

**DEVELOPING ECOREGION-BASED HEIGHT-
DIAMETER MODELS AND REFERENCE-AGE
INVARIANT POLYMORPHIC HEIGHT AND SITE INDEX
CURVES FOR BLACK SPRUCE AND JACK PINE IN
MANITOBA**

BY

WENLI XU

**A thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of**

Master of Science

**Department of Botany
University of Manitoba
Winnipeg, Manitoba
R3T 2N2**

© Wenli Xu 2004

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

**DEVELOPING ECOREGION-BASED HEIGHT-
DIAMETER MODELS AND REFERENCE-AGE
INVARIANT POLYMORPHIC HEIGHT AND SITE INDEX
CURVES FOR BLACK SPRUCE AND JACK PINE IN
MANITOBA**

BY

WENLI XU

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree
Of
MASTER OF SCIENCE**

Wenli Xu © 2004

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Acknowledgements

First and foremost I wish to acknowledge the invaluable and generous teaching, advice, support and patience of my thesis advisors, Drs. G. Geoff Wang and Norman C. Kenkel. I also greatly appreciate the advice, constructive criticism and time donated by my committee members, Drs. J. Thomas Booth and Xikui Wang.

Thanks go to Ed East, Gerry Becker, Greg Carlson and Jianwei Liu for giving me a lot of support and help during the past years and my very capable field assistants were Dave Vande Vyvere, Erica Schweitzer, Jane Bell and Dave Chetybuk.

Thanks also go to Dr. Shongming Huang, Dr. Yuqing Yang, Dr. Chuankaun Wang, Dr. Zhehong Ye, Kevin Kemball and Sophan Chhin for giving me a lot of helps and suggestions during the past years.

Finally, I would like to thank my family and friends, particularly my parents, wife and son, for their love, encouragement and support.

Abstract

Eight different height-diameter models, five height prediction models with stand variables as predictors and six height and site index models were selected, examined, compared and developed for black spruce (*Picea mariana* [Mill.] B.S.P.) and jack pine (*Pinus banksiana* Lamb.) in Manitoba.

Eight different height-diameter models were fitted using nonlinear modeling techniques and compared in each of the five ecoregions in Manitoba: Churchill River Upland (Ecoregion 88); Hayes River Upland (Ecoregion 89); Lac Seul Upland (Ecoregion 90); Lake of the Woods (Ecoregion 91); Mid-Boreal Lowland (Ecoregion 148). Results suggested that the Weibull-type and Chapman-Richards models were the most suitable models. Differences of the height-diameter relationship among and between ecoregions were tested. Testing results suggested that height-diameter models significantly differed between ecoregions, indicating ecoregion-based or 'local' height-diameter models are needed for prediction purposes. The ecoregion-based height-diameter models developed in this study may provide more accurate information for developing forest growth and yield models.

Five height prediction models were examined with the addition of stand density variables into the base height-diameter model. Adding stand variable resulted in increased prediction accuracy.

Six height and site index models were examined and compared for black spruce and jack pine based on the provincial stem analysis data and the most suitable models were selected for Manitoba and the provincial height and site index prediction tables were produced.

TABLES OF CONTENTS

	Page
ACKNOWLEDGEMENT	i
ABSTRACT	ii
TABLES OF CONTENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
LIST OF APPENDICES	xi
CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW.....	1
1.1 Introduction.....	1
1.2 Height-diameter models	3
1.3 Site index models.....	6
1.4 Overall research objectives	7
CHAPTER 2. ECOREGION-BASED HEIGHT-DIAMETER MODELS DEVELOPMENT FOR BLACK SPRUCE AND JACK PINE	9
2.1 Introduction.....	9
2.2 Study Area	11
2.3 The data	16
2.4 Model Selections and Analysis.....	18
2.4.1 Selected height-diameter models	18
2.4.2 Error structure of the height-diameter models.....	18
2.4.3 Comparison of models among the five eco-regions and between each paired ecoregions	22
2.4.4 Model fitting.....	24
2.4.5 Model validation.....	26
2.5 Results	27
2.5.1 Model fitting and comparison for each ecoregion	27
2.5.1.1 Ecoregion 88	27
2.5.1.2 Ecoregion 89	33
2.5.1.3 Ecoregion 90	39
2.5.1.4 Ecoregion 91	45
2.5.1.5 Ecoregion 148	51

2.5.2 Model validation.....	57
2.5.2.1 Model validation for black spruce.....	57
2.5.2.2 Model validation for jack pine.....	58
2.5.3 Testing height-diameter relationships among and between ecoregions....	64
2.5.3.1 Testing height-diameter relationships among the five ecoregions	64
2.5.3.2 Testing height-diameter relationships between paired ecoregions...	64
2.5.4 Prediction errors of applying the combined model to each ecoregion.....	68
2.5.5 Prediction errors of applying Alberta provincial height-diameter model.	70
2.6 Discussion and Conclusions	72

CHAPTER 3. FITTING HEIGHT-DIAMETER MODELS INCORPORATED WITH STAND VARIABLES FOR JACK PINE ... 74

3.1 Introduction.....	74
3.2 Study Area	75
3.3 The Data.....	75
3.4 Selecting base model for introducing stand density variables	77
3.5 Model fitting and analysis	79
3.6 Results	80
3.6.1 Fitting models incorporated with stand variables	80
3.6.2 Model validation.....	83
3.7 Discussion and conclusions.....	84

CHAPTER 4. SITE INDEX MODELS DEVELOPMENT FOR BLACK SPRUCE AND JACK PINE 85

4.1 Introduction.....	85
4.2 The data	86
4.3 Model Selections and Analysis.....	92
4.3.1 Selected height growth and site index models	92
4.3.2 Data structure.....	93
4.3.3 Model fitting and analysis	93
4.3.4 Model validation.....	94
4.3.4.1 The paired t test	94
4.3.4.2 The Separate t tests.....	94
4.3.4.3 Simultaneous F test.....	95
4.3.4.4 The novel test.....	95
4.4 Results	97
4.4.1 Fitting models	97
4.4.2 Model comparison by prediction statistics	100
4.4.3 Model comparison by graphical performance	100
4.4.4 Model validation.....	114
4.4.5 Model re-fitting	118
4.5 Discussion and Conclusions	124

LITERATURE CITED 129
APPENDIX 1 136
APPENDIX 2 140

LIST OF TABLES

Table 2.1. Summary of the data used for height-diameter model development by species and ecoregions.	17
Table 2.2. Estimated parameters and RMSE of models [2.1] to [2.8] for black spruce and jack pine in ecoregion 88.	29
Table 2.3. Prediction statistics of models [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 88.	30
Table 2.4. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 89.	35
Table 2.5. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 89.	36
Table 2.6. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 90.	41
Table 2.7. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 90.	42
Table 2.8. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 91.	47
Table 2.9. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 91.	48
Table 2.10. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 148.	53
Table 2.11. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 148.	54
Table 2.12. Prediction statistics based on the validation data set, calculated for model [2.1] to [2.8] and the five ecoregions for black spruce and jack pine.	61
Table 2.13. F-test of the differences for the height-diameter among five forest ecoregions.	64
Table 2.14. F-test of the differences for the height-diameter between each paired ecoregions in Manitoba.	66
Table 2.15. Predicting errors of applying the combined model to each ecoregion.	69

Table 2.16. Predicting errors of applying Alberta provincial height-dbh model to each ecoregion.....	71
Table 3.1. PSP summary statistics for stands with jack pine in ecoregion 91.	76
Table 3.2. Fit statistics of weighted nonlinear height-diameter model [3.1] for jack pine in ecoregion 91.....	81
Table 3.3. Fit statistics for jack pine height prediction model [3.6] in ecoregion 91.....	81
Table 4.1. Summary of the data used for site index model development of black spruce.	90
Table 4.2. Summary of the data used for site index model development of jack pine.	91
Table 4.3. Estimated parameters and fitted statistics of model [4.1] to [4.6] for black spruce and jack pine.....	98
Table 4.4. Prediction statistics of model [4.1] to [4.6] for black spruce and jack pine.....	99
Table 4.5. Height prediction statistics of model [4.1] to [4.6] based on the validation data set for black spruce and jack pine.	116
Table 4.6. Results of four different tests based on the validation data for site index prediction of model [4.1] to [4.3] and [4.5] to [4.6] for black spruce and jack pine.	117
Table 4.7. Estimated parameters and fitted statistics of model [4.2] for black spruce and jack pine based on the pooled data.	118
Table 4.8. Accuracy and precision of site index predictions from model [4.2] for black spruce.	120
Table 4.9. Accuracy and precision of site index predictions from model [4.2] for jack pine.....	121
Table 4.10. Accuracy and precision of height predictions from model [4.2] for black spruce.	122
Table 4.11. Accuracy and precision of height predictions from model [4.2] for jack pine.....	123

LIST OF FIGURES

Figure 2.1. Distribution of the five studied ecoregion in Manitoba: Churchill River Upland (Ecoregion 88); Hayes River Upland (Ecoregion 89); Lac Seul Upland (Ecoregion 90); Lake of the Woods (Ecoregion 91); Mid-Boreal Lowland (Ecoregion 148).	12
Figure 2.2. Plots of residual against predicted height for black spruce in ecoregion 91 (a, b, c, d, e, f, g, h represent models (2.1) , (2.2) , (2.3) , (2.4) , (2.5) , (2.6) , (2.7) , (2.8)).	20
Figure 2.3. Plots of studentized residual against predicted height for black spruce in ecoregion 91 (a, b, c, d, e, f, g, h represent model (2.1) , (2.2) , (2.3) , (2.4) , (2.5) , (2.6) , (2.7) , (2.8)).	21
Figure 2.4. Plots of total tree height against DBH for black spruce in ecoregion 88 based on model [2.2].	31
Figure 2.5. Plots of studentized residual against predicted height for black spruce in ecoregion 88 based on model [2.2].	31
Figure 2.6. Plots of total tree height against DBH for jack pine in ecoregion 88 based on model [2.3].	32
Figure 2.7. Plots of studentized residual against predicted height for jack pine in ecoregion 88 based on model [2.3].	32
Figure 2.8. Plots of total tree height against DBH for black spruce in ecoregion 89 based on model [2.3].	37
Figure 2.9. Plots of studentized residual against predicted height for black spruce in ecoregion 89 based on model [2.3].	37
Figure 2.10. Plots of total tree height against DBH for jack pine in ecoregion 89 based on model [2.2].	38
Figure 2.11. Plots of studentized residual against predicted height for jack pine in ecoregion 89 based on model [2.2].	38
Figure 2.12. Plots of total tree height against DBH for black spruce in ecoregion 90 based on model [2.4].	43
Figure 2.13. Plots of studentized residual against predicted height for black spruce in ecoregion 90 based on model [2.4].	43

Figure 2.14. Plots of total tree height against DBH for jack pine in ecoregion 90 based on model [2.3].....	44
Figure 2.15. Plots of studentized residual against predicted height for jack pine in ecoregion 90 based on model [2.3].....	44
Figure 2.16. Plots of total tree height against DBH for black spruce in ecoregion 91 based on model [2.2].....	49
Figure 2.17. Plots of studentized residual against predicted height for black spruce in ecoregion 91 based on model [2.2].....	49
Figure 2.18. Plots of total tree height against DBH for jack pine in ecoregion 91 based on model [2.2].....	50
Figure 2.19. Plots of studentized residual against predicted height for jack pine in ecoregion 91 based on model [2.2].....	50
Figure 2.20. Plots of total tree height against DBH for black spruce in ecoregion 148 based on model [2.1].....	55
Figure 2.21. Plots of studentized residual against predicted height for black spruce in ecoregion 148 based on model [2.1].....	55
Figure 2.22. Plots of total tree height against DBH for jack pine in ecoregion 148 based on model [2.2].....	56
Figure 2.23. Plots of studentized residual against predicted height for jack pine in ecoregion 148 based on model [2.2].....	56
Figure 2.24. Height-diameter curves overlay by the best fitted models in five ecoregions. Above is for black spruce and below is for jack pine.....	67
Figure 3.1. Plot of studentized residuals against the predicted height for jack pine from weighted nonlinear least squares of model [3.6].	82
Figure 4.1. Site index and height distribution of the black spruce and jack pine trees used for model fitting. (BS) Calculated site index versus sectioned tree height data of black spruce. (JP) Calculated site index versus sectioned tree height data of jack pine.....	88
Figure 4.2. Residuals plotted against predicted height of black spruce.....	103
Figure 4.3. Residuals plotted against predicted height of jack pine.	104
Figure 4.4. Site index prediction biases by breast height age classes of black spruce from model [4.1] to [4.6]. Each line represents a tree on the graph.	105

Figure 4.5. Average site index prediction biases by breast height age classes of black spruce from model [4.1] to [4.6]. The solid line in the middle represents the average height prediction errors and the two dashed lines ($\bar{e} \pm$ standard deviation) connected the standard deviations associated with the average errors. 106

Figure 4.6. Average height prediction biases of model [4.1] to [4.6] by breast height age classes for black spruce. The solid line in the middle represents the average height prediction errors and the two dashed lines ($\bar{e} \pm$ standard deviation) connected the standard deviations associated with the average errors. 107

Figure 4.7. Site index curves generated from model [4.1], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves. 108

Figure 4.8. Site index curves generated from model [4.2], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves. 109

Figure 4.9. Site index curves generated from model [4.3], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves. 110

Figure 4.10. Site index curves generated from model [4.4], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves. 111

Figure 4.11. Site index curves generated from model [4.5], overlaid with the observed tree sectioning data of black black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves. 112

Figure 4.12. Site index curves generated from model [4.6], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves. 113

Figure 4.13. Model [4.2], Nigh et al. (2002), Huang et al. (1994) and Ker and Bowling (1991) height –age curves for black spruce plotted against breast height age for site indices 5, 10 and 15 m. 128

Figure 4.14. Model [4.2], Huang et al. (1994), Ker and Bowling (1991) and Carmean (2001) height –age curves for jack pine plotted against breast height age for site indices 5, 10,15 and 20 m. 128

LIST OF APPENDICES

Appendix 1	Height and site index prediction table for black spruce.....	136
Appendix 2	Height and site index prediction table for jack pine	140

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Both jack pine (*Pinus banksiana* Lamb) and black spruce [*Picea mariana* (Mill.) B.S.P.] are among the most widely distributed and economically important tree species in Manitoba. The total merchantable volume of black spruce is 196.1 million m³ and that of jack pine is 140.5 million m³ (Manitoba Conservation, 2001)

Black spruce (*Picea mariana* (Mill.)B.S.P.) is a main tree species in the boreal forests of Canada, and its geographic range includes much of Canada and the North-eastern United States. The northern boundary of black spruce follows the tree line, extending east from the west coast of Alaska, through British Columbia to the Atlantic coast of Labrador. The southern boundary includes central British Columbia, Alberta, Saskatchewan, Manitoba, Minnesota, Wisconsin (Viereck and Johnston 1990, Farrar 1995).

Black spruce is commonly found in pure stands on lowland or wet organic soils, but it can also be found in mixed stands on a variety of mineral soils (Viereck and Johnston 1990) with jack pine (*Pinus banksiana* Lamb) and trembling aspen (*Populus tremuloides* Michx.). Because of the lateral growth of root systems, black spruce can survive in very shallow soils or in soils with shallow oxygenated zones. The majority root biomass of black spruce is located within the upper 20 cm of a soil profile (Viereck and Johnston 1990). Black spruce is an important resource for the pulp wood industry and it is also used for lumber (Viereck and Johnston 1990).

The mean height of black spruce is 12 to 20 meters at maturity on relatively high quality sites. On poor quality sites, the mean height is somewhat lower at about 8 to 12

meters (Viereck and Johnston 1990). The character of shade tolerance allows black spruce to develop under a closed canopy, although growth is better under open canopy (Farrar 1995).

Jack pine (*Pinus banksiana* Lamb) is also a very widely distributed tree species in North America. It is an important source of pulpwood, lumber, and round timber (Rudolph and Laidly 1990). The form of jack pine is quite variable, depending on soil and crown conditions (Hosie 1979). In open conditions, it has a conical open crown of ascending and arching branches and a tapered trunk. In the closed stand, the live-crown becomes reduced, usually covering less than 1/5 of the length of the tree; the trunk is slender, straight, and with little taper. On the poor soils and rocky sites, jack pine tree is often short and twisted. The root systems are usually widespread and moderate in depth (Hosie 1979).

Jack pine is a boreal forest tree species and usually grows in pure stands or in mixture with black spruce. It can also be found with white spruce, balsam fir, trembling aspen, balsam poplar and white birch in mixed stands. Usually, jack pine establishes and maintains even-aged pure stands after forest fires (Sweda et al. 1979).

The mean height of jack pine is 14 to 24 meters at maturity on relatively high quality sites (Hosie 1979). Jack pine is a pioneer, shade-intolerant species typically found on dry sandy soils. It survives well on poor soils such as rocky sites. The wood of jack pine is moderately hard and heavy, but not particularly strong, and is used for general construction, pulp, railway ties and mine timbers. When mature, even-aged stands of jack pine are commonly harvested by clear-cutting. Regeneration of jack pine often comes from natural seed-fall, direct seeding or by planting seedlings.

The provincial site index models and ecoregion-based height-diameter models developed in this study may provide more accurate information for forest stand productivity classifications, and improve the accuracy and precision of forest inventory projections and provincial wood supply analysis.

1.2 Height-diameter models

Foresters often use height-diameter models to predict total tree heights based on observed diameter at breast height (DBH) for estimating tree or stand volume and site quality (e.g., Stout and Shumway 1982). Therefore, estimates of tree or stand volume and site quality heavily rely on the accuracy of height-diameter models. Many postulated height-diameter models have been developed for different species and regions since Meyer (1940) suggested the exponential height-diameter relationship.

