<td>0.975 0.003 16.8 0.196</td>
<td>0.975 0.003 17.7 0.194</td>
</tr>
<tr>
<td>0.3721</td>
<td>0.810 18.0</td>
<td>0.888 0.007 23.2 0.330</td>
<td>0.887 0.007 23.8 0.328</td>
</tr>
<tr>
<td>0.4175</td>
<td>0.680 18.7</td>
<td>0.754 0.010 26.3 0.399</td>
<td>0.754 0.010 26.8 0.396</td>
</tr>
<tr>
<td>0.5335</td>
<td>0.320 15.9</td>
<td>0.322 0.010 26.1 0.412</td>
<td>0.336 0.011 27.8 0.414</td>
</tr>
<tr>
<td>0.6077</td>
<td>0.190 12.9</td>
<td>0.175 0.008 22.5 0.350</td>
<td>0.186 0.009 25.1 0.353</td>
</tr>
<tr>
<td>0.8000</td>
<td>0.050 7.7</td>
<td>0.041 0.004 18.1 0.254</td>
<td>0.046 0.005 20.5 0.261</td>
</tr>
<tr>
<td>1.0735</td>
<td>0.010 4.6</td>
<td>0.007 0.002 14.1 0.110</td>
<td>0.007 0.002 13.9 0.112</td>
</tr>
<tr>
<td>1.4667</td>
<td>0.000 2.8</td>
<td>0.001 0.001 11.0 0.078</td>
<td>0.002 0.001 11.5 0.076</td>
</tr>
<tr>
<td>2.0376</td>
<td>0.000 1.8</td>
<td>0.000 0.000 3.8 0.035</td>
<td>0.001 0.001 9.0 0.032</td>
</tr>
<tr>
<td>2.8741</td>
<td>0.000 1.2</td>
<td>0.000 0.000 2.9 0.048</td>
<td>0.000 0.000 3.9 0.046</td>
</tr>
<tr>
<td>4.1087</td>
<td>0.000 0.8</td>
<td>0.000 0.000 2.4 0.032</td>
<td>0.000 0.000 3.4 0.036</td>
</tr>
<tr>
<td>5.9481</td>
<td>0.000 0.5</td>
<td>0.000 0.000 2.0 0.031</td>
<td>0.000 0.000 3.0 0.029</td>
</tr>
</tbody>
</table>

Note: The values were based on 2,000 iterations. $\sigma^2$ represents the possible values for the true error variance of a volume model; $OC$ and $ASN$ are the OC and ASN values determined from the approximate equations; $OC_m$ and $ASN_m$ are Monte Carlo estimated OC and ASN values, respectively; $SE_{OC_m}$ and $SE_{ASN_m}$ are the standard errors of the Monte Carlo estimated OC and ASN, respectively.
Table 4.12: Comparisons of the Monte Carlo OC and ASN curves of the SATP I and III and the approximate equations with $\sigma_0^2 = 1.50, \sigma_1^2 = 2.25, \gamma = \alpha = 0.05$ and $\beta = 0.10$.

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>Approximate Equations</th>
<th>SATP I</th>
<th>SATP III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\overline{OC}$</td>
<td>$\overline{ASN}$</td>
<td>$\overline{OC}_m$</td>
</tr>
<tr>
<td>0.9028</td>
<td>1.000</td>
<td>22.0</td>
<td>1.000</td>
</tr>
<tr>
<td>0.9747</td>
<td>1.000</td>
<td>23.8</td>
<td>1.000</td>
</tr>
<tr>
<td>1.0556</td>
<td>1.000</td>
<td>26.3</td>
<td>1.000</td>
</tr>
<tr>
<td>1.1468</td>
<td>1.000</td>
<td>29.8</td>
<td>0.999</td>
</tr>
<tr>
<td>1.2500</td>
<td>1.000</td>
<td>35.0</td>
<td>1.000</td>
</tr>
<tr>
<td>1.3670</td>
<td>0.990</td>
<td>43.0</td>
<td>0.990</td>
</tr>
<tr>
<td>1.5000</td>
<td>0.950</td>
<td>55.3</td>
<td>0.968</td>
</tr>
<tr>
<td>1.6515</td>
<td>0.830</td>
<td>71.0</td>
<td>0.867</td>
</tr>
<tr>
<td>1.7352</td>
<td>0.710</td>
<td>77.1</td>
<td>0.739</td>
</tr>
<tr>
<td>1.9203</td>
<td>0.410</td>
<td>76.0</td>
<td>0.433</td>
</tr>
<tr>
<td>2.0227</td>
<td>0.270</td>
<td>68.6</td>
<td>0.278</td>
</tr>
<tr>
<td>2.2500</td>
<td>0.100</td>
<td>50.3</td>
<td>0.093</td>
</tr>
<tr>
<td>2.5114</td>
<td>0.030</td>
<td>35.6</td>
<td>0.026</td>
</tr>
<tr>
<td>2.8125</td>
<td>0.010</td>
<td>25.8</td>
<td>0.012</td>
</tr>
<tr>
<td>3.1602</td>
<td>0.000</td>
<td>19.4</td>
<td>0.006</td>
</tr>
<tr>
<td>3.5625</td>
<td>0.000</td>
<td>14.9</td>
<td>0.003</td>
</tr>
<tr>
<td>4.0288</td>
<td>0.000</td>
<td>11.8</td>
<td>0.000</td>
</tr>
<tr>
<td>4.5703</td>
<td>0.000</td>
<td>9.50</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: The values were based on 2,000 iterations. $\sigma^2$ represents the possible values for the true error variance of a volume model; $\overline{OC}$ and $\overline{ASN}$ are the OC and ASN values determined from the approximate equations; $\overline{OC}_m$ and $\overline{ASN}_m$ are Monte Carlo estimated OC and ASN values, respectively; $SE_{\overline{OC}_m}$ and $SE_{\overline{ASN}_m}$ are the standard errors of the Monte Carlo estimated OC and ASN, respectively.
Figure 4.3: Comparison of the Monte Carlo OC and ASN curves of the SATP I and III and the Wald’s approximated equations with $\sigma_0^2 = 0.3, \sigma_1^2 = 0.8, \alpha = \beta = 0.05$. (a) Comparison of OC curves; (b) Comparison of ASN curves.
Figure 4.4: Comparison of the Monte Carlo OC and ASN curves of the SATP I and III and Wald's approximated equations with $\sigma_0^2 = 1.50, \sigma_1^2 = 2.25, \alpha = 0.05$ and $\beta = 0.10$. (a) Comparison of OC curves; (b) Comparison of ASN curves.
assumptions may not be practical for the purpose of model testing. Taking random observations singly will increase the travel times between the sampling units, and this could greatly degrade the expected cost-saving advantage of the SATP procedures. Also, if the data obtained include field measurement and lab analysis, single observation selection is not practical for most applications. In addition, if no upper limit is set for the number of observations sampled, the SATP procedures may result in a large sample size that is even larger than that considered large enough for an equally reliable fixed sample size procedure.