Predicting total tree height based on observed diameter at breast height outside bark (DBH, 1.3 meters above ground) is routinely required in forest management and silvicultural research work (Arabatzis and Burkhart 1992; Huang et al. 2000). Examples include volume sampling programs, pre-harvest surveys. However, the height-diameter relationship of a given species heavily depends on local environmental conditions and varies within a geographic region (Huang et al. 2000; Zhang et al. 2002).

In Ontario, the Chapman-Richards growth function has been used to model jack pine (*Pinus banksiana* Lamb.) tree height-diameter relationships at provincial, regional, and ecoregional levels (Zhang, et al. 2002). Tree height-diameter relationships of jack pine are significantly different among the geographic regions of Ontario, reflecting a dependence on local climatic, soil, and ecological conditions. The provincial and regional height-diameter models are therefore not appropriate for predicting tree heights at the

ecoregional level (Zhang, et al. 2002). Applying a specific ecoregional model to other ecoregions resulted in significant biases in the prediction of local tree heights (Zhang, et al. 2002).

In order to increase prediction accuracy, the height-diameter base models which use DBH as a predictor are modified to include stand variables, such as stand density, basal area of a stand and species compositions using the parameter prediction approach (Clutter et al. 1983; Huang and Titus 1994; Staudhammer and LeMay 2000). Because of the cost of obtaining accurate age data, the tree or stand ages are typically not introduced as additional predictors in height-diameter base models (Huang and Titus 1994).

To select appropriate height-diameter models, Huang et al. (1992) plotted height against DBH for white spruce (*Picea glauca* Voss.) and aspen (*Populus tremuloides* Michx.). Height-diameter relationship for white spruce resembled a typical sigmoid or S-shaped curve with an inflection point occurring in the lower portion of the data points. The height-diameter relationship for aspen indicated a concave curve with no apparent inflection point. Based on these findings, Huang et al. (1992) proposed that functions generating concave shapes and sigmoid shapes should be considered in the process of model selection in order to obtain an accurate height-diameter model. They selected and compared 20 published, nonlinear height-diameter functions for 16 tree species in Alberta. Based on the t-statistics for significance of parameters, mean squared error values, and the plot of studentized residuals against the predicted heights, all the equations recommended for height-diameter relationships from their study produced S-shaped curves with inflection points. Fang and Bailey (1998) investigated 33 height-diameter equations, including S-shaped and concave-shaped curves, for tropical forests on

Hainan Island in southern China. They suggested the selection of height-diameter relationships should consider both data-related and reasonable biological criteria, such as monotonic increment, functional inflection point, and asymptotic value.

Lei (1998) suggested that the data for modeling the height-diameter relationship should include early height growth. In addition, functions rendering a concave curve cannot describe tree growth behaviours appropriately, especially for fast-growth tree species. However, if a data set includes only larger or older trees beyond the inflection point, then a model generating a concave curve will probably work best. Even though a sigmoid model is correct biologically, it may not provide a reasonable fit if the range of the data is beyond the point of inflection for the species and growing conditions (Lei 1998).

The Bertalanffy-Richards growth function (Richards 1959) and the Schnute (1981) model are probably the most flexible and versatile functions available for modeling height-diameter relationships (Lei 1998). Both models are able to assume various shapes with different parameter values, and they produce satisfactory curves under a wide range of biological and ecological conditions. The Schnute model is usually easy to fit and achieve convergence quickly (Lei 1998).

1.3 Site index models

In North America, site index has been commonly used as a measure of forest productivity (e.g., Carmean 1975; Monserud 1984; Huang, et al. 1994b; Wang et al. 2000) and is a 'driving' variable in many growth and yield models (e.g., Clutter et al. 1983; Davis and Johnson 1987) for even aged stands. Site index is usually defined as the mean height of the dominant and co-dominant trees 50 years after they reached breast height (1.3m) (Huang, et al. 1994b; Philip 1998; Carmean 2001). Breast height age was used because early height growth is erratic and does not necessarily reflect site productivity (Clutter et al. 1983). Site index is the mean height of dominant and co-dominant trees of a stand in Manitoba. Site index provides standardized comparisons of productive potential between sites, across a broad range of existing stand conditions.

Given a set of stem analysis data obtained from dominant and co-dominant trees, as site index proceeds from low to high, the shape of height-age curves will change as well as their magnitude. Usually, trees with the lowest site index display a flatter, more linear curve while trees with the highest site index approach the asymptote rapidly; thus the curves for different site classes diverge for young trees, then they tend to become parallel or converge at older ages. If individual curves are fit to relatively narrow classes of site index, most of the variability of curve shape across site will be recovered (Carmean 1972). However, if an equation is fit to data pooled across site index, the trend in curve shape across site index will be underestimated. Goelz and Burk (1992) indicated that this observation was a "regression towards the mean" phenomenon, although it is more appropriate to term it an "errors-in-variables" problem (Fuller 1987) as it results

from the assumption that height at base age (site index) is measured without error. This problem could become important when subsets of residuals based on age and site index are plotted. Goelz and Burk (1992) used an ad hoc procedure in an attempt to remove the errors-in-variables problem.

The potential statistical problem of heteroscedasticity and inherent dependence (autocorrelation of the tree sectioning data, caused by taking several measurements from the same trees at different locations) may render the usual regression hypothesis-testing procedures and interval estimations invalid. The problem can be solved by using appropriate generalized least squares techniques (Monseru 1984; Goelz and Burk 1992; Huang 1992). However, since the ordinary nonlinear least squares estimator of the parameters is asymptotically unbiased and consistent even if both heteroscedasticity and autocorrelation are present (Gallant 1987; Judge et al. 1988), the problem should not bias the coefficient estimates of the height growth and site index model. As a result, the estimated coefficients from ordinary nonlinear least squares techniques will still provide unbiased estimates of the predicted values for the dependent variable. The problem of autocorrelations is also ignored, mainly because it does not cause biased parameter estimates and result in biased predictions. This attitude towards autocorrelation has been widely accepted by many researchers in taper modeling, height growth and site index studies and is appropriate provided that only the point estimates are required (Curtis et al. 1974; Monserud 1984; Huang 1997; Kozak 1997).

1.4 Overall research objectives

This study has three objectives: (1) to quantify the relationship between height and DBH (diameter at breast height) of black spruce and jack pine, and to test if separate

height-diameter models are needed for different ecoregions in Manitoba; (2) to test if stand variables can be used, in addition to DBH, to improve height prediction when developing height-DBH models, and (3) to develop reference-age invariant polymorphic height growth and site index models for black spruce and jack pine in Manitoba .

The results of this research are reported in three, relatively independent chapters. Chapter 2 deals with objective (1). Chapter 3 deals with objective (2). Chapter 4 deals with objective (3).

CHAPTER 2. ECOREGION-BASED HEIGHT-DIAMETER MODELS DEVELOPMENT FOR BLACK SPRUCE AND JACK PINE

2.1 Introduction

The prediction of forest growth and yield is the basis for developing forest management plan, analyzing wood supply, and determining provincial annual allowable cut (AAC) to achieve sustainable forest management in Manitoba.

Predicting total tree height based on observed diameter at breast height outside bark (DBH, 1.3 meter above ground) is routinely required in practical management and silvicultural research work, such as volume sampling and pre-harvest surveys. Because data of tree heights are relatively more difficult and costly to obtain, measuring all sample trees for DBH and a portion of these trees for height is a common practice when collecting data from both permanent and temporary sample plots. Currently, foresters in Manitoba are using Alberta provincial black spruce and jack pine height-diameter models to estimate the total tree height because of the lack of locally developed models. As a result, the Manitoba growth and yield program has identified an urgent need to develop local tree height-diameter equations that can be used to estimate tree heights and consequently tree volumes.

A single height-diameter model developed for an entire province may not account for variation among different ecological regions. As a result, applying the provincial model to different ecoregions may result in biased predictions of tree heights (Huang et al. 2000; Peng et al. 2001). For example, the incorrect application of height-diameter models from one ecoregion to another resulted in tree height overestimations between 1.10% and 29.05%, or underestimations between 1.92% and 21.92% (Huang et al. 2000).

The objectives of the this study were (1) to quantify the relationship between height and DBH of black spruce and jack pine in Manitoba, (2) to test if separate height-diameter models are needed for different ecoregions in Manitoba, and (3) to develop ecoregion-based height-diameter models if necessary.

2.2 Study Area

The study included the following ecoregions in Manitoba (Smith et al 1998): Churchill River Upland (Ecoregion 88); Hayes River Upland (Ecoregion 89); Lac Seul Upland (Ecoregion 90); Lake of the Woods (Ecoregion 91); Mid-Boreal Lowland (Ecoregion 148) (Figure 2.1).

The Churchill River Upland Ecoregion is located on the southern edge of the Precambrian Shield. It is characterized by closed Boreal forest of black spruce (*Picea mariana*) and jack pine (*Pinus banksiana*) and white spruce (*Picea glauca*) (Smith et al 1998). It has cool summers and very cold winters. The mean annual temperature is about -2.5°C . The mean summer temperature is 12.5°C and the mean winter temperature is -18.5°C . The mean annual precipitation ranges 400–500 mm (ESWG 1995). This ecoregion is classified as having a subhumid high boreal ecoclimate. It forms part of the continuous coniferous boreal forest that extends from northwestern Ontario to Great Slave Lake in the southern Northwest Territories. The predominant vegetation consists of closed stands of black spruce and jack pine with a shrub layer and a ground cover of mosses and lichens and black spruce is the climatic climax species (ESWG 1995). Permafrost is distributed throughout the ecoregion, but is only widespread in organic deposits. Exposed bedrock occurs throughout the ecoregion and is locally prominent. On level and in depressional areas, Gleysolic soils are associated with clayey sediments, whereas Mesisols and Organic Cryosols are associated with shallow to deep peatlands (ESWG 1995).

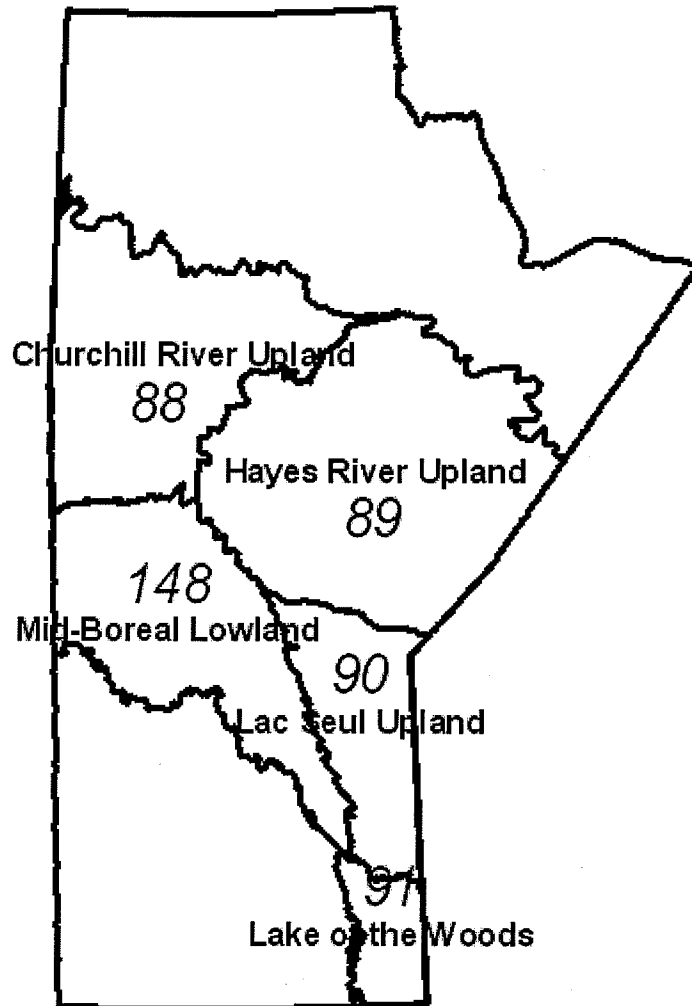


Figure 2.1. Distribution of the five studied ecoregion in Manitoba: Churchill River Upland (Ecoregion 88); Hayes River Upland (Ecoregion 89); Lac Seul Upland (Ecoregion 90); Lake of the Woods (Ecoregion 91); Mid-Boreal Lowland (Ecoregion 148).

The Hayes River Upland Ecoregion is dominated by shallow peat soils and bogs, and by medium tall closed stands of black spruce and jack pine with some paper birch (*Betula papyrifera* Marsh.) (Smith et al.1998). It has cool summers and very cold winters. The mean annual temperature is about -4°C . The mean summer temperature is 11.5°C and the mean winter temperature is -20°C . The mean annual precipitation ranges from 400 mm in the northwest to 600 mm in the southeast (ESWG 1995). This ecoregion is classified as having a sub-humid high boreal ecoclimate. Black spruce is the climatic climax tree species (ESWG 1995). Bedrock exposures have fewer trees and are covered with lichens. Closed to open stands of stunted black spruce with ericaceous shrubs and a ground cover of sphagnum moss dominate poorly drained peat-filled depressions. Permafrost is found throughout the ecoregion, but is only widespread in organic deposits (ESWG 1995). Eutric Brunisols are dominant on calcareous loamy till and calcareous sandy deposits, whereas Dystric Brunisols are associated with noncalcareous fluvio-glacial materials (ESWG 1995).

This Lac Seul Upland Ecoregion is marked by cool summers and very cold winters (ESWG 1995). The mean annual temperature ranges from -4°C to 0.5°C ; the mean summer temperature ranges from 11.5°C to 14°C ; and the mean winter temperature ranges from -20.5°C to -14.5°C . Mean annual precipitation ranges from 350-700 mm, with the wettest areas being in the south-eastern portions of the Lac Seul Upland (ESWG 1995). Permafrost occurs sporadically throughout this ecoregion, except in the area of the Lac Seul Upland, which has a warmer climate (ESWG 1995). Forests of this ecoregion are dominated by stands of black spruce and jack pine, with a shrub layer of ericaceous shrubs (Ericaceae) and a ground cover of moss and lichens (ESWG 1995). Characteristic

vegetation includes white spruce, balsam fir (*Abies balsamea* (L.) Mill.), and black spruce with some trembling aspen (*Populus tremuloides* Michx.) and balsam poplar (*Populus balsamifera* L.), although jack pine and black spruce are more common on moderately well-to imperfectly drained sites. Poorly drained areas are covered by fens and bogs and are dominated by black spruce. Dystric Brunisolic soils are dominant, and Gray Luvisolic and Gleysolic soils occur on finer glaciolacustrine sediments (ESWG 1995).

The Lake of the Woods ecoregion is more closely identified with the warmer, more humid southeastern mixed forest region, than with the colder, drier boreal regions to the north and it has warm summers and cold winters (ESWG 1995). This region is characterized by moist low boreal and subhumid transitional low boreal ecoclimates. The mean annual temperature is about 1.5°C. The mean summer temperature is 15°C and the mean winter temperature is -13°C. The mean annual precipitation ranges from 500 mm to 700 mm (ESWG 1995). Characteristic vegetation includes a succession from trembling aspen, paper birch, and jack pine to white spruce, black spruce, and balsam fir. The ecoregion is underlain by massive, crystalline, acidic, Archean bedrock, forming hummocky, broadly sloping uplands and lowlands. Characteristic vegetation is a mixed forest that includes a succession from trembling aspen, paper birch (*Betula papyrifera*), jack pine, white spruce, black spruce and balsam fir. Forest species assemblages are highly influenced by drainage characteristics and topography (ESWG 1995).

The Mid-Boreal Lowland ecoregion is marked by short, warm summers and cold winters. The mean annual temperature is about -1°C. The mean summer temperature is 13.5°C and the mean winter temperature is -17°C. The mean annual precipitation ranges from 375 mm to 625 mm (ESWG 1995). The cold and poorly drained fens and bogs are

covered with tamarack and black spruce. The mixed deciduous and coniferous forest is characterized by medium to tall, closed stands of trembling aspen and balsam poplar with white spruce, black spruce, and balsam fir occurring in late successional stages (ESWG 1995).

2.3 The data

A total of 125,231 trees were selected from provincial temporary sample plots (TSP), permanent sample plots (PSP), and stem analysis data across the five ecoregions (Ecoregion 88, Ecoregion 89, Ecoregion 90, Ecoregion 91 and Ecoregion 148) in Manitoba (Smith et al 1998). Of these, 74,410 trees are black spruce and 50,821 trees are jack pine. Tree height (m) and outside bark diameter at breast height (DBH, cm) were measured for all sampled trees. Trees with broken tops or forks were not included in this study. The methods used for collecting the data are described in Manitoba Forest Inventory Field Instruction Manual (Manitoba Conservation 1998). The summary statistics, including mean, maximum, minimum, standard deviation of DBH and height are shown in Table 2.1. For model validation purpose, 80 % of all the data were randomly selected and used for model fitting and the remaining 20% of the data were used for model validation.

Table 2.1. Summary of the data used for height-diameter model development by species and ecoregions.