Methods suggested by Wald (1947) to solve these problems were described in Section 2.4.1.4. Instead of taking observations singly, a group of $k$ observations may be taken at each stage of sampling (i.e., group selection). Also, a maximum number of observations ($n_0$, called the truncation point) may be preset to stop the sequential testing process when $n_0$ is reached. A final decision to accept or reject the null hypothesis is then made by comparing the observed test statistic, $T_{n_0}$ (e.g., $\sum_{i=1}^{n_0} d_i^2$ when SATP I is applied), with the average of the acceptance and rejection values at this point (i.e., Wald’s rule of truncation). These modifications will affect the derived decision boundaries of the SATP procedure, and therefore, the resulting error probabilities. The reliability of the SATP under these situations is unknown.

To investigate the effect of group selection, one testing plan, $\sigma_0^2 = 1.221, \sigma_1^2 = 2.747$ and $\gamma = \alpha = \beta = 0.05$, was arbitrarily chosen. In choosing this plan, the mean square error associated with Spurr’s (1952) estimated model from the Douglas-fir data set was used for $\sigma_0^2$, and $\sigma_1^2$ was then arbitrarily set to $2.25\sigma_0^2$. Three group sizes were selected as 2, 5 and 10 (i.e., 2, 5 or 10 observations were taken at each stage of simulated sampling). For comparison, a group size of one (observations taken singly) was also used. To study the effect of truncation, $\sigma_0^2 = 1.221$ and $\sigma_1^2 = 2.747$ were used along with three combinations of error probabilities of the type I and II error (i.e., $\alpha = \beta = 0.05$, $\alpha = 0.05$ and $\beta = 0.10$,
and $\alpha = 0.10$ and $\beta = 0.05$). These three selected plans had a specified type I error ($\alpha$) that was equal, less than, or greater than the specified type II error ($\beta$). To set the upper limits of the number of observations sampled, $n_0$, the approximate ASN equation (Equation 3.69 or 3.70) was used to determine ASN when $\sigma_0^2$ was the true variance (i.e., $E_{\sigma_0^2}(n)$) for each of three testing plans. Based on this calculated ASN value, four levels for the truncation points ($n_0$) were set as $n_0 = E_{\sigma_0^2}(n)$, $1.5E_{\sigma_0^2}(n)$, $2E_{\sigma_0^2}(n)$ and $3E_{\sigma_0^2}(n)$. For comparison, the case of no truncation (i.e., $n_0 = \infty$) was also used. In order to provide some indication of the effects of truncation and group selection on the performance of the SATP procedure, the Monte Carlo estimates of the actual probabilities of the type I and II error, $\alpha_m$ and $\beta_m$, and associated ASN values were obtained using the procedures described in Section 4.2.1 for the various group sizes and truncation points of the testing plans described above.

As group size increased from 1 to 10, the resulting type I error (i.e., $\alpha_m = 1 - OC_{\sigma_0^2}$) decreased from 0.026 to 0.016, and the resulting type II error ($\beta_m = OC_{\sigma_1^2}$) decreased from 0.047 to 0.025 (Table 4.13). This means that as the number of observations taken at each stage increased, the difference (error) between the actual error probabilities and their nominal values (0.05) increased rapidly. Also, as the group size increased from 1 to 10, the resulting ASN (average sample number) increased from 23.1 to 30.0 when $\sigma_0^2$ was true, and from 23.9 to 34.4 when $\sigma_1^2$ was true. This result is explainable since the decision boundaries of the SATP procedures were obtained by ignoring possible overshooting of the decision boundaries. As the number of observations taken at each stage increases, overshooting of the decision boundaries increases. The consequence is that the resulting error probabilities are smaller than their desired values, and more observations are taken than are necessary.

In general, the resulting error probabilities decreased as the values of the truncation points of the number increased (Table 4.14). For the type I error, a truncation point of
Table 4.13: Comparison of Monte Carlo estimates of $\alpha$, $\beta$ and associated ASN values of SATP I for various group sizes with $\sigma_0^2 = 1.221$, $\sigma_1^2 = 2.747$ and $\alpha = \beta = 0.05$.

<table>
<thead>
<tr>
<th>Monte Carlo Estimate</th>
<th>Group Sampling Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_m$</td>
<td>0.026</td>
</tr>
<tr>
<td>$A\hat{SN}_{\sigma_0^2}$</td>
<td>23.1</td>
</tr>
<tr>
<td>$SE_{A\hat{SN}}$</td>
<td>0.181</td>
</tr>
<tr>
<td>$\beta_m$</td>
<td>0.047</td>
</tr>
<tr>
<td>$A\hat{SN}_{\sigma_1^2}$</td>
<td>23.9</td>
</tr>
<tr>
<td>$SE_{A\hat{SN}}$</td>
<td>0.177</td>
</tr>
</tbody>
</table>

Note: The values were based on 2,000 iterations. $\alpha_m = 1 - OC_{\sigma_0^2}$ and $\beta_m = OC_{\sigma_1^2}$, and $OC_{\sigma_0^2}$ and $OC_{\sigma_1^2}$ are Monte Carlo estimated OC values when $\sigma_0^2$ and $\sigma_1^2$ are true, respectively; $A\hat{SN}_{\sigma_0^2}$ and $A\hat{SN}_{\sigma_1^2}$ are Monte Carlo estimate of average sample number when $\sigma_0^2$ or $\sigma_1^2$ is true; $SE_{A\hat{SN}}$ is the standard error of the estimated ASN.
1.5E_{e^2}(n) or 2E_{e^2}(n) would be enough to obtain a resulting probability (\alpha_m) equal to or less than the desired value (\alpha). However, for type II error, a truncation point of 3E_{e^2}(n) would be required to obtain a resulting probability (\beta_m) equal to or less than the desired value (\beta). Therefore, based on the results of this simulation, 3E_{e^2}(n) appears to be a reasonable truncation point, when Wald's (1947) rule of truncation is used to force the SATP procedure to reach a decision before a terminal decision is reached.