Ecoregion	Spp	MD	MH	MinD	MinH	MaxD	MaxH	StdD	StdH	n
88	BS	9.3	9.3	0.1	1.5	73	24.5	6.26	4.91	10,987
89	BS	6.5	7	0.4	1.5	34.3	29	6.1	5.29	6,512
90	BS	13.2	12.3	0.1	1.5	55.1	28	5.64	3.86	16,870
91	BS	11.3	10.5	0.4	1.5	33.7	23	5.94	4.34	5,811
148	BS	11.4	10.7	0.2	1.5	41.3	26	5.84	4.62	19,346
88	JP	8.7	8.5	0.2	1.5	71	24.5	7.6	5.32	7,541
89	JP	7.3	7.3	0.2	1.5	47	24	6.93	5.45	2,654
90	JP	17.4	14.2	0.3	1.5	53.2	28	6.98	4.22	13,334
91	JP	15.2	12.6	0.3	1.4	48.5	24.5	8.29	4.53	5,373
148	JP	16.5	13.3	0.6	1.5	49.4	25	7.24	4.43	9,557

Note: Spp-species, MD-mean dbh (cm), MH-mean height (m),
 MinD-minimum dbh (cm), MinH-minimum height (m),
 MaxD-maximum dbh (cm), MaxH-maximum height (m),
 StdD- standard deviation of dbh, StdH- standard deviation of height,
 n-sample size (number of trees), BS-black spruce, JP-jack pine.

2.4 Model Selections and Analysis

2.4.1 Selected height-diameter models

Based on literature review, the following eight height–diameter models were selected for evaluation in this study, they include those used by Curtis (1967), Huang et al. (1992), Arabatzis and Burkhart (1992), Moore et al. (1996), Zhang (2002), and Fang and Bailey (1998).

$$[2.1] \quad H = 1.3 + a(1 - e^{-bD})^c \quad (\text{Richards 1959})$$

$$[2.2] \quad H = 1.3 + a(1 - e^{-bD^c}) \quad (\text{Yang et al. 1978})$$

$$[2.3] \quad H = 1.3 + \frac{a}{(1 + \frac{1}{bD^c})} \quad (\text{Ratkowsky and Reedy 1986})$$

$$[2.4] \quad H = 1.3 + ae^{b/(D+c)} \quad (\text{Ratkowsky 1990})$$

$$[2.5] \quad H = 1.3 + ae^{b/D} \quad (\text{Burkhart and Strub 1974; Burk and Burkhart 1984; Buford 1986})$$

$$[2.6] \quad H = 1.3 + e^{(a+b/(D+1))} \quad (\text{Wykoff et al. 1982})$$

$$[2.7] \quad H = 1.3 + D^2 / (a + bD)^2 \quad (\text{Curtis 1967; Prodan 1968})$$

$$[2.8] \quad H = 1.3 + aD / (D + 1) + bD \quad (\text{Watts 1983; Larson 1986})$$

where H is the total height of a tree (m), D is the diameter at breast height outside bark (cm), a , b , c are the parameters to be estimated, e is the base of the natural logarithm ($\cong 2.71828$).

2.4.2 Error structure of the height-diameter models

A fundamental least squares assumption is that the errors are independent and identically distributed with zero mean and constant variance. However, in many forest modeling situations, there is a common pattern that the error variation increases as the values of the dependent variable increase (Huang et al. 1992). This trend is apparent in our data (Figure 2.2). When the problem of unequal error variances occurs, weighted nonlinear least squares (WLS) is applied, with the weights selected to be inversely proportional to the variance of the error term. Based on the studentized residuals analysis (Figure 2.3), a weight factor of $1/dbh$ was found the most suitable and selected in all the WLS. For a correctly identified function, when assumptions of the regression analysis are met, the studentized residuals have zero mean and constant variance. When plotting the studentized residuals against the predicted height, it will show a homogenous band. Studentized residuals are the scaled version of residuals that are obtained by dividing each residual by its standard error (SAS/STAT 1990). They are designed to take into account that un-standardized residuals have intrinsically unequal variances even though the theoretical error term is assumed to have constant variance (Draper and Smith 1981; Rawlings 1988; Neter *et al.* 1990; Huang *et al.* 1992).

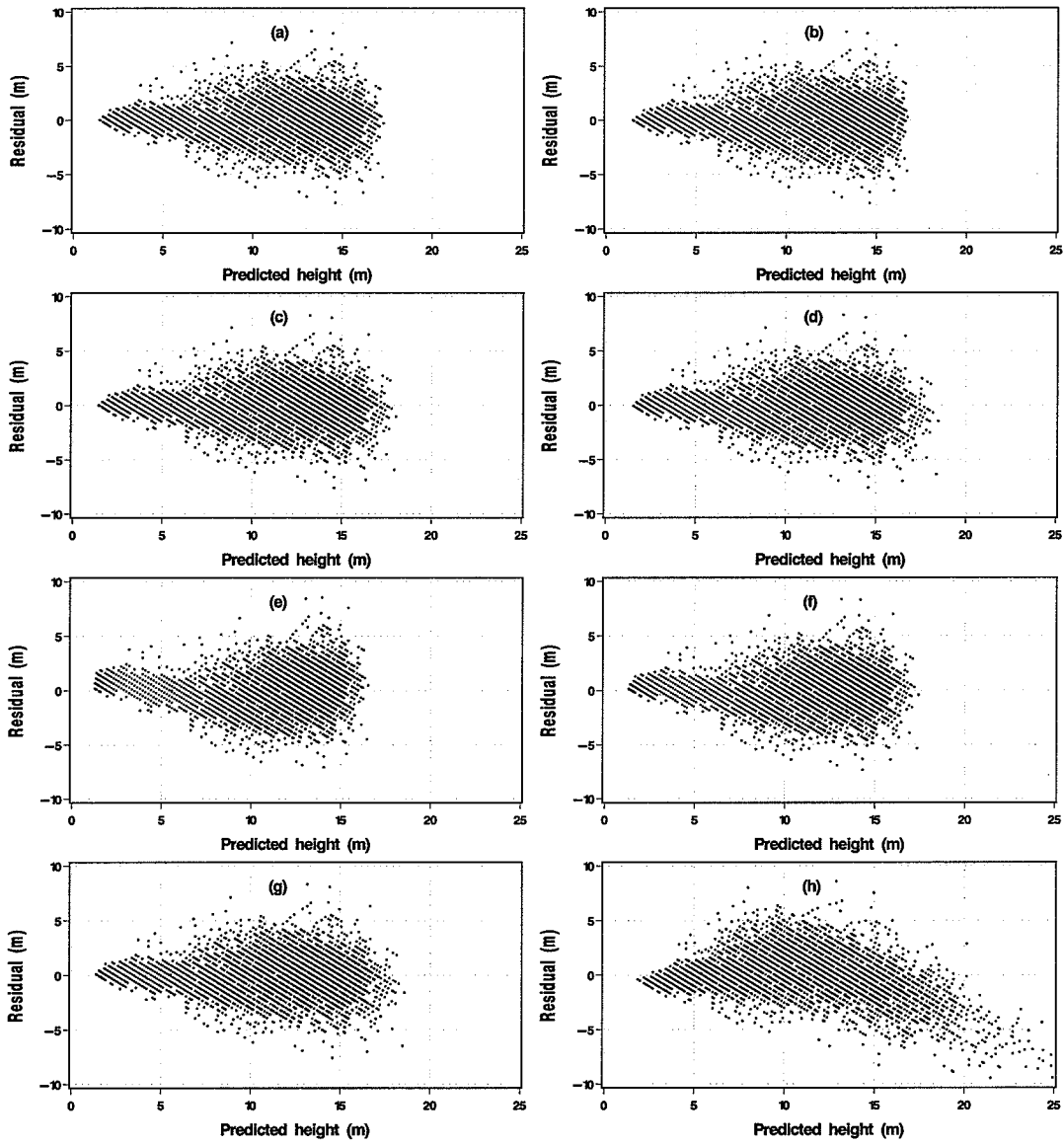


Figure 2.2. Plots of residual against predicted height for black spruce in ecoregion 91 (a, b, c, d, e, f, g, h represent modesl (2.1) , (2.2) , (2.3) , (2.4) , (2.5) , (2.6) , (2.7) , (2.8)).

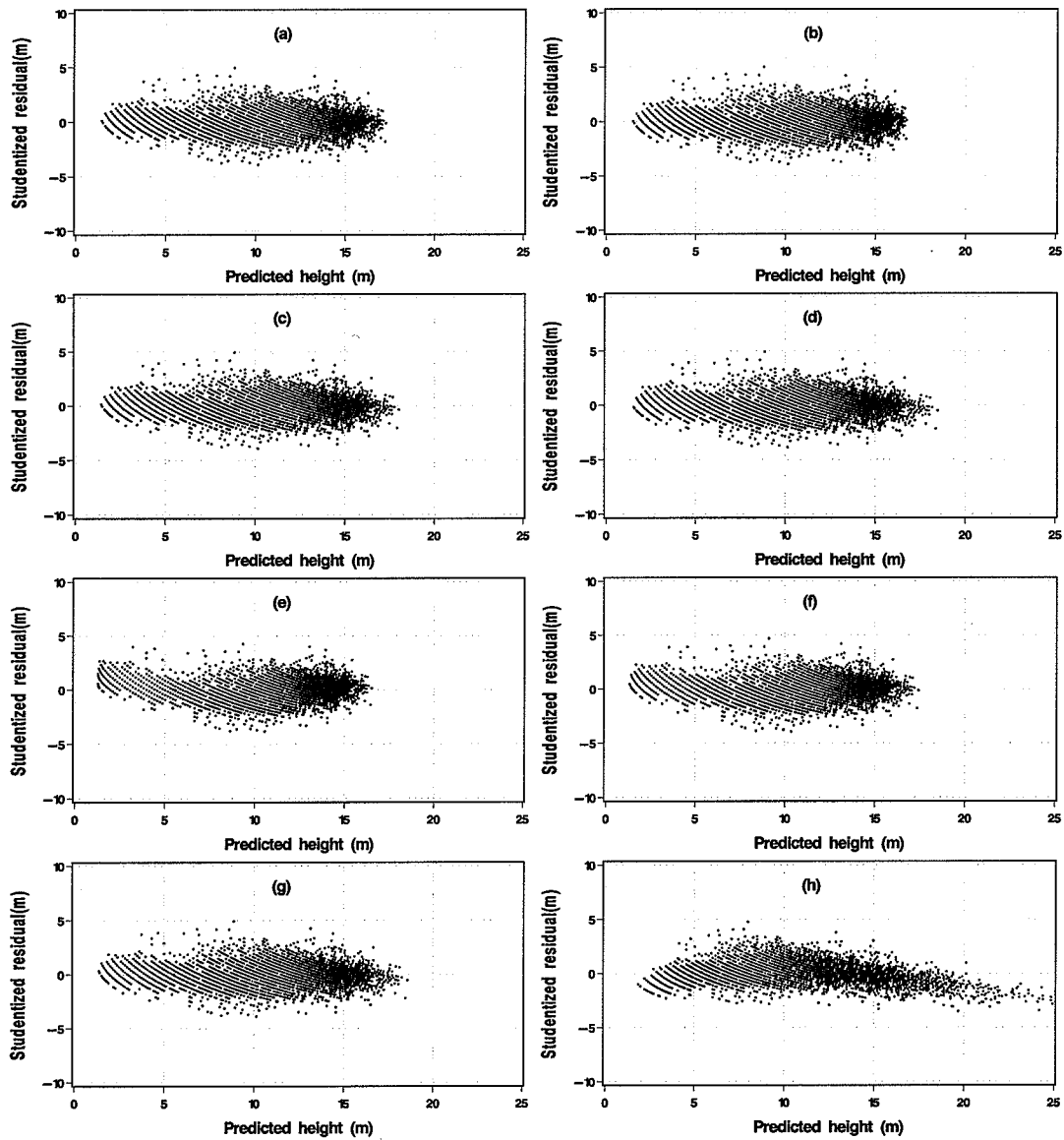


Figure 2.3. Plots of studentized residual against predicted height for black spruce in ecoregion 91 with weight = $1/dbh$ (a, b, c, d, e, f, g, h represent model (2.1) , (2.2) , (2.3) , (2.4) , (2.5) , (2.6) , (2.7) , (2.8)).

2.4.3 Comparison of models among the five eco-regions and between each paired ecoregions

To compare the differences of the height-diameter models among different ecoregions, the nonlinear extra sum of squares method was used (Bates and Watts, 1988; Judge et al., 1988; Huang et al., 2000). This method has been used to compare differences among regionalized taper equations (Huang, 1994), single tree volume equations (Pillsbury et al., 1995) and height-diameter equations (Huang et al., 2000; Zhang et al., 2002).

The method requires the fitting of full and reduced models. The full model corresponds to different sets of parameters for each of the ecoregions involved, while the reduced model corresponds to the same set of parameters for all ecoregions. By using the indicator (dummy) variables approach (Bates and Watts, 1988; Huang et al., 2000; Zhang et al., 2002), the full model of the Weibull-type function [Eq.2.2], for the 5 ecoregions is written as:

$$[2.9] \quad H=13+(a+a_1x_1+a_2x_2+a_3x_3+a_4x_4)\left[1-e^{-(b+b_1x_1+b_2x_2+b_3x_3+b_4x_4)D^{(c+c_1x_1+c_2x_2+c_3x_3+c_4x_4)}}\right]$$

Four indicator variables (x_1 - x_4) are needed in equation [2.9] for five ecoregions. They are defined as follows:

If ecoregion=88 then $x_1=1, x_2=0, x_3=0, x_4=0$;

If ecoregion=89 then $x_1=0, x_2=1, x_3=0, x_4=0$;

If ecoregion=90 then $x_1=0, x_2=0, x_3=1, x_4=0$;

If ecoregion=91 then $x_1=0, x_2=0, x_3=0, x_4=1$;

If ecoregion=148 then $x_1=0, x_2=0, x_3=0, x_4=0$.

The full model has 15 parameters ($m_F = 15$). It has an error sum of squares denoted as SSE (F), with $df_F = (n-15)$ degrees of freedom, where n is the total number of observations from all five ecoregions combined. The reduced model of the height-diameter function, for which the same set of parameters were used for all five ecoregions, take the form of function [2.2], with its error sum of squares denoted as SSE(R). There are $df_R = (n-3)$ degrees of freedom associated with SSE(R), where the number of parameters is 3 ($m_R = 3$) in function [2.2]. Equality of the two models, [2.2] and [2.10], is tested by considering the following hypotheses:

$$H_0 : a_1 = a_2 = a_3 = a_4 = b_1 = b_2 = b_3 = b_4 = c_1 = c_2 = c_3 = c_4 = 0 \text{ against}$$

H_a : At least one of the equalities in H_0 is not true.

The appropriate test statistic is an F-test given by Bates and Watts (1988):

$$[2.10] \quad F = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

If $F > F_{critical} (1 - \alpha ; df_R - df_F, df_F)$, reject H_0 , separate models are required for separate ecoregions; If $F \leq F_{critical} (1 - \alpha ; df_R - df_F, df_F)$, accept H_0 , the reduced model is appropriate for combined ecoregions.

For the paired two ecoregion comparisons (total 10 pairs), by using the indicator (dummy) variables approach, the full model of the Weibull-type model (Eq.2.2) for the paired two ecoregions are written as:

$$[2.11] \quad H = 1.3 + (a + a_1 x_1) [1 - e^{-(b + b_1 x_1) D^{(c + c_1 x_1)}}]$$

One indicator variables x_1 is needed in the equation for the two ecoregions. They are defined as follows:

If ecoregion=90 then $x_1=0$;

If ecoregion=91 then $x_1=1$.

The full model has 6 parameters ($m_F = 6$). It has an error sum of squares denoted as $SSE(F)$, with $df_F = (n-6)$ degrees of freedom, where n is the total number of observations from all two ecoregions combined. The reduced model for which the same set of parameters is used for the two ecoregions, take the form of equation [2.2], with its error sum of squares denoted as $SSE(R)$. There are $df_R = (n-3)$ degrees of freedom associated with $SSE(R)$, where the number of parameters is 3 ($m_R = 3$) in (eq.2.2). Equality of the two models, equation [2.2] and equation [2.11], is tested by considering the following hypotheses:

$H_0 : a_1 = b_1 = c_1 = 0$ against

H_a : At least one of the equalities in H_0 is not true.

Using the F test statistic calculated by [2.10],

If $F > F_{critical} (1 - \alpha ; df_R - df_F, df_F)$, reject H_0 , separate models are required for separate ecoregions; If $F \leq F_{critical} (1 - \alpha ; df_R - df_F, df_F)$, accept H_0 , the reduced model is appropriate for combined ecoregions.

2.4.4 Model fitting

All models were fitted using the Gauss-Newton method of the SAS non-linear least squares procedures (SAS/STAT, SAS Institute Inc.1990). To ensure the solutions of the non-linear regressions are global rather local least squares solutions, different starting values of the model parameters were provided for the fits. Following the model fitting, studentized residual plots were examined for outliers, lack-of-fit and unequal variance. After the residual analysis, weighted non-linear least squares were applied to refit the

models. Models were compared using the R-squared values (R^2), root mean squared error (RMSE), average bias (\bar{e}), bias % (bias_per) and mean absolute deviation (MAD), which are calculated as:

$$[2.12] \quad R^2 = 1 - \frac{\sum_{i=1}^n w_i (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n w_i (Y_i - \bar{Y})^2}$$

$$[2.13] \quad RMSE = \sqrt{\frac{\sum_{i=1}^n w_i (y_i - \hat{y}_i)^2}{n-m}}$$

$$[2.14] \quad \bar{e} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)}{n}$$

$$[2.15] \quad MAD = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n}$$

$$[2.16] \quad bias_per = (\bar{e} / \bar{y}) \times 100 = \frac{\bar{e}}{\sum y_i / n} \times 100$$

where y_i is the measured height, \hat{y}_i is the estimated height, and \bar{y} is the average of the measured heights, w_i is the weight, n is the number of observations, m is the number of parameters in the model.