4.7 Conclusion

Five simulation studies were conducted to examine the developed SATP procedures using normal generated populations of the errors of tree volume models. Based on the results of these simulation studies, the following conclusions may be made:

1. If the errors of a tested model are iid normally distributed variables with a zero mean (estimation is unbiased) and a constant variance, \sigma^2, both SATP I and III are reliable for classifying a volume model into two wide classes (i.e., acceptable or unacceptable) based on two preset limits of error variance. Also, the probability of making a correct decision (OC values) of SATP I and III are very similar, but the average sample required (ASN values) of SATP III is slightly larger than that of SATP I (usually, within 5 observations). This occurs since the sample mean of errors must be estimated by SATP III.

2. If the errors, d, are iid normally distributed variables with non-zero means (estimations are biased) and constant variance, the testing decision to accept or reject a model reached by SATP I depends not only on the variance of errors, but also on the magnitude of bias. When the value of the bias is larger than that of the error variance, the probability of rejecting a model is very high even for very small values
Table 4.14: Comparison of Monte Carlo estimates of $\alpha$, $\beta$ and ASN values of SATP I for various truncation points and different combinations of $\alpha$ and $\beta$ with $\sigma_0^2 = 1.221$, $\sigma_1^2 = 2.747$.

<table>
<thead>
<tr>
<th>$(\alpha, \beta)$ Values</th>
<th>Monte Carlo Estimates</th>
<th>Truncation Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\alpha = 0.05, \beta = 0.05)$</td>
<td>$\alpha_m$</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>$\beta_m$</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td>$ASN_{\sigma_0^2}$</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>$ASN_{\sigma_1^2}$</td>
<td>19.7</td>
</tr>
<tr>
<td>$(\alpha = 0.05, \beta = 0.10)$</td>
<td>$\alpha_m$</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>$\beta_m$</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>$ASN_{\sigma_0^2}$</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>$ASN_{\sigma_1^2}$</td>
<td>14.7</td>
</tr>
<tr>
<td>$(\alpha = 0.10, \beta = 0.05)$</td>
<td>$\alpha_m$</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>$\beta_m$</td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td>$ASN_{\sigma_0^2}$</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>$ASN_{\sigma_1^2}$</td>
<td>17.9</td>
</tr>
</tbody>
</table>

Note: The values were based on 2,000 iterations. Truncation point means that if the number of observations taken is equal to this value and no terminal decision can be reached, the average of acceptance and rejection value at this stage is used to make the final decision based on Wald's rule of truncation; $E_{\sigma_0^2}(n)$ is the calculated ASN value when $\sigma_0^2$ is true using Wald's approximated ASN equation; $\infty$ means that no truncation is made; $\alpha_m$ and $\beta_m$ are Monte Carlo estimates of the actual $\alpha$ and $\beta$, respectively; $ASN_{\sigma_0^2}$ and $ASN_{\sigma_1^2}$ are Monte Carlo estimates of ASN when $\sigma_0^2$ and $\sigma_1^2$ are true, respectively.
of error variance. For a given value of bias, the probability of accepting a model obtained by SATP I is always less than or equal to those of SATP III. The probability obtained by SATP III did not change as the values of the introduced bias increased. This confirms that SATP I will reject an inaccurate model regardless of the source of errors (larger bias, lower precision, or both), whereas SATP III will determine the accuracy of a model after bias-correction. In practical applications, the use of SATP III for accuracy testing after bias-correction requires that the biases are the same for all estimated values, otherwise the bias in the estimation may not be removeable (e.g., the bias for the different levels of model's independent variables is not constant, but a random variable).

3. If the distribution of the errors of a volume model can be assumed to be a mixture of a finite number of normal distributions from different dbh classes with zero means and variances being proportional to dbh or \((\text{dbh}^2)\), SATP I and III are still applicable for determining the applicability of volume models when random observations are uniformly taken from every dbh class, even though the assumption of normality of the model errors is violated.

4. The nominal probabilities of type I and II errors \((\alpha \text{ and } \beta)\) used to calibrate an SATP were found to be larger than the resulting error probabilities when the SATP procedures are actually performed. This means that the SATP procedures are conservative, and always result in a higher testing precision with possibly a higher sample size than that required to meet the nominal error probabilities.

5. The suggested ASN equation for the SATP procedures based on Wald's (1947) general ASN function of an SPRT plan always underestimated the actual ASN. However, in general, the suggested OC equation worked well in approximating the actual OC values for the SATP procedures.
6. Under the normality assumption of model errors with a zero mean and a constant variance, when random group selection of observations is used instead of single selection, the resulting probabilities of type I and II errors will decrease with an associated increase in the average sample size required, as the group size increases.

7. If Wald's (1947) rule of truncation is used to force the SATP procedures to obtain a final decision when a preset maximum number of observations sampled is reached, a reasonable truncation point will be three times the expected sample size from the ASN equation.
Chapter 5

Field Application Procedures And Example

In Chapter 3, three sequential accuracy testing plans (SATP) were developed for applicability testing of tree volume models. The simulation results of Chapter 4 showed that the SATP procedures were reliable for classifying a model into two predefined classes (acceptable or unacceptable) under the normality assumption of the error, \( d \), or the relative error, \( d/y \). However, as with Freese's (1960) original procedure, care must be exercised to insure that the unconditional distribution of \( d \) or \( d/y \) is normal when applying the SATP procedures. Also, the appropriate application procedures should be followed in order to obtain the expected advantage of the SATP procedures in sampling cost savings.

This chapter is designed to meet the third objective of this research, that is, to propose and illustrate appropriate field procedures for applying the SATP procedures.

5.1 Field Application Procedures

5.1.1 Preparation

Before an SATP for determining the accuracy of a model can be carried out, information about this model and the application population should be collected. The desired information includes (1) the standard error of estimate \( (SE_E) \) from model building, the results of testing the regression assumptions, especially the normality of errors (residuals); (2) the mean difference \( (\bar{d}) \), the mean absolute difference \( (|\bar{d}|) \) and mean squared difference \( (MSD) \) of the estimated and observed values in the previous applications of the model;
Chapter 5. Application Example

and (3) the ranges of the dependent and independent variables in the application population (e.g., the range of tree dbh or height). This information is required for choosing the appropriate hypothesized error limits for model acceptance and rejection for a given application (e.g., $e_0$ and $e_1$, or $p_0$ and $p_1$), and for setting the desired probability level of model accuracy ($\gamma$). The information above is also helpful for selecting an appropriate group sampling design. The sampling cost constraint, or the calculated sample size for applying a fixed sample size test (e.g., Freese's (1960) procedure) is helpful for setting the truncation point of an SATP. Based on available information, the test parameters can then be specified to calibrate an SATP for a given application.

In order to determine an SATP, five parameters need to be specified. These are: (i) $e_0$, the acceptable error limit (or $p_0$, the acceptable percent error limit if SATP II is used); (ii) $e_1$, the unacceptable error limit (or $p_1$, unacceptable percent error limit if SATP II is used); (iii) the probability level of model accuracy associated with the $e_0$ or $p_0$ specified, $\gamma$; and (iv) the probability levels of the type I and II errors for the test, $\alpha$ and $\beta$. The choice of $e_0$ and $e_1$ (or $p_0$ and $p_1$) is not straightforward; it depends on the user's knowledge about the nature of model errors, and the expected sample size for making a terminal decision (i.e., the wider the interval between these class limits, the smaller the sample sizes required).

Based on the specified values for the parameters, the test hypotheses of an SATP are stated, and the decision boundaries can then be determined.

1. For applying SATP I and III, the test hypotheses are:

$$H_0 : \sigma^2 = \sigma_0^2 = \frac{e_0^2}{z_{1-\gamma/2}}$$

$$H_1 : \sigma^2 = \sigma_1^2 = \frac{e_1^2}{z_{1-\gamma/2}}$$

if SATP I is applied, $\mu_d = 0$ should also be added to the hypotheses above.
2. For applying SATP II, the test hypotheses are:

\[
H_0 : \sigma^2_{\delta/y} = \frac{p_3^2}{(100z_{1-\gamma/2})^2}
\]

\[
H_1 : \sigma^2_{\delta/y} = \frac{p_1^2}{(100z_{1-\gamma/2})^2}
\]

3. To determine the decision lines of an SATP, the intercepts, \( h_1 \) and \( h_2 \), and the common slope \( s \) must be calculated using the specified values of \( e_0, e_1 \) (or \( p_0, p_1 \) when applying SATP II), \( \gamma, \alpha \) and \( \beta \) with the formulae presented in Chapter 3. Based on these calculated intercepts and slope, the acceptance line is \( h_1 + ns \), and the rejection line is \( h_2 + sn \).

For selecting random observations from the population, random group sampling is suggested. To do the group sampling, first, the desired group size, \( k \) (e.g., 5, 10, etc.), should be preset; then \( k \) observations would be randomly taken from the population at each stage of sequential sampling. Under field conditions, \( k \) trees or plots (sampling units) could be taken at randomly selected points in the stand or forest (population). Also, a truncation point \( n_0 \) should be predetermined based on (1) the consideration of maximum sampling cost, or the calculated sample size of an equally reliable fixed-sample procedure; or (2) \( n_0 \geq 3E_{\alpha}(n) \), (i.e., three times the calculated ASN value using the approximate ASN equation, Equation 3.69 or 3.70). However, if the first method is used to determine \( n_0 \), Wald's (1947) rule of truncation cannot be used to obtain a final decision. In this case, a fixed sample size procedure should be used to draw the final decision based on the observations sampled.

The last preparation is to prepare a field decision graph (e.g., Figure 5.5) or a field decision table (e.g., Table 5.15), since the decision boundaries of the SATP procedures are completely determined when the values of two variance limits and error probabilities have been specified. This will simplify the field calculations when the SATP procedure
5. Application Example

125

5.1.2 Sampling and testing process

The sampling and testing process for the SATP procedure is:

1. Take a group of $k$ random observations (e.g., $k$ trees or plots) from the application population. Measure the values of the dependent ($y_i$) and independent ($x_i$) variables for $i = 1 \ldots k$;

2. Obtain an estimate using the tested model ($\hat{y}_i$) and compute the error for each estimate, $d_i = y_i - \hat{y}_i$;

Figure 5.5: Field decision graph for applying the SATP procedures.
Table 5.15: Field decision table for applying the SATP procedures.

<table>
<thead>
<tr>
<th>Number of Observations $(n)$</th>
<th>Accept $H_0$ (* *)</th>
<th>Acceptance Value $h_1 + sn$</th>
<th>Test Statistics $(T)$</th>
<th>Rejection Value $h_2 + sn$</th>
<th>Reject $H_0$ (* *)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>10</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>15</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>20</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>25</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>30</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>35</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>40</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>45</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>50</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>...</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
3. Calculate the cumulative squared error as the test statistic, \( T = \sum_{i=1}^{n} d_i^2 \) (or \( \sum_{i=1}^{n} (d_i/y_i)^2 \) when SATP II is applied); and

4. Compare the observed test statistic to the acceptance \((h_1 + sn)\) or rejection \((h_2 + sn)\) value at \(n^{th}\) stage of sampling, then stop sampling, or to continue by taking an additional group of observations.

Since all three SATP testing procedures require the same information (the observed error, \(d_i\) or transformed error, \(d_i/y_i\) for \(i = 1 \ldots n_0\)), it is suggested that all three sequential testing plans be used simultaneously in the field. This can be done by simply using three field decision sheets or graphs to record the testing process for each SATP procedure, or a computer program can be developed, and a portable computer can then be used to facilitate the process of testing.