A t-test was used to test the null hypothesis that the mean prediction error equals zero by:

$$[2.17] \quad S_e = \sqrt{\frac{\sum_{i=1}^n w_i (e_i - \bar{e})^2}{n-1}}$$

$$[2.18] \quad t = \frac{\bar{e}}{S_e / \sqrt{n}}$$

where e_i is the i th prediction error, S_e is the standard deviation of the prediction errors and n is the number of observations.

2.4.5 Model validation

The ultimate goal of model validation is to increase the credibility and gain sufficient confidence about a model. There are many model validation methods, but no one is universally “good”. Several methods should be applied in order to evaluate the fitted models thoroughly. The commonly used prediction statistics such as the R-squared values (R_p^2), root mean squared error ($RMSE_p$), mean bias (\bar{e}), bias % (bias_per), mean absolute deviation (MAD), standard deviation of the prediction errors (S_e) were also used to evaluate the fitted models using the validation data set. Both mean bias and bias % give an average measure of the differences between the predicted and the observed tree height, with mean prediction bias in absolute terms and percent bias in relative terms. RMSE incorporates both bias and variation and is a better measure of a model performance. In addition, the fitted height-diameter curves were plotted and examined to ensure the fitted model did not violate biological meaning, i.e., its curve should be a sigmoid shape.

2.5 Results

2.5.1 Model fitting and comparison for each ecoregion

2.5.1.1 Ecoregion 88

The fitted nonlinear least squares estimates of the parameters, root mean squared error (RMSE, the coefficient of determination (R^2)) are shown in Table 2.2. The asymptotic t-statistics, average bias (\bar{e}), mean absolute deviation (MAD) are shown in Table 2.3.

Judging from the results shown in Table 2.2 and Table 2.3, model [2.2], [2.1], [2.3], [2.4] performed quite well for black spruce. The R^2 values were high and the RMSE were small, and model [2.2] had the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were significant at $\alpha=0.05$ level. The bias_per of model [2.1]-[2.4] are less than 0.03%, with model [2.2] being the lowest (0.0052%). Model [2.5]-[2.8] showed higher bias_per, greater than 0.9%. The paired t tests of model [2.1]-[2.4] and [2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. Model [2.5] got the highest bias_per 2.21%. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per is about 0.000%, the MAD and Se values are the highest, 1.3039 and 1.9160, respectively. Furthermore, model [2.8] graphically displayed like a line, which does not make biological sense. Overall, model [2.2] performed the best. The graphical presentation of model [2.2] is given in Figure 2.4, along with height and diameter data. The studentized residual plot of model [2.2] is given in Figure 2.5.

For jack pine, model [2.3], [2.1], [2.2], performed quite well. The R^2 values were high and the RMSE were small, and model [2.3] had the highest R^2 and lowest RMSE.

All the t-statistics for the parameters of these eight models were significant at $\alpha=0.05$ level. The bias_per of model [2.1]-[2.3] are less than 0.03% and model [2.3] was the lowest (0.022%). Model [2.5]-[2.7] showed higher bias_per, greater than 0.5%. The paired t tests of model [2.1]-[2.4] and [2.8] showed that the mean prediction bias are not significantly different from zero at $\alpha=0.05$ level. Model [2.5] got the highest bias_per 3.35 %. For model [2.8], although the t-statistics for the parameters are significant at $\alpha=0.05$ level and Bias_per is about 0.000%, but the mean absolute deviation (MAD) and Se values are the highest, 1.4205 and 2.1740, respectively. Furthermore, model [2.8] graphically displayed like a line, which does not make biologically sense. Overall, model [2.3] and [2.2] performed the best. The graphically presentation of model [2.3] is given in Figure 2.6, along with height and diameter data. The studentized residual plot of model [2.3] is given in Figure 2.7, showing an almost homogenous band of the studentized residuals with zero mean, and it indicates the assumptions of the regression analysis are almost met.

Table 2.2. Estimated parameters and RMSE of models [2.1] to [2.8] for black spruce and jack pine in ecoregion 88.

Spp	Model	a	b	c	n	RMSE	R²
BS	[2.1]	19.3035	0.0865	1.3038	10987	0.4911	0.977
BS	[2.2]	18.0804	0.0433	1.2345	10987	0.4906	0.977
BS	[2.3]	24.9712	0.0305	1.3007	10987	0.4919	0.977
BS	[2.4]	26.149	-12.762	2.6367	10987	0.4936	0.977
BS	[2.5]	17.9247	-5.9035		10987	0.6054	0.965
BS	[2.6]	3.0127	-8.0298		10987	0.5329	0.973
BS	[2.7]	1.2095	0.2013		10987	0.5048	0.976
BS	[2.8]	0.5512	0.8135		10987	0.5705	0.969
JP	[2.1]	18.7595	0.0878	1.2627	7541	0.5148	0.977
JP	[2.2]	18.2047	0.0479	1.1874	7541	0.5165	0.976
JP	[2.3]	23.5853	0.0353	1.2824	7541	0.5135	0.977
JP	[2.4]	23.4709	-10.7775	2.1532	7541	0.5193	0.976
JP	[2.5]	16.705	-5.1587		7541	0.6263	0.965
JP	[2.6]	2.9693	-7.4739		7541	0.5466	0.974
JP	[2.7]	1.1164	0.209		7541	0.5218	0.976
JP	[2.8]	0.8854	0.7462		7541	0.6387	0.964

Note: RMSE - root mean square error of prediction; R² - prediction coefficient of determination.

Table 2.3. Prediction statistics of models [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 88.

Spp	Model	\bar{e}	MAD	Bias_per	Se	t	n
BS	[2.1]	-0.0015	1.075	-0.0156	1.453	-0.1053	10987
BS	[2.2]	-0.0005	1.073	-0.0052	1.451	-0.0348	10987
BS	[2.3]	-0.0026	1.078	-0.0283	1.458	-0.1899	10987
BS	[2.4]	0.0013	1.078	0.0136	1.455	0.0913	10987
BS	[2.5]	0.2058	1.28	2.2057	1.646	13.109	10987
BS	[2.6]	0.0917	1.159	0.9831	1.536	6.261	10987
BS	[2.7]	0.0309	1.093	0.331	1.467	2.2077	10987
BS	[2.8]	0	1.304	0	1.916	0	10987
JP	[2.1]	0.0026	1.098	0.03	1.543	0.1437	7541
JP	[2.2]	0.0019	1.101	0.0225	1.544	0.1079	7541
JP	[2.3]	0.0019	1.096	0.022	1.546	0.1055	7541
JP	[2.4]	0.0238	1.11	0.2792	1.565	1.3199	7541
JP	[2.5]	0.2854	1.365	3.3488	1.893	13.0952	7541
JP	[2.6]	0.124	1.187	1.4544	1.691	6.3642	7541
JP	[2.7]	0.0493	1.115	0.5779	1.583	2.7029	7541
JP	[2.8]	0	1.421	0	2.174	0	7541

Note: \bar{e} - average bias; MAD - mean absolute deviation; bias_per – average bias percent; Se - standard deviation of the prediction errors; t-the paired t test statistics.

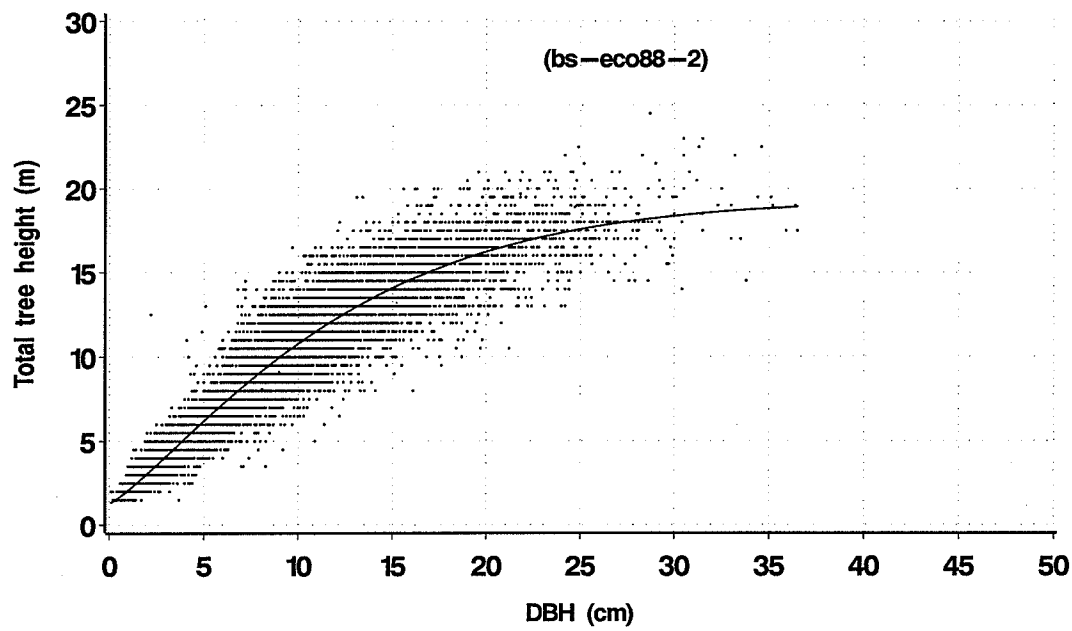


Figure 2.4. Plots of total tree height against DBH for black spruce in ecoregion 88 based on model [2.2].

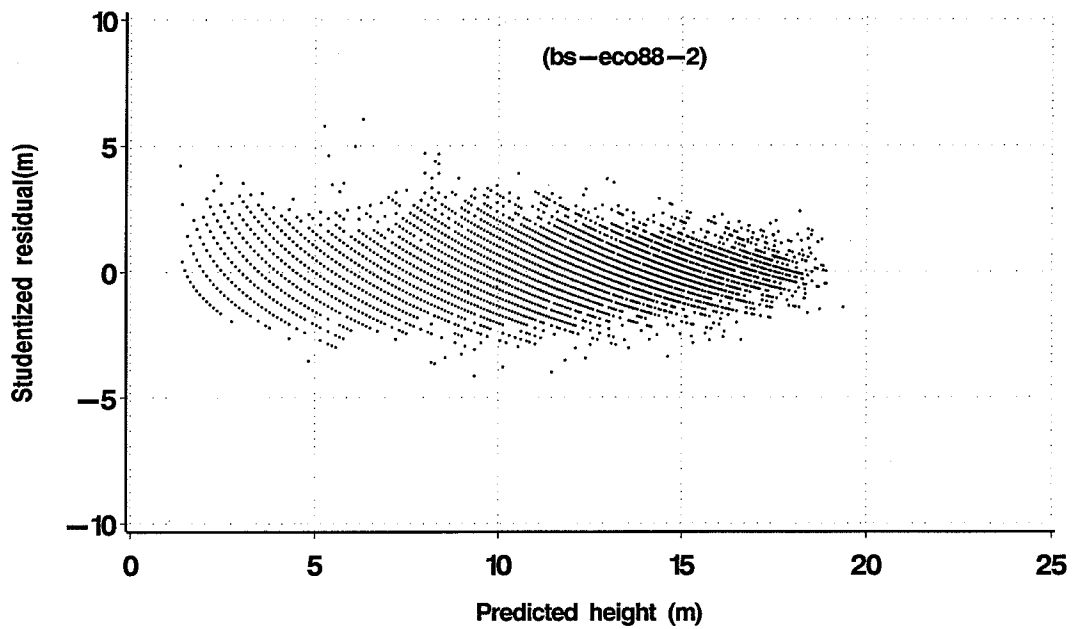


Figure 2.5. Plots of studentized residual against predicted height for black spruce in ecoregion 88 based on model [2.2].

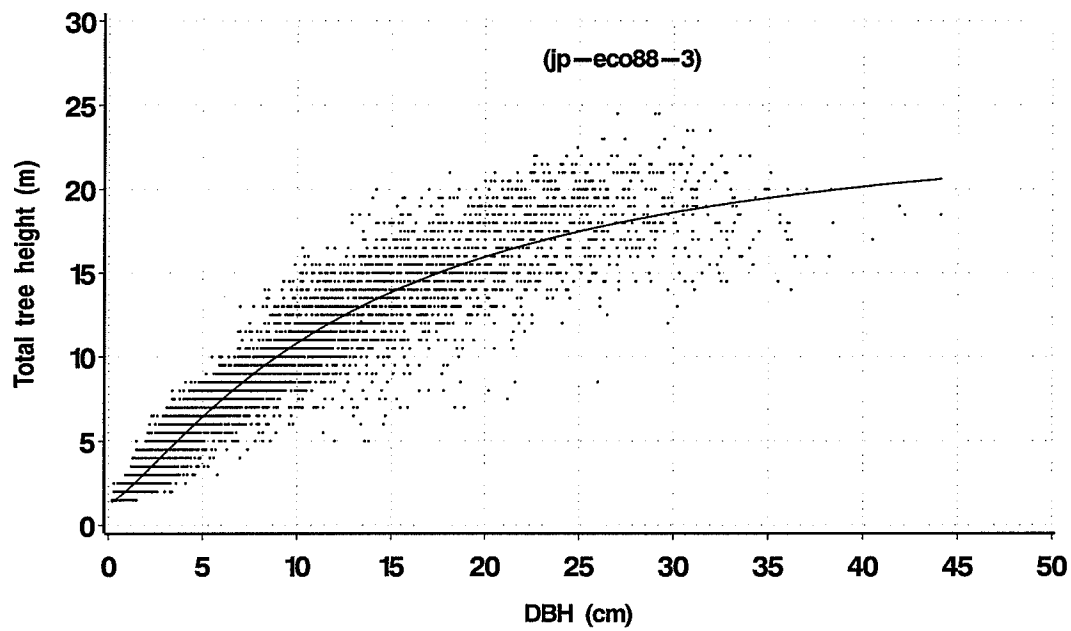


Figure 2.6. Plots of total tree height against DBH for jack pine in ecoregion 88 based on model [2.3].

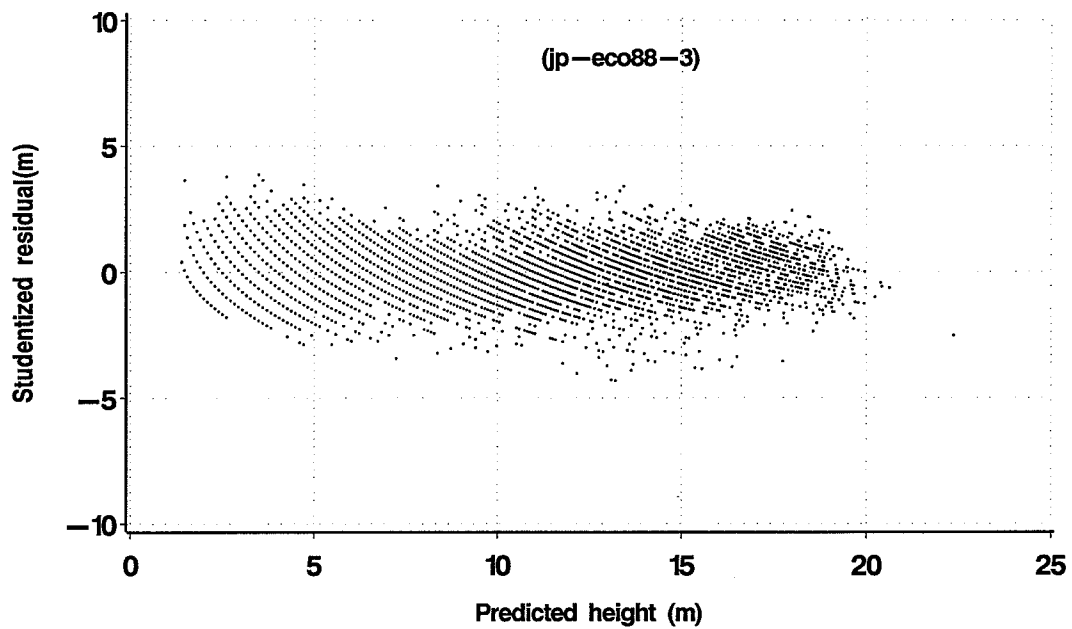


Figure 2.7. Plots of studentized residual against predicted height for jack pine in ecoregion 88 based on model [2.3].

2.5.1.2 Ecoregion 89

The fitted nonlinear least squares estimates of the parameters, root mean squared error (RMSE), and the coefficient of determination (R^2) are shown in Table 2.4. The asymptotic t-statistics, average bias (\bar{e}), mean absolute deviation (MAD) are shown in Table 2.5.

Judging from the results shown in Table 2.4 and Table 2.5, model [2.3], [2.1], [2.4], [2.2] performed quite well for black spruce. The R^2 values were higher and the RMSE were smaller, and model [2.3] had the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models are significant at $\alpha=0.05$ level. The bias_per of model [2.1]-[2.4] are less than 0.09%, with model [2.2] being the lowest (0.0108%). Model [2.5]-[2.7] showed higher bias_per, greater than 0.45%. The paired t tests of model [2.1]-[2.4] and [2.8] showed that the mean prediction bias are not significantly different from zero at $\alpha=0.05$ level. Model [2.5] had the highest bias_per 3.63%. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, the mean absolute deviation (MAD) and Se values were relatively higher, 0.8587 and 1.3809, respectively. Furthermore, model [2.8] graphically displayed like a line, which does not make biological sense. Overall, model [2.3] and [2.1] performed the best.. The graphically presentation of model [2.3] is given in Figure 2.8, along with height and diameter data. The studentized residual plot of model [2.3] is given in Figure 2.9.