It should be indicated that this is not equal to a multiple testing plan although three developed sequential testing plans are used simultaneously. The reason for this is that these three testing plans do not test the same null hypothesis, but three different null hypotheses (see Section 5.1.1). In other words, they are actually three separate testing plans for the same problem (i.e., determine the applicability of a model), and there is no interest in the joint probability level of the test. The simultaneous use is only for overcoming the problem that the underlying assumptions of \(d\) and \(d/y\) cannot be tested until data are collected.

In general, the sampling and testing process of Steps 1 to 4 are repeated. Whenever one of three testing plans has the terminal decision, the post-tests for the normality and unbiasedness assumptions suggested in the next section will be carried out for the observed errors, \(d\) or \(d/y\), depended on whether SATP I, III, or II terminates. For example, if the SATP I or III terminates, the post-tests for normality assumption should be conducted for the observed error, \(d\). Also if the test shows that \(d\) is normally distributed,
the testing process is completed. However, if \( d \) is not normally distributed, SATP I and III should be ignored. The testing process is continued until the SATP II has the terminal decision, and the normality assumption of \( d/y \) will then be checked. Finally, if no terminal decision can be reached by either SATP I, II or III, the sampling and testing process will be continued until the truncation point, \( n_0 \), is reached.

5.1.3 Post-tests for the underlying assumptions

To validate the testing decisions reached by the SATP procedure, the normality assumption of \( d \) or \( d/y \) must be confirmed through post-tests. The test approaches for normality of errors suggested for Freese's (1960) procedure and Reynolds' (1984) estimation procedures (Reynolds 1984) are also suggested as the post-tests for normality of the SATP procedure. These are two goodness-of-fit testing procedures for normality based on the empirical distribution function (edf) of the errors, and named the Cramér-von Mises (\( W^2 \)) and Anderson-Darling (\( A^2 \)) tests. To apply these tests, the observed errors, \( d_i \) or \( d_i/y_i \), must first be sorted in ascending order as:

\[
d_{(1)} \leq d_{(2)} \leq \ldots \leq d_{(n)}
\]

Then the edf for the sorted errors is given by:

\[
F_n(d) = \begin{cases} 
\frac{i-1}{n} & \text{if } d_{(i-1)} \leq d < d_{(i)}, \, i = 1, 2, \ldots, n \\
1 & \text{if } d \geq d_{(n)} 
\end{cases}
\]

The test statistic of the Cramér-von Mises test is:

\[
W^2 = \sum_{i=1}^{n} \left( F[d_{(i)}] - \frac{2i - 1}{2n} \right)^2 + \frac{1}{12n} \tag{5.78}
\]

and the test statistic for the Anderson-Darling test is:

\[
A^2 = -\frac{1}{n} \sum_{i=1}^{n} (2i - 1)[\ln F(d_{(i)}) - \ln \{1 - F(d_{(n+1-i)})\}] - n \tag{5.79}
\]
The critical values for these tests are given by Stephens (1974) when the mean and/or the variance must be estimated from the sample. If both the mean ($\mu_d$) and the variance ($\sigma^2$) of $d$ or $d/y$ must be estimated, another test suggested for normality is the correlation coefficient ($r$) test based on normal probability plot suggested by Filliben (1975). In an actual application, these three procedures of normality can be all carried out for the same set of model errors to increase the reliability of the decision.

To test the unbiasedness of the observed error, $d$ or $d/y$, the hypothesis $H_0 : \mu_d = 0$ for error, $d$ (or $H_0 : \mu_{d/y} = 0$ for error, $d/y$) should be tested. The appropriate test statistic suggested for this hypothesis by Gregoire and Reynolds (1988) is:

$$T_p = \frac{\bar{d}}{S_d}$$

(5.80)

where $S_d$ is the standard error of $\bar{d}$; and $T_p$ follows the student’s $t$ distribution with $n - 1$ degrees of freedom.

If $H_0 : \mu_d = 0$ is rejected, it indicates that the estimates of the tested model are biased. Under this situation, the decision of SATP III will provide some indication of the model accuracy after bias-correction for the error, $d$.

It should be noted that since the observed model errors ($d$ or $d/y$) used to carry out the suggested post-tests are obtained from a sequential sampling process, and the sample size is a random variable instead a predetermined value, these post-tests are only approximate tests used for providing some indications of the underlying distributions of $d$ or $d/y$.

5.1.4 Making the final testing decision

After the post-tests for checking the normality assumption, a final testing decision on the applicability of the model can be made.
Chapter 5. Application Example

1. If the conclusion from the post-tests is that the observed $d$ is normally distributed, but $d/y$ is not. The decision reached by the SATP I or III is valid for making the final decision about the accuracy of the model. The decision reached by SATP II should be ignored;

2. If the post-tests concludes that the observed $d/y$ is normally distributed, but $d$ is not, then the decision reached by SATP II should be used as the final decision and the decision reached by SATP I and III should be ignored;

3. If the post-tests conclude that both $d$ and $d/y$ are not normal, the decisions reached by all three proposed SATP procedures should be ignored. In this case, the non-parametric procedures suggested by Reynolds (1984) may be used to provide an ad hoc final decision based on the observations collected in the sequential testing process;

4. If the hypothesis $H_0 : \mu_d = 0$ is rejected, but $d$ is normally distributed, then the final decisions about the accuracy of a model should be based on the decision reached by SATP III, which represents the accuracy after bias-correction.

5. Finally, if $n_0$ is reached (i.e., no terminal decision is reached by the procedure), the final decision depends on how $n_0$ is set. If $n_0 \geq 3E_{\sigma^2}(n)$, Wald's (1947) rule of truncation can then be used to obtain the final decision by comparing the value of test statistic $(T)$ with the average of the acceptance and the rejection value at $n_0$. If $n_0$ is the calculated sample size of an equally reliable fixed sample size procedure, then it means that there is no gain from using the SATP procedures, and a fixed sample size should be used, or the model should be fitted using the new data.
5.2 Application Example

In order to illustrate the proposed field application methods of the SATP procedures, an example of applicability testing for applying the B. C. standard volume model to a subpopulation was constructed.

5.2.1 Problem and data

The tree volume model suggested by the B.C. Ministry of Forests is:

\[
\log(V) = -4.34950 + 1.82276 \times \log(dbh) + 1.10812 \times \log(\text{Height})
\]  

(5.81)

for estimating the total tree volume of lodgepole pine (Watts 1983), where \(\log\) is the base 10 logarithm. This model was developed using a sample of 2,846 felled trees, and suggested for estimating the total volume of lodgepole pine trees for all age classes over most of the province (\(i.e.,\) FIZ's A to J). The percent standard error of estimate in volume units \(S_v(\%)\) was \(\pm 9.3\%\). The problem of interest was to determine whether this model was sufficiently accurate for estimating tree volume in FIZ E.

In constructing this example, the lodgepole pine trees collected from FIZ E described in Section 4.1 were used. The subset was viewed as representative of lodgepole pine growing in this zone. The calculated statistics for this data set are given in Table 5.16.

5.2.2 Specification of test parameters

To specify the acceptance and non-acceptance limits of error variance, \(\sigma_0^2\) and \(\sigma^2\), the known percentage standard error of estimate in volume units \((S_v(\%)\) was used for \(\sigma_0^2\). Since the errors \(d\) or \(d/y\) in the example would be in logarithmic units, \(S_v(\%)\) had to be represented in logarithmic units. This was obtained by rearranging Equation 2.11 to
Table 5.16: Calculated statistics of the lodgepole pine trees collected from FIZE in B.C.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (dbh cm)</td>
<td>192</td>
<td>47.2</td>
<td>25.21</td>
<td>15.0</td>
<td>163.0</td>
</tr>
<tr>
<td>dbh (cm)</td>
<td>192</td>
<td>13.1</td>
<td>9.33</td>
<td>3.8</td>
<td>43.0</td>
</tr>
<tr>
<td>Height (m)</td>
<td>192</td>
<td>13.4</td>
<td>7.90</td>
<td>3.1</td>
<td>36.8</td>
</tr>
<tr>
<td>Volume (m$^3$)</td>
<td>192</td>
<td>0.18968</td>
<td>0.30440</td>
<td>0.00240</td>
<td>1.97060</td>
</tr>
</tbody>
</table>

obtain:

$$SE_E = \log\left(\frac{0.3}{100} + 1\right) = 0.03862$$

(5.82)

This meant that the variance of residual ($MSE$) of the fitted model (Equation 5.81) in logarithmic units was $SE_E^2 = (0.03862)^2 = 0.00149$. Using this value, $\sigma_0^2$ was set to 0.0015. Under the assumption of normality of errors and a $\gamma$ of 0.05, the acceptance error limit in logarithmic units was $e_0 = 1.96 \times 0.03862 = 0.08$. The non-acceptable error limit was then arbitrarily set to $e_1 = 1.25 \times e_0 = 0.10$. This, in turn gave $\sigma_1^2 = e_1^2 / 1.96^2 = 0.0026$.

To set the truncation point ($n_0$), Equation 3.69 was used to calculate the expected sample size (ASN value) when $\sigma_0^2$ was true. The calculated value was $E_{\sigma_0^2}(n) = 13.8$, and $3E_{\sigma_0^2}(n) = 41.4$ trees. The truncation point was then set to 50 trees. To carry out SATP II, the percentage acceptance and non-acceptance limits, $p_0$ and $p_1$, were arbitrarily set to 10 and 15, respectively. Finally, both the probabilities of the type I and II error were
Chapter 5. Application Example

Each set to 0.05. In summary, the specified parameters of this example were:

- Acceptable error limit: \( e_0 = 0.08 \) (logarithmic units)
- Unacceptable error limit: \( e_1 = 0.10 \) (logarithmic units)
- Probability level of accuracy required: \( \gamma = 0.05 \)
- Percent acceptable error limit: \( p_0 = 10 \)
- Percent unacceptable error limit: \( p_1 = 15 \)
- Desired type I error: \( \alpha = 0.05 \)
- Desired type II error: \( \beta = 0.05 \)
- Truncation point: \( n_0 = 50 \)

5.2.3 Testing results

These specified values were then used to compute the intercepts \( h_1 \) and \( h_2 \) and slope \( s \) of the decision boundaries of SATP procedure using the formulae presented in Chapter 3. These calculated intercepts and slopes were:

- For SATP I and III: \( h_1 = -0.02725, h_2 = 0.02725, \) and \( s = 0.00207 \);
- For SATP II: \( h_1 = -0.02760, h_2 = 0.02760, \) and \( s = 0.00380 \).

With these determined decision boundaries, the SATP procedure was carried out by taking five randomly selected observations from the FIZ E lodgepole pine data at each stage of sampling, and by calculating the sum of squared errors for this group of observations. The cumulative squared errors up to this stage of sampling were then calculated, and compared to the predetermined decision boundaries for each of three SATP procedure.

Three developed SATP procedures were used simultaneously in this example. At the 4\(^{th} \) stage of sampling, SATP II reached its terminal decision of accepting the model (Figure 5.6), and 20 observations were taken. The post-tests were then carried out to check the normality assumption of the relative error, \( d/y \) based on these 20 observed
relative errors. The calculated statistics for Cramér-von Mises ($W^2$) and Anderson-Darling ($A^2$) tests were 0.2224 and 1.3627, respectively. The modified test statistics for these two tests determined from Stephens' (1974, p.732) Case 3 were 0.2280 and 1.5501, respectively. Both these modified values were greater than the critical values at 0.01 significance level. The calculated correlation coefficient based on the probability plot was 0.8599. From the Table 1 of Filliben (1975, p.113) with $n = 20$, this calculated value was found to be significant even at 0.005 significance level. Therefore, all these three tests rejected the normality assumption of $d/y$.