For jack pine, model [2.2], [2.1], [2.4], [2.7] performed quite well. The R^2 values are higher and the RMSE were smaller, and model [2.4] had the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were significant at

$\alpha=0.05$ level. The bias_per of model [2.1]-[2.2] were less than 0.08%, with model [2.2] being the lowest (0.051%). Model [2.5]-[2.6] showed higher bias_per, greater than 1.1%. The paired t tests of model [2.1]-[2.4] and [2.7]-[2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. Model [2.5] got the highest bias_per (3.1047 %). For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, but the mean absolute deviation (MAD) and Se values were the highest, 1.1892 and 1.9150 respectively. Furthermore, model [2.8] graphically displayed like a line, which does not make biological sense. Overall, model [2.2] and [2.1] performed the best. The graphically presentation of model [2.2] is given in Figure 2.10, along with height and diameter data. The studentized residual plot of model [2.2] is given in Figure 2.11, showing an almost homogenous band of the studentized residuals with zero mean, and it indicates the assumptions of the regression analysis are almost met.

Table 2.4. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 89.

Spp	Model	a	b	c	n	RMSE	R²
BS	[2.1]	21.3847	0.0833	1.4029	6512	0.4051	0.9797
BS	[2.2]	19.7083	0.0331	1.3053	6512	0.4063	0.9796
BS	[2.3]	27.0978	0.0232	1.3743	6512	0.4046	0.9798
BS	[2.4]	28.843	-13.7336	2.633	6512	0.4055	0.9797
BS	[2.5]	17.661	-5.5407		6512	0.5518	0.9623
BS	[2.6]	3.0571	-8.2435		6512	0.4519	0.9747
BS	[2.7]	1.286	0.1904		6512	0.4139	0.9788
BS	[2.8]	-0.3066	0.9059		6512	0.4566	0.9742
JP	[2.1]	18.072	0.1045	1.4798	2654	0.4612	0.9866
JP	[2.2]	17.2036	0.0374	1.3325	2654	0.4657	0.9863
JP	[2.3]	21.5955	0.0277	1.4555	2654	0.4588	0.9867
JP	[2.4]	24.1255	-10.9635	1.8955	2654	0.4544	0.987
JP	[2.5]	17.6844	-5.7836		2654	0.5419	0.9815
JP	[2.6]	3.0184	-8.124		2654	0.4751	0.9858
JP	[2.7]	1.2438	0.1991		2654	0.4575	0.9868
JP	[2.8]	-0.1471	0.8471		2654	0.5686	0.9796

Note: RMSE - root mean square error of prediction; R² - prediction coefficient of determination.

Table 2.5. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 89.

Spp	Model	\bar{e}	MAD	Bias_per	Se	t	n
BS	[2.1]	0.001	0.7339	0.0142	1.1227	0.0709	6512
BS	[2.2]	0.0008	0.7365	0.0108	1.1279	0.0539	6512
BS	[2.3]	0.0008	0.7332	0.0118	1.1201	0.0593	6512
BS	[2.4]	0.0056	0.7345	0.0808	1.1267	0.403	6512
BS	[2.5]	0.2533	1.0271	3.6366	1.4952	13.6711	6512
BS	[2.6]	0.0977	0.8369	1.4024	1.2821	6.1483	6512
BS	[2.7]	0.0327	0.7511	0.4688	1.1453	2.3009	6512
BS	[2.8]	0	0.8587	0	1.3809	0	6512
JP	[2.1]	0.0057	0.9432	0.0772	1.4013	0.2084	2654
JP	[2.2]	0.0037	0.9538	0.0505	1.4111	0.1352	2654
JP	[2.3]	0.006	0.937	0.082	1.3967	0.222	2654
JP	[2.4]	0.0118	0.9282	0.1602	1.3997	0.4329	2654
JP	[2.5]	0.2279	1.0704	3.1047	1.6054	7.3133	2654
JP	[2.6]	0.0878	0.9607	1.1955	1.4765	3.0621	2654
JP	[2.7]	0.0224	0.9281	0.3048	1.3927	0.8277	2654
JP	[2.8]	0	1.1892	0	1.915	0	2654

Note: \bar{e} - average bias; MAD - mean absolute deviation; bias_per – average bias percent; Se - standard deviation of the prediction errors; t-the paired t test statistics.

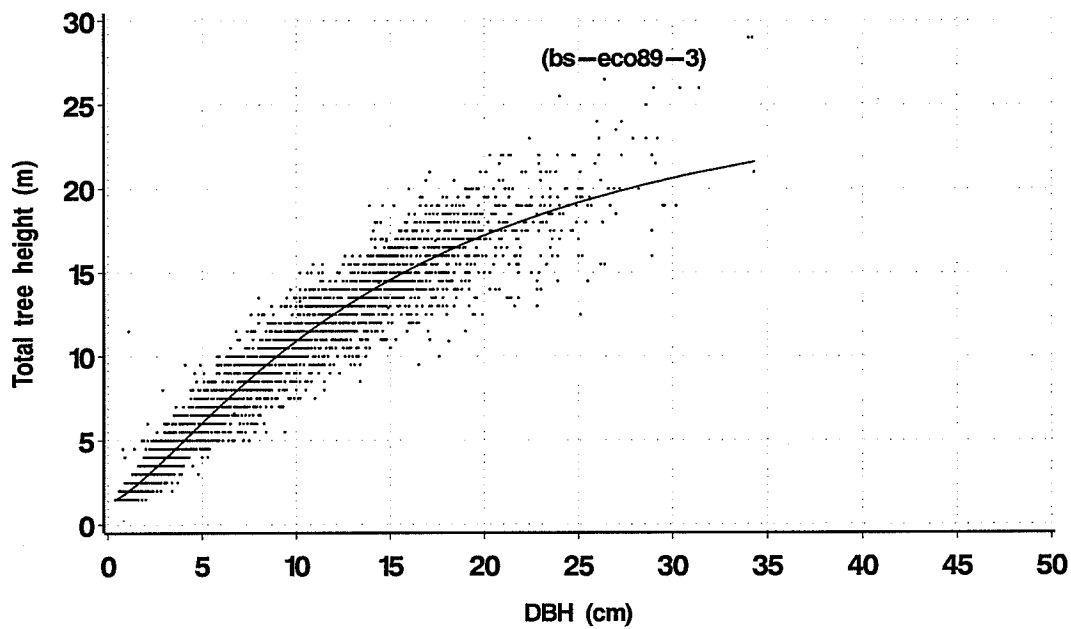


Figure 2.8. Plots of total tree height against DBH for black spruce in ecoregion 89 based on model [2.3].

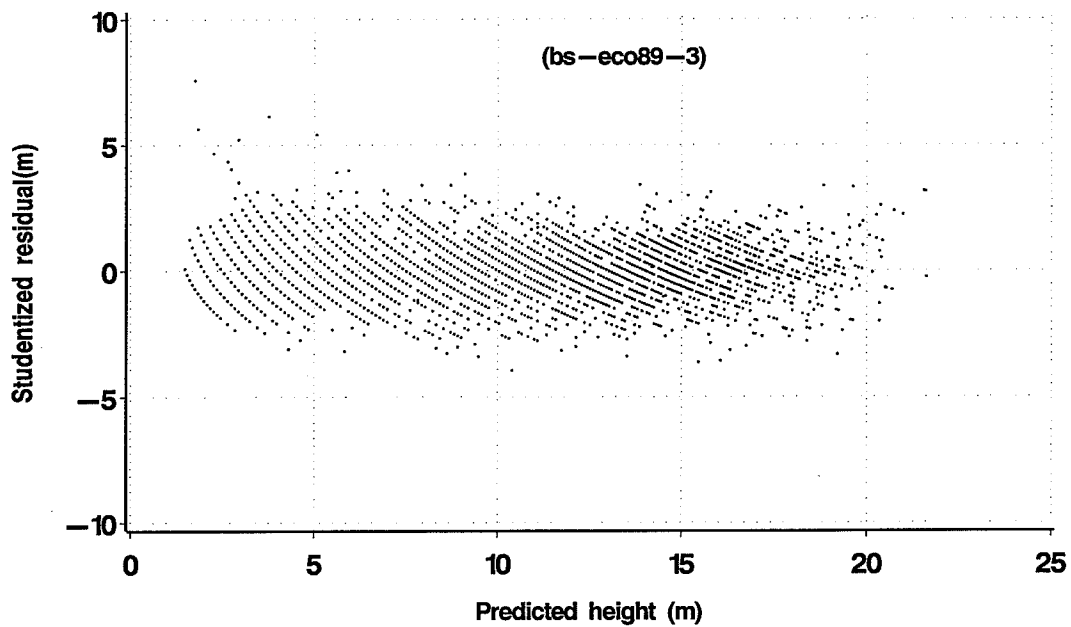


Figure 2.9. Plots of studentized residual against predicted height for black spruce in ecoregion 89 based on model [2.3].

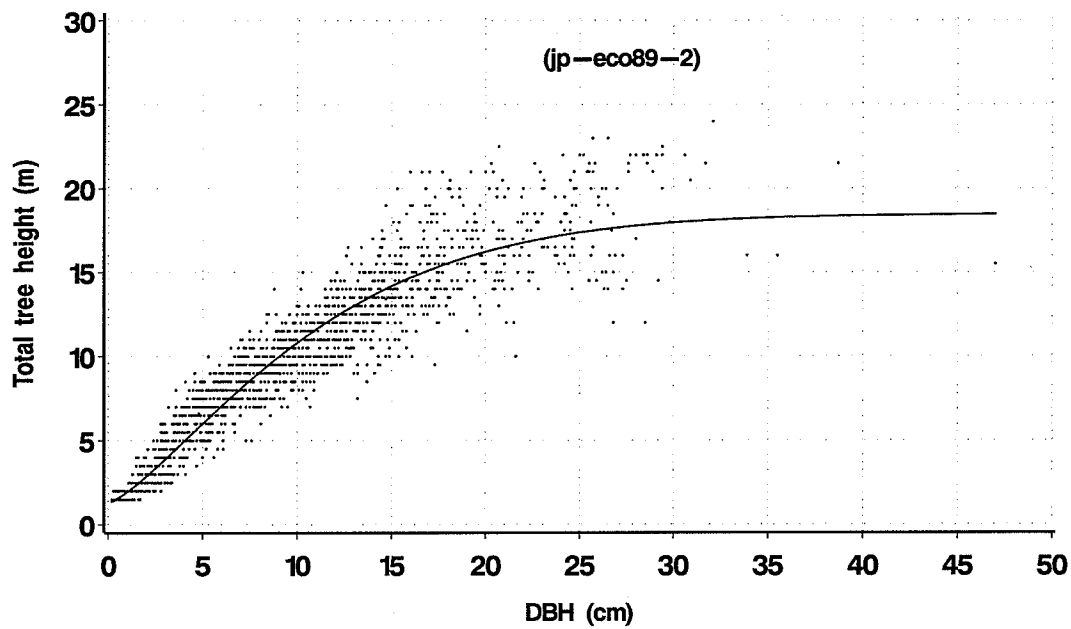


Figure 2.10. Plots of total tree height against DBH for jack pine in ecoregion 89 based on model [2.2].

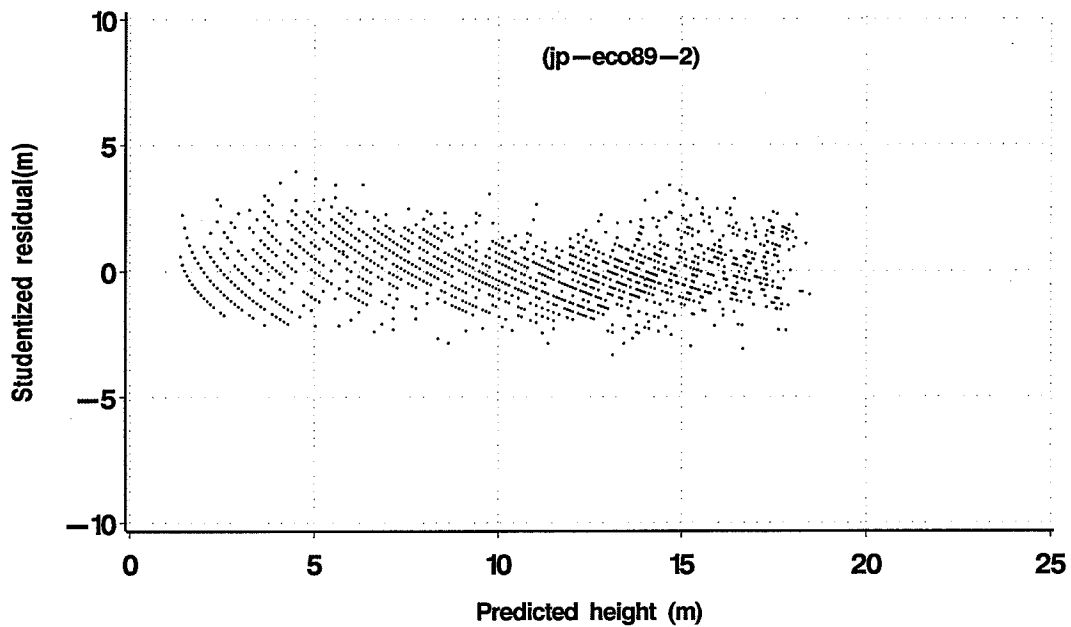


Figure 2.11. Plots of studentized residual against predicted height for jack pine in ecoregion 89 based on model [2.2].

2.5.1.3 Ecoregion 90

The fitted nonlinear least squares estimates of the parameters, root mean squared error (RMSE), the coefficient of determination (R^2) are shown in Table 2.6. The asymptotic t-statistics, average bias (\bar{e}), mean absolute deviation (MAD) are shown in Table 2.7.

Judging from the results shown in Table 2.6 and Table 2.7, model [2.4], [2.7], [2.3], [2.1], [2.2] performed quite well for black spruce. The R^2 values were higher and the RMSE were smaller, and model [2.4] had the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were significant at $\alpha=0.05$ level. The bias_per of model [2.1]-[2.4] and [2.7] were less than 0.03% and model [2.4] was the lowest 0.008%. Model [2.5]-[2.6] showed higher bias_per, greater than 0.15%. The paired t tests of model [2.1]-[2.4] and [2.6]-[2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. Model [2.5] got the highest bias_per 0.366 %. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, but the mean absolute deviation (MAD) and Se values were relatively higher, 1.6618 and 2.1938 respectively. Further more, model [2.8] graphically displayed like a line, which does not make biologically sense. Overall, model [2.4] and [2.2] performed the best. The graphically presentation of model [2.4] is given in Figure 2.12, along with height and diameter data. The studentized residual plot of model [2.4] is given in Figure 2.13.

For jack pine, model [2.3], [2.1], [2.2] performed quite well. The R^2 values were higher and the RMSE were smaller, model [2.3] showed the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were significant at

$\alpha=0.05$ level. The bias_per of model [2.1]-[2.3] were less than 0.009% and model [2.3] showed the lowest 0.0022%. Model [2.5]-[2.6] showed higher bias_per, greater than 0.3%. The paired t tests of model [2.1]-[2.4] and [2.7]-[2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. Model [2.5] got the highest bias_per 0.5027 %. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, but the mean absolute deviation (MAD) and Se values were the highest, 2.2349 and 2.8574 respectively. Further more, model [2.8] graphically displayed like a line, which does not make biologically sense. Overall, the model [2.3], [2.1] and [2.2] performed the best. The graphically presentation of model [2.3] is given in Figure 2.14, along with height and diameter data. The studentized residual plot of model [2.3] is given in Figure 2.15, showing an almost homogenous band of the studentized residuals with zero mean, and it indicates the assumptions of the regression analysis are almost met.

Table 2.6. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 90.

Spp	Model	a	b	c	n	RMSE	R2
BS	[2.1]	18.5518	0.1001	1.486	16870	0.5444	0.887
BS	[2.2]	17.9268	0.0378	1.2911	16870	0.5454	0.8866
BS	[2.3]	21.9847	0.0242	1.4901	16870	0.5437	0.8872
BS	[2.4]	25.743	-12.0679	2.0284	16870	0.5432	0.8875
BS	[2.5]	21.9824	-8.2368		16870	0.5547	0.8826
BS	[2.6]	3.1668	-10.0207		16870	0.5459	0.8863
BS	[2.7]	1.3437	0.1916		16870	0.5434	0.8874
BS	[2.8]	2.4351	0.6622		16870	0.6129	0.8567
JP	[2.1]	20.2479	0.0714	1.1696	13334	0.6371	0.8613
JP	[2.2]	19.9286	0.048	1.1116	13334	0.6373	0.8612
JP	[2.3]	25.4147	0.0331	1.24	13334	0.6369	0.8614
JP	[2.4]	26.2484	-13.3759	2.8959	13334	0.638	0.8609
JP	[2.5]	22.195	-8.4359		13334	0.6577	0.8522
JP	[2.6]	3.1502	-9.8907		13334	0.6463	0.8572
JP	[2.7]	1.3059	0.1963		13334	0.6384	0.8607
JP	[2.8]	3.7918	0.5374		13334	0.6993	0.8328

Note: RMSE - root mean square error of prediction; R^2 - prediction coefficient of determination.

Table 2.7. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 90.

Spp	Model	\bar{e}	MAD	Bias_per	Se	t	n
BS	[2.1]	0.0028	1.477	0.0226	1.9253	0.187	16870
BS	[2.2]	0.0028	1.48	0.023	1.9276	0.1905	16870
BS	[2.3]	0.0018	1.477	0.0144	1.9227	0.1194	16870
BS	[2.4]	0.001	1.476	0.0082	1.9216	0.0683	16870
BS	[2.5]	0.045	1.5	0.3661	1.9475	3.0012	16870
BS	[2.6]	0.0193	1.481	0.1574	1.9285	1.3028	16870
BS	[2.7]	-0.0028	1.477	-0.0229	1.9217	-0.1903	16870
BS	[2.8]	0	1.662	0	2.1938	0	16870
JP	[2.1]	0.0011	2.057	0.0081	2.6186	0.0506	13334
JP	[2.2]	0.0011	2.057	0.0079	2.6186	0.0493	13334
JP	[2.3]	-0.0003	2.058	-0.0022	2.6192	-0.0136	13334
JP	[2.4]	0.005	2.059	0.0354	2.6205	0.2218	13334
JP	[2.5]	0.0714	2.095	0.5027	2.6573	3.1048	13334
JP	[2.6]	0.0458	2.075	0.3221	2.6393	2.003	13334
JP	[2.7]	0.012	2.06	0.0844	2.6223	0.5282	13334
JP	[2.8]	0	2.235	0	2.8574	0	13334

Note: \bar{e} - average bias; MAD - mean absolute deviation; bias_per – average bias percent; Se - standard deviation of the prediction errors; t-the paired t test statistics.

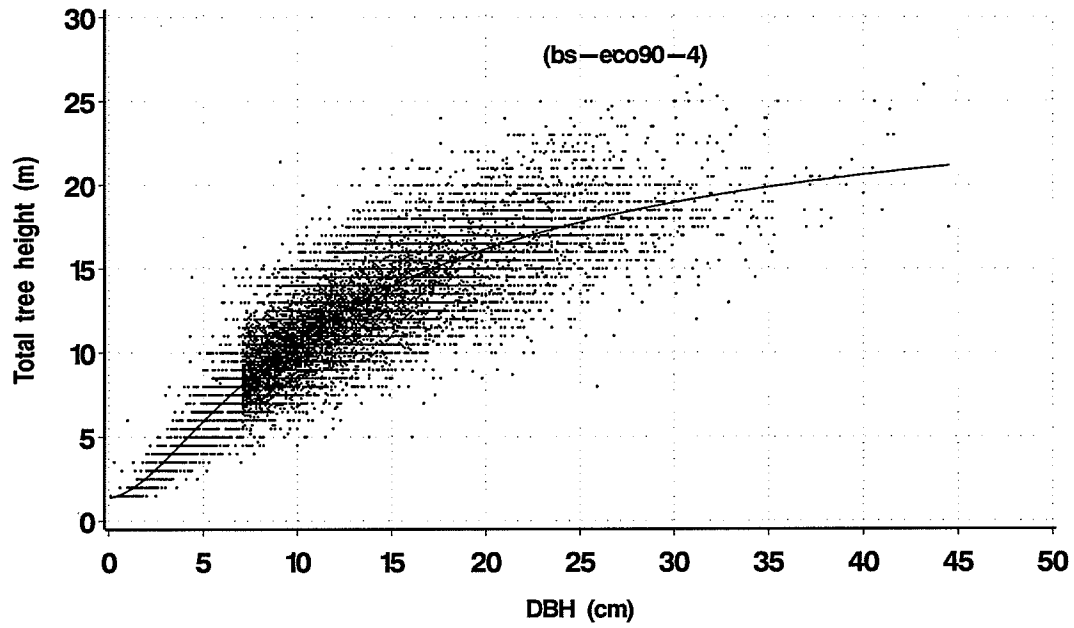


Figure 2.12. Plots of total tree height against DBH for black spruce in ecoregion 90 based on model [2.4].

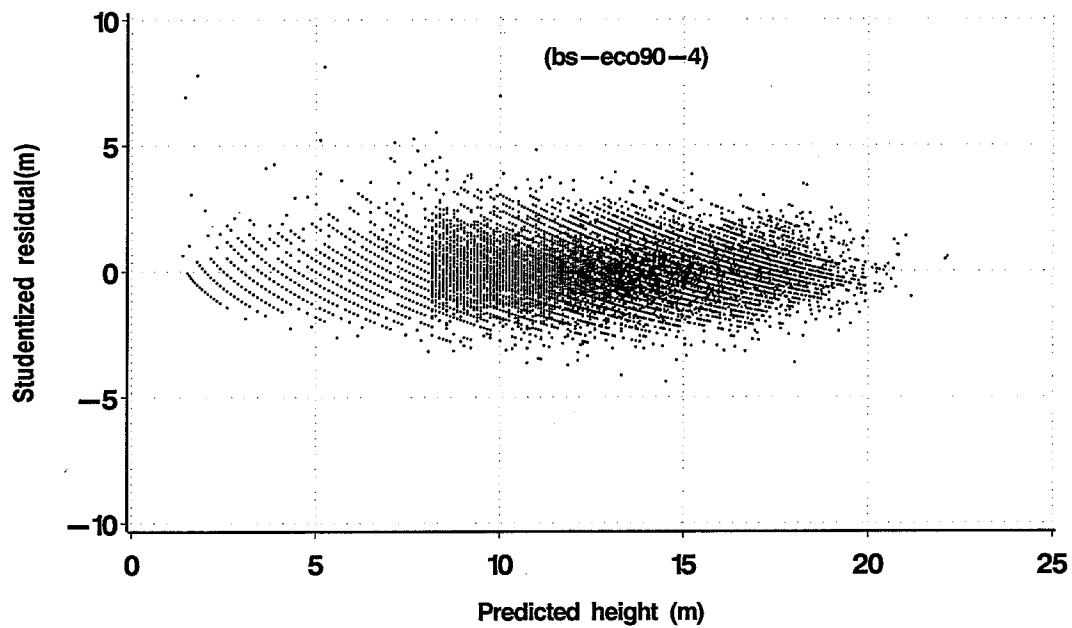


Figure 2.13. Plots of studentized residual against predicted height for black spruce in ecoregion 90 based on model [2.4].

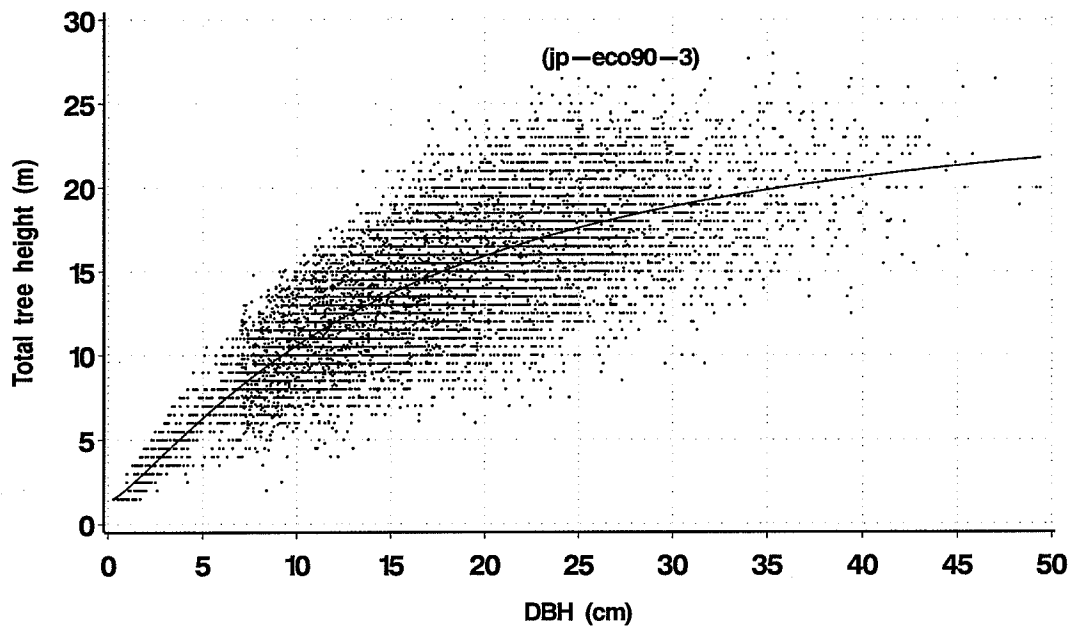


Figure 2.14. Plots of total tree height against DBH for jack pine in ecoregion 90 based on model [2.3].

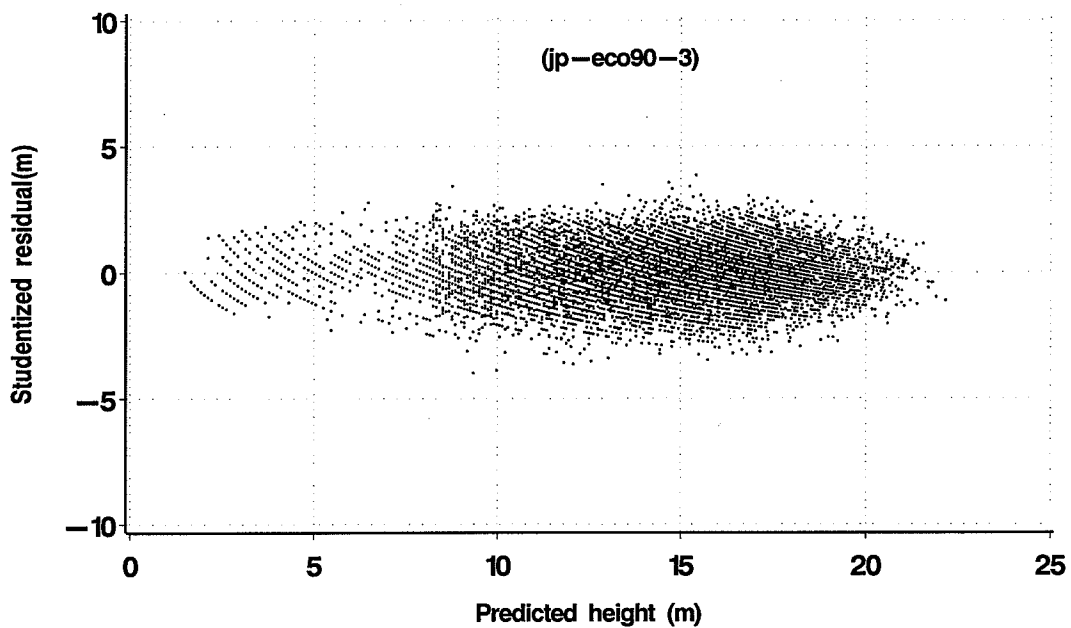


Figure 2.15. Plots of studentized residual against predicted height for jack pine in ecoregion 90 based on model [2.3].

2.5.1.4 Ecoregion 91

The fitted nonlinear least squares estimates of the parameters, root mean squared error (RMSE), the coefficient of determination (R^2) are shown in Table 2.8. The asymptotic t-statistics, average bias (\bar{e}), mean absolute deviation (MAD) are shown in Table 2.9.

Judging from the results shown in Table 2.8 and Table 2.9, model [2.2], [2.1], [2.3] performed quite well for black spruce. The R^2 values were higher and the RMSE were smaller, model [2.2] had the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were significant at $\alpha=0.05$ level. The bias_per of model [2.1]-[2.4] and [2.7] were less than 0.05% and model [2.1] showed the lowest -0.009 %. Model [2.5]-[2.6] showed higher bias_per, greater than 0.3%. The paired t tests of model [2.1]-[2.4] and [2.6]-[2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. Model [2.5] showed the highest bias_per 0.94 %. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, but the mean absolute deviation (MAD) and Se values are relatively higher, 1.5294 and 2.0741 respectively. Further more, model [2.8] graphically displayed like a line, which does not make biological sense. Overall, model [2.2] and [2.1] performed the best. The graphically presentation of model [2.2] is given in Figure 2.16, along with height and diameter data. The studentized residual plot of model [2.2] is given in Figure 2.17.

For jack pine, model [2.2], [2.1], [2.3] performed quite well. The R^2 values were higher and the RMSE were smaller, model [2.2] showed the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were significant at

$\alpha=0.05$ level. The bias_per of model [2.1]-[2.3] were less than -0.2% and model [2.2] showed the lowest -0.0443 %. Model [2.5] showed higher bias_per, greater than 0.77 %. The paired t tests of model [2.1]-[2.4] and [2.7]-[2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, but the mean absolute deviation (MAD) and Se values were the highest, 2.3452 and 3.0463 respectively. Further more, model [2.8] graphically displayed like a line, which does not make biological sense. Overall, model [2.2] and [2.1] performed the best. The graphically presentation of model [2.2] is given in Figure 2.18, along with height and diameter data. The studentized residual plot of model [2.2] is given in Figure 2.19, showing an almost homogenous band of the studentized residuals with zero mean, and it indicates the assumptions of the regression analysis are almost met.

Table 2.8. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 91.

Spp	Model	a	b	c	n	RMSE	R2
BS	[2.1]	16.53	0.1132	1.4983	5811	0.5118	0.9484
BS	[2.2]	15.5862	0.0412	1.339	5811	0.5115	0.9485
BS	[2.3]	20.0293	0.0292	1.4662	5811	0.5128	0.9482
BS	[2.4]	23.6169	-11.4032	2.1134	5811	0.5141	0.948
BS	[2.5]	18.5146	-6.5797		5811	0.5548	0.9394
BS	[2.6]	3.0301	-8.6662		5811	0.5237	0.946
BS	[2.7]	1.2666	0.2029		5811	0.5143	0.9479
BS	[2.8]	0.9954	0.7357		5811	0.5982	0.9296
JP	[2.1]	16.0895	0.1212	1.3085	5373	0.6135	0.9379
JP	[2.2]	15.6786	0.0638	1.2128	5373	0.6123	0.9381
JP	[2.3]	18.8304	0.0478	1.387	5373	0.6172	0.9371
JP	[2.4]	21.3946	-9.0144	1.9367	5373	0.6186	0.9368
JP	[2.5]	18.2536	-5.506		5373	0.6588	0.9284
JP	[2.6]	2.9814	-7.1509		5373	0.6271	0.9351
JP	[2.7]	0.9691	0.2175		5373	0.6225	0.936
JP	[2.8]	3.5971	0.5334		5373	0.7812	0.8993

Note: RMSE - root mean square error of prediction; R^2 - prediction coefficient of determination.

Table 2.9. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 91.

Spp	Model	\bar{e}	MAD	Bias_per	Se	t	n
BS	[2.1]	-0.0022	1.2822	-0.0211	1.696	-0.0993	5811
BS	[2.2]	-0.001	1.28	-0.0093	1.693	-0.044	5811
BS	[2.3]	-0.0037	1.2861	-0.0349	1.701	-0.1637	5811
BS	[2.4]	-0.0044	1.2911	-0.0418	1.706	-0.1957	5811
BS	[2.5]	0.0984	1.3745	0.94	1.76	4.2616	5811
BS	[2.6]	0.0347	1.3119	0.3318	1.716	1.5431	5811
BS	[2.7]	-0.002	1.2929	-0.019	1.708	-0.0887	5811
BS	[2.8]	0	1.5294	0	2.074	0	5811
JP	[2.1]	-0.0107	1.7625	-0.0843	2.299	-0.3397	5373
JP	[2.2]	-0.0056	1.7582	-0.0443	2.295	-0.1788	5373
JP	[2.3]	-0.0205	1.7748	-0.1625	2.313	-0.6509	5373
JP	[2.4]	-0.0346	1.7898	-0.274	2.329	-1.0899	5373
JP	[2.5]	0.0977	1.8084	0.7729	2.334	3.068	5373
JP	[2.6]	0.0183	1.7812	0.1452	2.317	0.5804	5373
JP	[2.7]	-0.0198	1.7881	-0.1565	2.326	-0.6231	5373
JP	[2.8]	0	2.3452	0	3.046	0	5373

Note: \bar{e} - average bias; MAD - mean absolute deviation; bias_per – average bias percent; Se - standard deviation of the prediction errors; t-the paired t test statistics.

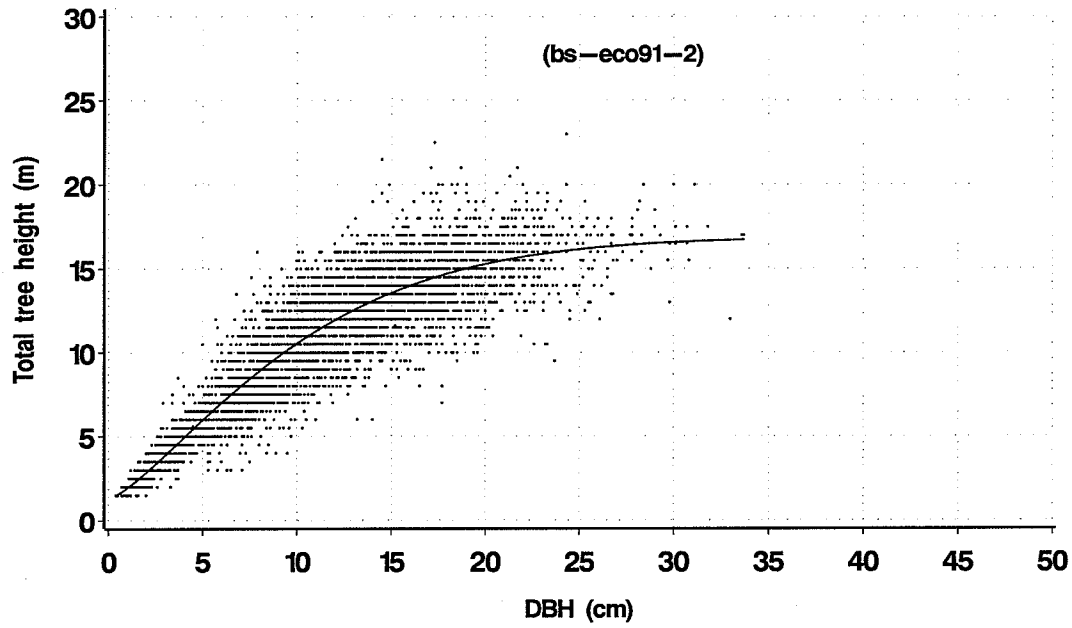


Figure 2.16. Plots of total tree height against DBH for black spruce in ecoregion 91 based on model [2.2].