With the rejection of the normality of $d/y$, the decision reached by SATP II was ignored. The sampling and testing process were continued using SATP I and III. SATP III reached the terminal decision of accepting the model at 5th stage of sampling (Figure 5.7), and 25 observations were sampled. Again the post-tests were carried out to check the normality assumption for the error, $d$, based on the 25 observed errors. The calculated statistics for Cramér-von Mises ($W^2$) and Anderson-Darling ($A^2$) tests were 0.0280 and 0.1924, respectively. These values were less than the critical values in Stephens' (1974) Case 3 table even at the 0.15 significance level. The calculated correlation coefficient based on the probability plot was 0.9922. From the Table 1 of Filliben (1975, p.113) with $n = 25$, this calculated value was found to be not significant even at the 0.25 significance level. Therefore, all these three tests confirmed that normality assumption of $d$ was reasonable. This means that the testing decision reached by SATP III is valid.

Since SATP III is an accuracy test after bias-correction, we would like to know whether the volume estimates of the tested model are biased. This requires a test of the unbiasedness assumption of $d$ (i.e., to test the hypothesis $H_0 : \mu_d = 0$). The $t$-test statistic ($T_p$) was calculated as -3.9578 using the 25 observed errors. Since $|T_p|$ was greater than the 95 two-tails percentile of the $t$-distribution with 24 degrees of freedom, the null hypothesis $H_0$ was rejected. The conclusion was that the volume estimates from the tested
Figure 5.6: Field decision graph for SATP II when applied to test the applicability of B.C. standard volume model of lodgepole pine for the application to the subregion, FIZ E with specified parameters: $p_0 = 10$ and $p_1 = 15$, $\gamma = 0.05$, $\alpha = \beta = 0.05$ and $n_0 = 50$.

model were biased. With the information about the model errors obtained, the sequential testing process could be completed without to continue the testing process of SATP I.

5.2.4 Final testing decision

Since post-testing of normality indicated that the distribution of $d$ was normal, but $d/y$ was not, the testing decision reached by the SATP III was valid, but that of the SATP II was not. In this example, the information about the accuracy of the tested model based on SATP III was as follows. The volume estimates were biased, the average bias ($\bar{d}$)
Figure 5.7: Field decision graph for SATP III when applied to test the applicability of B.C. standard volume model of lodgepole pine for the application to the subregion, FIZ E with specified parameters: $e_0 = 0.08$ and $e_1 = 0.10$ in logarithmic units, $\gamma = 0.05$, $\alpha = \beta = 0.05$ and $n_0 = 50$. 
and its standard error \( (S_d) \) were -0.0234 and 0.00592 in logarithmic units, respectively. However, if the model bias could be assumed to be constant across dbh classes, the accuracy of the model after bias-correction would be considered to be acceptable against the accuracy requirement specified.

Therefore, the conclusion could be made as follows. We would expect that at least 95 percent of the absolute values of the model errors represented in logarithmic units to be less than 0.08 when the model is applied to FIZE, if the bias in model estimates was corrected. This is equivalent to saying that the true variance of errors (variation of model estimates) is less 0.0015 in logarithmic units. However, the estimated volumes from the model overestimated the actual tree volumes with an average bias of -0.0234 in logarithmic units. If the user considered such a bias as tolerable, the model estimates would be acceptable without modification. Otherwise, a modification of \( y_i - 0.0234 \) for the estimated logarithmic volumes could be used when the model is applied in FIZE.

To confirm the results obtained, the errors were calculated for all trees in FIZE. The actual accuracy of the model after correcting bias (i.e., \( d' = d - d = d + 0.0234 \)) was calculated. The results were \( P(|d'| \leq 0.08) = 0.9531 \) and the true variance of \( d' \), \( \sigma^2 \) was 0.00143. These results are consistent with the conclusion reached by the SATP procedures. Finally, if Freese’s (1960) test I or III was used instead of SATP procedures, the sample size required (Equation 4.73) to have the desired error probabilities, \( \alpha = \beta = 0.05 \), and the ratio \( \sigma_1^2/\sigma_0^2 = 0.0026/0.0015 = 1.733 \), would be 70. Therefore, the use of the developed SATP procedures in this example resulted in a 100\( (1 - 25/70) = 64.3\) saving of the sample cost.
Chapter 6

Conclusions and Recommendations

Mathematical models are widely used for estimating tree or stand volumes in forestry. In the development of a tree volume model, model specification, parameter estimation, and model validation are usually based on data collected over a large geographical region of forested land, and for a particular tree species. Developed in this way, a model will only be appropriate for providing accurate volume estimates of trees in the same population. However, in forest inventories, a volume model may be considered for application to a more local region (subpopulation), or to the same species in a different geographical region (new population). Under these situations, the accuracy of the model is uncertain. An applicability test should then be carried out to determine the actual accuracy of the model in order to use it with confidence.

Statistical testing procedures based on data with a predetermined, sufficiently large sample size (fixed sample size procedures) are conventionally used for the purpose of applicability testing or validation of forestry models. Since an applicability test is not used to provide accurate estimates for the parameters of model errors, but to classify a model into two wide classes only (i.e., acceptable or unacceptable for a given application), it is possible that a decision to accept or reject a model can be made with a small sample, especially when the actual accuracy of the model is far below or above the requirement of the user. This suggests that a fixed sample size procedure may not be efficient for the purpose of applicability testing of volume models. This is particularly of concern when data collection is expensive, time-consuming, or destructive, as with data required for
developing volume models.

The first objective of this research was to suggest alternative procedures to fixed sample size procedures for applicability testing of tree volume models, and other forestry estimation or prediction models. The second objective was to test the reliability of the suggested alternative testing procedures under different application conditions. The third objective was to suggest and illustrate appropriate field procedures for the application of the suggested alternative testing procedures.

Information from the literature was used to develop three new testing plans with a variable sample size, labeled as sequential accuracy testing plans (SATP) I, II and III. The SATP procedures are extensions of Freese’s (1960) procedure of accuracy testing to Wald’s (1947) sequential probability ratio test (SPRT) plan. Approximate Operating Characteristic (OC) and Average Sample Number (ASN) equations were also derived for calculating the probability of making a correct decision and the expected sample size for the SATP procedures, respectively based on Wald’s the OC and ASN functions for SPRT plans.