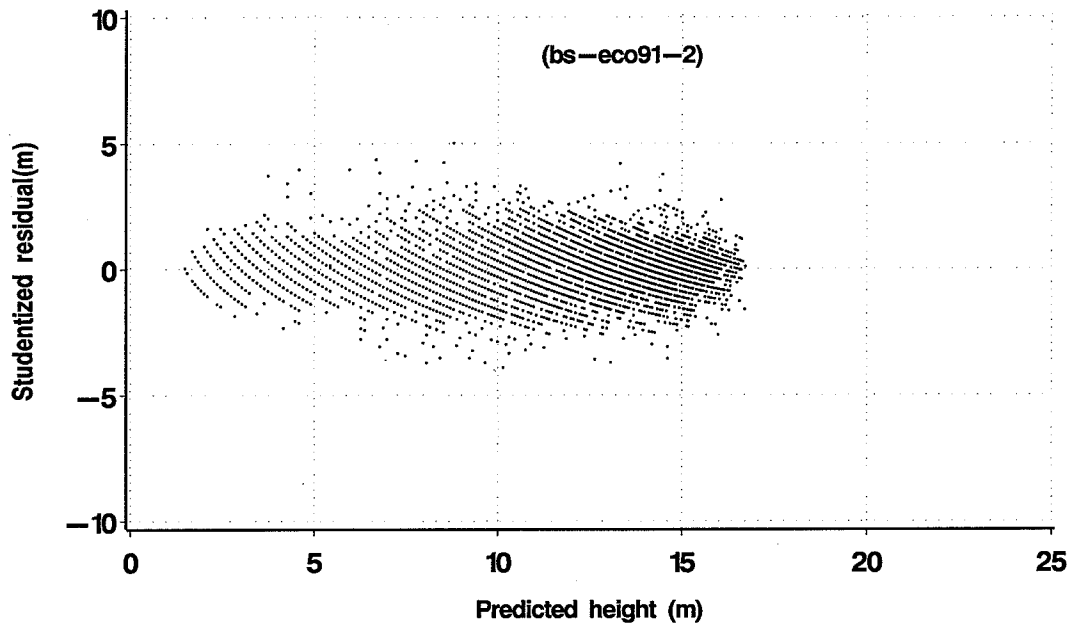


Figure 2.17. Plots of studentized residual against predicted height for black spruce in ecoregion 91 based on model [2.2].

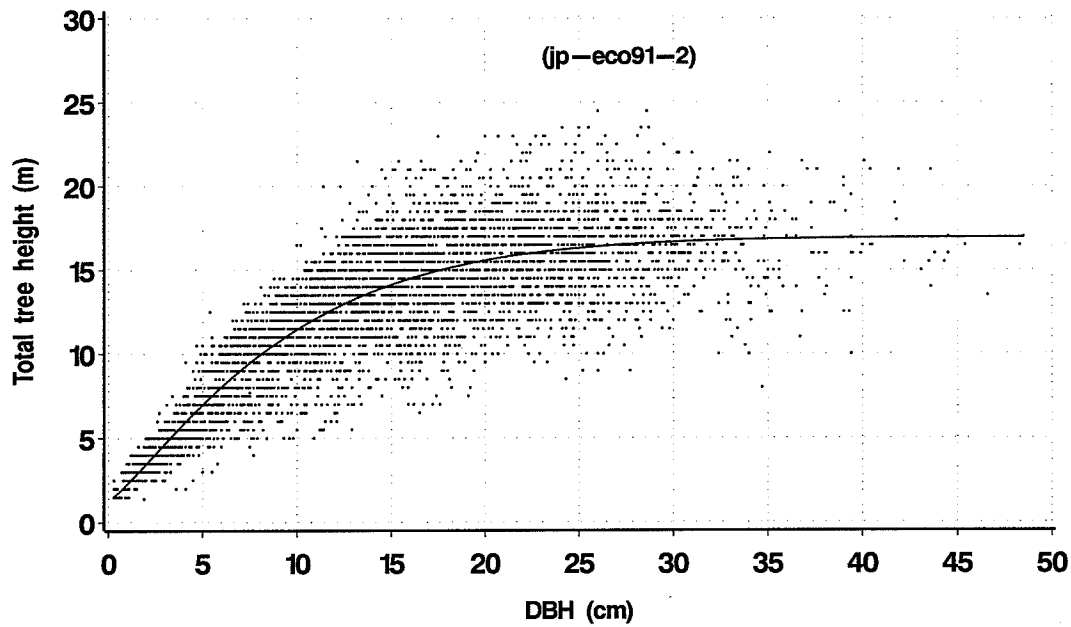


Figure 2.18. Plots of total tree height against DBH for jack pine in ecoregion 91 based on model [2.2].

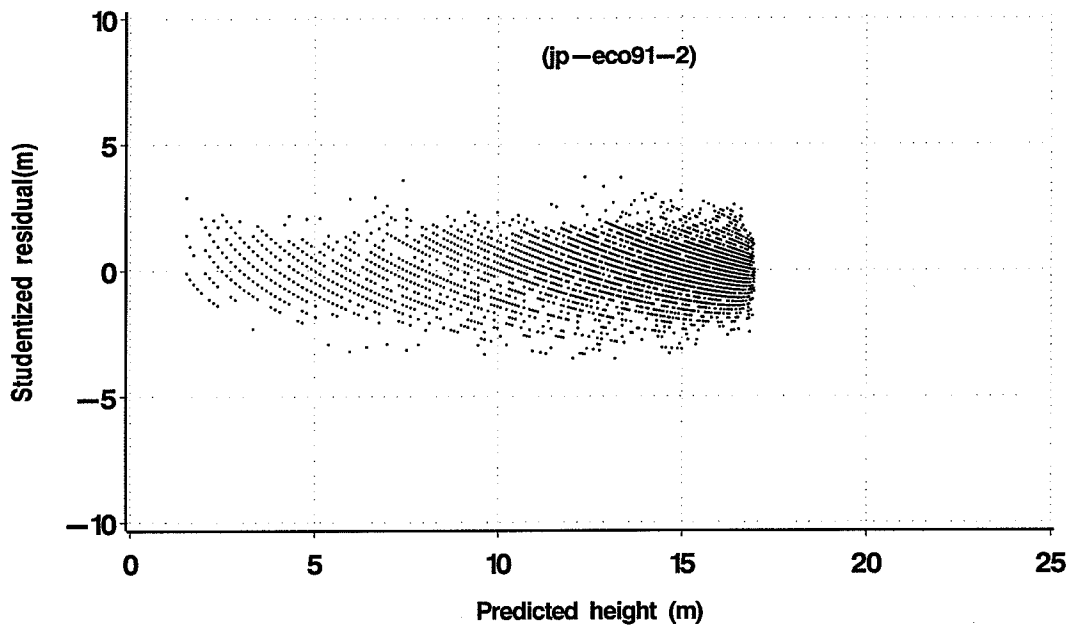


Figure 2.19. Plots of studentized residual against predicted height for jack pine in ecoregion 91 based on model [2.2].

2.5.1.5 Ecoregion 148

The fitted nonlinear least squares estimates of the parameters, root mean squared error (RMSE), the coefficient of determination (R^2) are shown in Table 2.10. The asymptotic t-statistics, average bias (\bar{e}), mean absolute deviation (MAD) are shown in Table 2.11.

Judging from the results shown in Table 2.10 and Table 2.11, model [2.1], [2.2], [2.3] performed quite well for black spruce. The R^2 values were higher and the RMSE were smaller, model [2.2] showed the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were significant at $\alpha=0.05$ level. The bias_per of model [2.1]-[2.4] and [2.7] were less than 0.1 % and model [2.1] showed the lowest - 0.000 %. Model [2.5]-[2.6] had higher bias_per, greater than 0.2 %. The paired t tests of model [2.1]-[2.4] and [2.7]-[2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. Model [2.5] had the highest bias_per 0.72 %. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, but the mean absolute deviation (MAD) and Se values were relatively higher, 1.6396 and 2.2528 respectively. Further more, model [2.8] graphically displayed like a line, which does not make biological sense. Overall, model [2.1] and [2.2] performed the best. The graphically presentation of model [2.1] is given in Figure 2.20, along with height and diameter data. The studentized residual plot of model [2.1] is given in Figure 2.21.

For jack pine, model [2.2], [2.1], [2.3] and [2.4] performed quite well. The R^2 values were higher and the RMSE were smaller, model [2.2] had the highest R^2 and lowest RMSE. All the t-statistics for the parameters of these eight models were

significant at $\alpha=0.05$ level. The bias_per of model [2.1]-[2.3] were less than -0.1 % and model [2.4] showed the lowest -0.0181 %. Model [2.5] showed higher bias_per, greater than 1.177 %. The paired t tests of model [2.1]-[2.4] and [2.7]-[2.8] showed that the mean prediction bias were not significantly different from zero at $\alpha=0.05$ level. For model [2.8], although the t-statistics for the parameters were significant at $\alpha=0.05$ level and Bias_per was about 0.000%, but the mean absolute deviation (MAD) and Se values were the highest, 1.9675 and 2.5948 respectively. Further more, model [2.8] graphically displayed like a line, which does not make biological sense. Overall, model [2.2] and [2.1] performed the best. The graphically presentation of model [2.2] is given in Figure 2.22, along with height and diameter data. The studentized residual plot of model [2.2] is given in Figure 2.23, showing an almost homogenous band of the studentized residuals with zero mean, and it indicates the assumptions of the regression analysis are almost met.

Table 2.10. Estimated parameters and RMSE of model [2.1] to [2.8] for black spruce and jack pine in ecoregion 148.

Spp	Model	a	b	c	n	RMSE	R2
BS	[2.1]	18.3838	0.1081	1.681	19346	0.561	0.9488
BS	[2.2]	17.0524	0.0276	1.445	19346	0.5606	0.9489
BS	[2.3]	21.7102	0.0185	1.597	19346	0.5612	0.9487
BS	[2.4]	27.6769	-13.5953	2.281	19346	0.5621	0.9486
BS	[2.5]	21.811	-8.3192		19346	0.5891	0.9435
BS	[2.6]	3.1825	-10.428		19346	0.5703	0.9471
BS	[2.7]	1.4816	0.1833		19346	0.5639	0.9482
BS	[2.8]	0.371	0.8009		19346	0.6353	0.9343
JP	[2.1]	21.4255	0.0552	1.006	9557	0.5748	0.919
JP	[2.2]	21.215	0.0542	1.011	9557	0.5748	0.919
JP	[2.3]	30.3918	0.0375	1.054	9557	0.576	0.9187
JP	[2.4]	27.5055	-15.9197	4.307	9557	0.5765	0.9186
JP	[2.5]	20.0236	-7.179		9557	0.6511	0.8961
JP	[2.6]	3.0524	-8.6307		9557	0.6156	0.9071
JP	[2.7]	1.2308	0.2038		9557	0.5935	0.9137
JP	[2.8]	2.9474	0.5663		9557	0.6291	0.903

Note: RMSE - root mean square error of prediction; R^2 - prediction coefficient of determination.

Table 2.11. Prediction statistics of model [2.1] to [2.8] based on the fitting data for black spruce and jack pine in ecoregion 148.

Spp	Model	\bar{e}	MAD	Bias_per	Se	t	n
BS	[2.1]	0	1.4266	0	1.8779	-0.0001	19346
BS	[2.2]	0.0014	1.4252	0.0132	1.8785	0.1045	19346
BS	[2.3]	-0.0009	1.4276	-0.0087	1.8787	-0.0687	19346
BS	[2.4]	-0.0061	1.4315	-0.0573	1.885	-0.4537	19346
BS	[2.5]	0.0766	1.4889	0.7145	1.9126	5.5713	19346
BS	[2.6]	0.0307	1.4499	0.286	1.8877	2.2591	19346
BS	[2.7]	-0.0076	1.4364	-0.0706	1.8915	-0.5566	19346
BS	[2.8]	0	1.6396	0	2.2528	0	19346
JP	[2.1]	-0.0031	1.8066	-0.023	2.3336	-0.1287	9557
JP	[2.2]	-0.0029	1.8063	-0.0216	2.333	-0.1205	9557
JP	[2.3]	-0.0046	1.8118	-0.0343	2.3411	-0.191	9557
JP	[2.4]	-0.0024	1.8095	-0.0181	2.3341	-0.101	9557
JP	[2.5]	0.1571	1.9645	1.1773	2.4392	6.2971	9557
JP	[2.6]	0.1039	1.911	0.7785	2.4	4.2317	9557
JP	[2.7]	0.0395	1.8489	0.2956	2.351	1.6405	9557
JP	[2.8]	0	1.9675	0	2.5948	0	9557

Note: \bar{e} - average bias; MAD - mean absolute deviation; bias_per – average bias percent; Se - standard deviation of the prediction errors; t-the paired t test statistics.

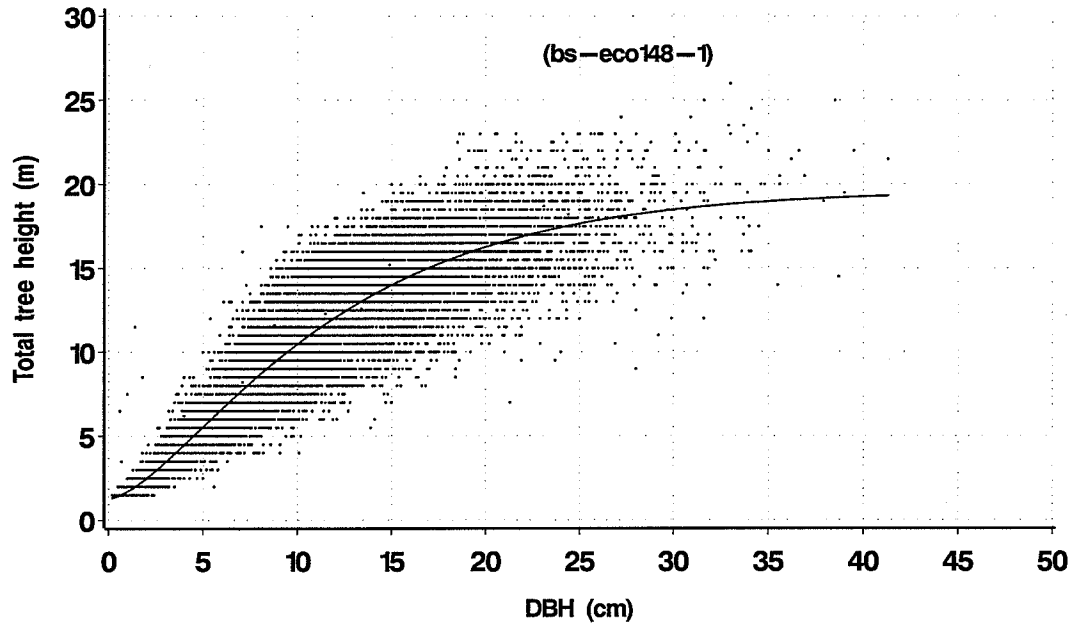


Figure 2.20. Plots of total tree height against DBH for black spruce in ecoregion 148 based on model [2.1].

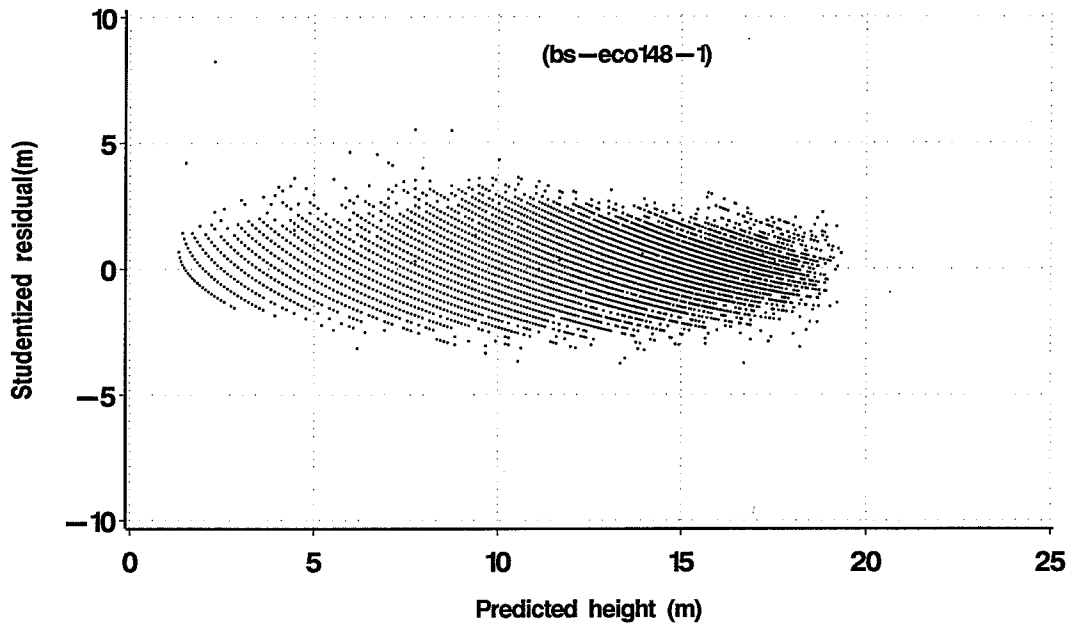


Figure 2.21. Plots of studentized residual against predicted height for black spruce in ecoregion 148 based on model [2.1].