The second objective of the research was met by addressing volume model testing under different error assumptions. For this purpose, Monte Carlo techniques were used to generate normally distributed errors, or a mixture of normally distributed errors for tree volume models. The conclusions reached from the simulations were as follows. First, if the errors of a model are iid normal variables with a zero mean (no estimation bias is present) and a constant variance, SATPs I and III are reliable for classifying the applicability of a model. The use of SATP I or SATP III will obtain a similar testing decision and expected sample size under this situation. The simulation results show that under this distributional assumption, SATPs I and III have smaller probabilities of making a wrong decision than those specified (i.e., α and β). The Monte Carlo estimates of the ASN values showed that in applying SATP I and III, an expected sample size for
making a decision with less than 30 observations can be obtained, if the ratio of $e_1/e_0$ is larger than 2.00 for $\alpha = \beta = 0.05$, or if $\alpha = \beta$ is larger than 0.10 when the the ratio of $e_1/e_0$ is somewhat less then 2.00. Also, by comparing the resulting ASN values to the sample sizes required by the equally reliable fixed sample size procedures (i.e., having the same $\alpha$ and $\beta$, and the ratio, $e_1/e_0$), the use of the SATP procedures would result in, on average, a 40 to 60% of sampling cost-saving.

Second, if the errors of a model are iid normal variables with a non-zero mean (estimation bias is present) and a constant variance, SATP I is still valid for classifying the applicability of an existing model based on the specified values of $e_0$, $e_1$ and $\gamma$. However, the use of SATP I will reject an inaccurate model regardless of the sources of inaccuracy (large bias, lack of precision, or both) under the distributional assumptions. If accuracy of a model after bias-correction is desired, SATP III should be used instead of SATP I. The use of SATP III requires an assumption that the biases in the model estimates are the same, and that the mean of the observed errors in the current sample is a good estimate for the average bias across the range of application. Since the basis of Freese's (1960) accuracy test (i.e., test to determine whether the true variance of model errors exceeds an user-supplied hypothesized variance) was incorporated into the derivation of the SATP I and II, the simulation results confirmed a statement made for Freese's test I and II by Reynold (1984). That is, when bias is present, the true variance of errors would need to be even smaller than the user-supplied hypothesized for model acceptance (i.e., $\sigma^2_0$ in the SATP procedures). If the bias is large enough, it may not be possible to accept the model no matter how small the true error variance.

If the errors of a model can be considered to be a mixture of a finite number of normal distributions with zero means, and the error variances are proportional to tree dbh, SATP I and III are still appropriate for determining the applicability of an existing model when the random observations are taken uniformly from each distribution.
Because the distributional assumptions of SATP II are exactly the same as those required by SATP I, all conclusions obtained for SATP I are also true for SATP II. However, in applying SATP II, the distributional assumptions are made for the relative errors of a model instead of the absolute errors. Also, percent error limits must be specified as the accuracy requirement.

Simulations were carried out to examine the accuracy of the suggested OC and ASN equations, and to investigate the effects of two modifications of Wald’s (1947) assumptions for SPRT plans on applicability testing of tree volume models. The results showed that the suggested OC equation worked well for approximating the actual OC values of SATP procedure in general. However, the suggested ASN equation consistently underestimated the actual ASN values of the SATP procedures. Hence, the ASN equation suggested can be used for providing some indication of the expected sample size of the SATP procedures only.

By modifying the SATP procedures and using group selection instead of single selection of random observations, the resulting probabilities of type I and II errors consistently decreased, and the average sample size required to make a decision consistently increased as the group size increased. In general, the use of Wald’s (1947) rule for setting a maximum number of observations sampled (truncation point) to stop an SATP procedure and obtain a final decision before a terminal decision is reached was reliable only when the truncation point was larger than three times the calculated ASN value using the suggested ASN equation.

For field applications, it was suggested that sampling continue until all three tests reached a terminal decision, or the preset truncation point was reached. Post-tests for normality and unbiasedness were also suggested.

Some problems exist in the SATP procedures and further studies on these problems are needed to improve the usefulness of sequential sampling procedures for forestry model
testing. First, the normality assumption for the unconditional distribution of model errors or relative errors is critical for applying the SATP procedures. This assumption may not be met for some applications. To overcome this problem, studies should be carried out to investigate the actual distributions of the errors for some common forestry models, and to suggest appropriate error transformation methods to meet the normality assumption. Another approach for solving this problem may be to develop sequential non-parametric testing procedures. Second, the random selection methods required to apply the SATP or other sequential sampling procedures (i.e., random single or random grouping selection) may be impractical for forestry model testing since it will increase the travel time between the sampled units, and degrade the usefulness of sequential procedures. Also, the observations obtained from random selection methods may not cover the entire range of the model, and this will reduce the representativeness of the observations for model testing. Therefore, other sampling methods may be appropriate for applying the sequential testing procedures. Finally, the data processing for each selected sampling unit (i.e., tree or plot) must be completed in the field in order to continue the testing process of the SATP procedures. If the measurements obtained involve complicated analyses or computation processes, the application of SATP procedures may become difficult. Appropriate computer programs should be developed and the use of a portable computer is strongly recommended for applying the SATP procedures.

As variable sample size procedures, the SATP procedure does not require a predetermined, sufficiently large sample set. This property of the SATP procedures insures that the sample size required to make a testing decision is minimum. The use of the SATP procedures do not preclude the use of other fixed sample size procedures, because sampling can be always continued to let the sample size be large enough for using a fixed sample size procedure. Therefore, the developed SATP procedures are considered to be
appropriate alternative testing procedures to the conventional fixed sample size procedures. Also, based on the similarity between the purpose of an applicability test and model validation, the developed SATP procedures should find fairly wide applicability in validating forestry estimation or prediction models, especially when data collection is highly expensive, time-consuming and destructive.

The main contribution made by this research was to lay the foundation for future research in applying sequential analysis approaches to forestry model testing and thereby bring to the forestry researcher's attention the possible wide applications of sequential analysis procedures.
References Cited


References Cited


References Cited


References Cited


Marshall, P. and Y. Wang. 1993. OPSILIR (Version 2.0): A computer program to determine the efficient sampling distribution for simple linear regression problems. Funding for the development of this program was supplied through the STDF-AGAR program of the Science Council of British Columbia, Canada.