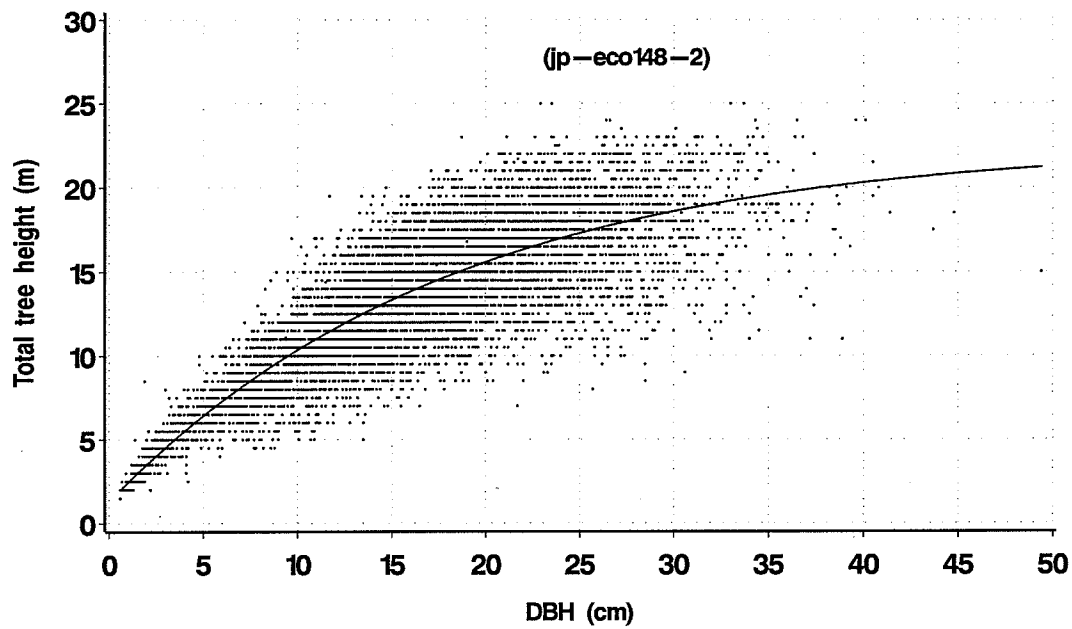


Figure 2.22. Plots of total tree height against DBH for jack pine in ecoregion 148 based on model [2.2].

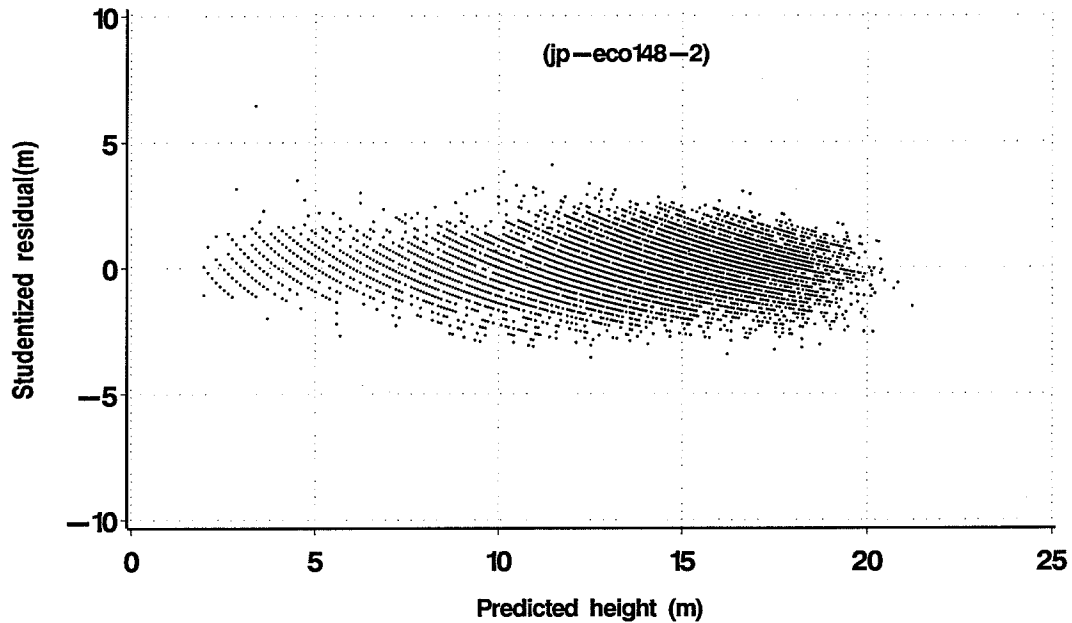


Figure 2.23. Plots of studentized residual against predicted height for jack pine in ecoregion 148 based on model [2.2].

2.5.2 Model validation

2.5.2.1 Model validation for black spruce

Based on the validation data, the measured height values from the testing data set were compared to values predicted by model [2.1] to [2.8] using the estimated parameters in Table 2.2 to Table 2.11. The bias of the prediction was obtained by subtracting the predicted height from the measured height. The prediction statistics calculated for model [1] to [8] and five ecoregions for black spruce and jack pine are given in Table 2.12.

For black spruce, in ecoregion 88, model [2.1] to [2.4] performed quite well and model [2.2] performed the best. For model [2.2], the mean bias (\bar{e}) was -0.049 ; the standard deviation of the prediction bias (S_e) was 1.438 ; the root mean square error of prediction ($RMSE_p$) was 0.479 ; the percent bias (Bias_per) was -0.53% ; the prediction coefficient of determination (R_p^2) is 0.9779 ; the paired t-test calculated $t = -1.804$ with 2747 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 89, model [2.1] to [2.4] performed quite well and model [2.1] performed the best. For model [2.1], the mean bias (\bar{e}) was -0.050 ; the standard deviation of the prediction bias (S_e) was 1.086 ; the root mean square error of prediction ($RMSE_p$) was 0.391 ; the percent bias (Bias_per) was -0.75% ; the prediction coefficient of determination (R_p^2) is 0.9789 ; the paired t-test calculated $t = -1.860$ with 1629 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 90, model [2.1] to [2.4] and [2.7] performed quite well and model [2.4] performed the best. For model [2.4], the mean bias (\bar{e}) was 0.007 ; the standard

deviation of the prediction bias (S_e) was 1.965; the root mean square error of prediction ($RMSE_p$) was 0.553; the percent bias ($Bias_per$) was 0.60%; the prediction coefficient of determination (R_p^2) is 0.8814; the paired t-test calculated $t = 0.236$ with 4218 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 91, model [2.1] to [2.4] and [2.6] to [2.7] performed quite well and model [2.2] performed the best. For model [2.2], the mean bias (\bar{e}) was -0.008; the standard deviation of the prediction bias (S_e) was 1.760; the root mean square error of prediction ($RMSE_p$) was 0.527; the percent bias ($Bias_per$) was -0.07%; the prediction coefficient of determination (R_p^2) is 0.9455; the paired t-test calculated $t = -0.166$ with 1453 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 148, model [2.1] to [2.3] performed quite well and model [2.2] performed the best. For model [2.2], the mean bias (\bar{e}) was -0.063; the standard deviation of the prediction bias (S_e) was 1.840; the root mean square error of prediction ($RMSE_p$) was 0.546; the percent bias ($Bias_per$) was -0.60%; the prediction coefficient of determination (R_p^2) is 0.9515; the paired t-test calculated $t = -2.379$ with 4837 degrees of freedom, was significant at $\alpha=0.05$ level, it means the mean prediction bias was significantly different from zero.

2.5.2.2 Model validation for jack pine

In ecoregion 88, model [2.1] to [2.4] performed quite well and model [2.3] performed the best. For model [2.3], the mean bias (\bar{e}) was -0.044; the standard deviation

of the prediction bias (S_e) was 1.581; the root mean square error of prediction ($RMSE_p$) was 0.52; the percent bias ($Bias_per$) was -0.50% ; the prediction coefficient of determination (R_p^2) is 0.976; the paired t-test calculated $t = -1.206$ with 1886 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 89, model [2.1] to [2.4] and model [2.7] performed quite well and model [2.1] performed the best. For model [2.1], the mean bias (\bar{e}) was -0.031 ; the standard deviation of the prediction bias (S_e) was 1.470; the root mean square error of prediction ($RMSE_p$) was 0.474; the percent bias ($Bias_per$) was -0.43% ; the prediction coefficient of determination (R_p^2) is 0.9869; the paired t-test calculated $t = -0.598$ with 664 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 90, model [2.1] to [2.4] and [2.7] performed quite well and model [2.1] performed the best. For model [2.1], the mean bias (\bar{e}) was -0.017 ; the standard deviation of the prediction bias (S_e) was 2.621; the root mean square error of prediction ($RMSE_p$) was 0.641; the percent bias ($Bias_per$) was -0.12% ; the prediction coefficient of determination (R_p^2) is 0.8438; the paired t-test calculated $t = -0.364$ with 3334 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 91, model [2.1] to [2.4] performed quite well and model [2.2] performed the best. For model [2.2], the mean bias (\bar{e}) was 0.053; the standard deviation of the prediction bias (S_e) was 2.261; the root mean square error of prediction ($RMSE_p$) was 0.606; the percent bias ($Bias_per$) was 0.42% ; the prediction coefficient of

determination (R_p^2) is 0.8438; the paired t-test calculated $t = -0.412$ with 3334 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero.

In ecoregion 148, model [2.1] to [2.4] performed quite well and model [2.2] performed the best. For model [2.2], the mean bias ($\bar{\epsilon}$) was -0.042; the standard deviation of the prediction bias (S_ϵ) was 2.268; the root mean square error of prediction ($RMSE_p$) was 0.561; the percent bias (Bias_per) was -0.32%; the prediction coefficient of determination (R_p^2) is 0.9201; the paired t-test calculated $t = -0.903$ with 2390 degrees of freedom, was significant at $\alpha=0.05$ level, it means the mean prediction bias was not significantly different from zero.

Table 2.12. Prediction statistics based on the validation data set, calculated for model [2.1] to [2.8] and the five ecoregions for black spruce and jack pine.

Eco-region	Spp	Model	n	R_p^2	RMSE _p	$\bar{\epsilon}$	MAD	Bias per	Se	t
88	BS	[2.1]	2747	0.9778	0.48	-0.051	1.08	-0.55	1.442	-1.864
88	BS	[2.2]	2747	0.9779	0.479	-0.049	1.08	-0.53	1.438	-1.804
88	BS	[2.3]	2747	0.9777	0.481	-0.053	1.09	-0.57	1.45	-1.927
88	BS	[2.4]	2747	0.9775	0.484	-0.047	1.09	-0.51	1.447	-1.72
88	BS	[2.5]	2747	0.9641	0.611	0.189	1.29	2.03	1.646	6.013
88	BS	[2.6]	2747	0.973	0.53	0.063	1.16	0.68	1.529	2.157
88	BS	[2.7]	2747	0.9761	0.498	-0.011	1.1	-0.11	1.459	-0.383
88	BS	[2.8]	2747	0.9698	0.561	-0.09	1.32	-0.97	1.927	-2.446
88	JP	[2.1]	1886	0.9758	0.521	-0.045	1.14	-0.51	1.582	-1.235
88	JP	[2.2]	1886	0.9757	0.522	-0.046	1.14	-0.52	1.583	-1.258
88	JP	[2.3]	1886	0.976	0.52	-0.044	1.13	-0.5	1.581	-1.206
88	JP	[2.4]	1886	0.9754	0.526	-0.022	1.14	-0.25	1.599	-0.596
88	JP	[2.5]	1886	0.9645	0.632	0.245	1.38	2.78	1.912	5.563
88	JP	[2.6]	1886	0.9726	0.554	0.082	1.21	0.93	1.716	2.081
88	JP	[2.7]	1886	0.975	0.53	0.002	1.15	0.02	1.614	0.051
88	JP	[2.8]	1886	0.964	0.636	-0.035	1.46	-0.4	2.177	-0.698
89	BS	[2.1]	1629	0.9789	0.391	-0.05	0.72	-0.75	1.086	-1.86
89	BS	[2.2]	1629	0.9787	0.392	-0.049	0.72	-0.73	1.091	-1.812
89	BS	[2.3]	1629	0.9789	0.39	-0.052	0.72	-0.77	1.084	-1.919
89	BS	[2.4]	1629	0.9789	0.391	-0.054	0.72	-0.8	1.088	-1.987
89	BS	[2.5]	1629	0.9605	0.535	0.113	1	1.7	1.417	3.231
89	BS	[2.6]	1629	0.9738	0.435	-0.009	0.81	-0.13	1.218	-0.285
89	BS	[2.7]	1629	0.978	0.399	-0.039	0.74	-0.58	1.102	-1.419
89	BS	[2.8]	1629	0.9726	0.446	0.003	0.83	0.05	1.378	0.098
89	JP	[2.1]	664	0.9869	0.474	-0.031	1.03	-0.4	1.47	-0.548
89	JP	[2.2]	664	0.9867	0.478	-0.034	1.04	-0.43	1.477	-0.594
89	JP	[2.3]	664	0.987	0.473	-0.029	1.02	-0.37	1.467	-0.516
89	JP	[2.4]	664	0.9871	0.472	-0.022	1.02	-0.28	1.476	-0.387
89	JP	[2.5]	664	0.9813	0.567	0.219	1.18	2.79	1.727	3.268
89	JP	[2.6]	664	0.9857	0.496	0.072	1.06	0.91	1.576	1.172
89	JP	[2.7]	664	0.987	0.472	-0.01	1.02	-0.13	1.467	-0.183
89	JP	[2.8]	664	0.9809	0.573	-0.078	1.28	-0.99	1.954	-1.029

(Table 2.12 continued next page)

Table 2.12 (continued).

90	BS	[2.1]	4218	0.881	0.554	0.009	1.52	0.07	1.967	0.293
90	BS	[2.2]	4218	0.8805	0.556	0.009	1.52	0.07	1.968	0.301
90	BS	[2.3]	4218	0.8812	0.554	0.008	1.52	0.07	1.965	0.269
90	BS	[2.4]	4218	0.8814	0.553	0.007	1.52	0.06	1.965	0.236
90	BS	[2.5]	4218	0.8786	0.56	0.048	1.54	0.39	1.979	1.564
90	BS	[2.6]	4218	0.8812	0.554	0.024	1.52	0.19	1.966	0.785
90	BS	[2.7]	4218	0.881	0.554	0.004	1.52	0.03	1.968	0.118
90	BS	[2.8]	4218	0.8473	0.628	0.017	1.72	0.14	2.241	0.502
90	JP	[2.1]	3334	0.8438	0.641	-0.017	2.06	-0.12	2.622	-0.364
90	JP	[2.2]	3334	0.8437	0.641	-0.016	2.06	-0.12	2.622	-0.359
90	JP	[2.3]	3334	0.8438	0.641	-0.019	2.07	-0.13	2.624	-0.412
90	JP	[2.4]	3334	0.8434	0.642	-0.014	2.07	-0.1	2.626	-0.309
90	JP	[2.5]	3334	0.8341	0.661	0.045	2.11	0.32	2.668	0.978
90	JP	[2.6]	3334	0.8394	0.65	0.021	2.09	0.15	2.649	0.459
90	JP	[2.7]	3334	0.8432	0.642	-0.009	2.07	-0.06	2.629	-0.192
90	JP	[2.8]	3334	0.8145	0.699	-0.002	2.23	-0.02	2.849	-0.045
91	BS	[2.1]	1453	0.9454	0.528	-0.01	1.34	-0.1	1.762	-0.219
91	BS	[2.2]	1453	0.9455	0.527	-0.008	1.34	-0.07	1.76	-0.166
91	BS	[2.3]	1453	0.9452	0.529	-0.013	1.35	-0.12	1.766	-0.286
91	BS	[2.4]	1453	0.9448	0.53	-0.015	1.35	-0.14	1.771	-0.323
91	BS	[2.5]	1453	0.9359	0.571	0.103	1.44	0.97	1.823	2.156
91	BS	[2.6]	1453	0.9427	0.541	0.033	1.38	0.31	1.78	0.702
91	BS	[2.7]	1453	0.9447	0.531	-0.012	1.35	-0.12	1.773	-0.267
91	BS	[2.8]	1453	0.9258	0.615	-0.051	1.6	-0.48	2.161	-0.893
91	JP	[2.1]	1344	0.9434	0.607	0.05	1.76	0.4	2.263	0.815
91	JP	[2.2]	1344	0.9435	0.606	0.053	1.76	0.42	2.261	0.862
91	JP	[2.3]	1344	0.9428	0.61	0.047	1.77	0.37	2.276	0.753
91	JP	[2.4]	1344	0.9425	0.611	0.037	1.78	0.29	2.287	0.592
91	JP	[2.5]	1344	0.9344	0.653	0.154	1.81	1.22	2.312	2.435
91	JP	[2.6]	1344	0.9407	0.621	0.081	1.78	0.65	2.286	1.302
91	JP	[2.7]	1344	0.9418	0.615	0.049	1.78	0.39	2.287	0.79
91	JP	[2.8]	1344	0.911	0.761	0.136	2.28	1.08	2.889	1.725

(Table 2.12 continued next page)

Table 2.12 (continued).

148	BS	[2.1]	4837	0.9514	0.546	-0.063	1.41	-0.6	1.842	-2.386
148	BS	[2.2]	4837	0.9515	0.546	-0.063	1.41	-0.6	1.84	-2.379
148	BS	[2.3]	4837	0.9513	0.547	-0.063	1.41	-0.6	1.845	-2.389
148	BS	[2.4]	4837	0.9511	0.548	-0.067	1.42	-0.63	1.854	-2.514
148	BS	[2.5]	4837	0.9459	0.576	0.007	1.46	0.07	1.872	0.257
148	BS	[2.6]	4837	0.9496	0.556	-0.035	1.43	-0.33	1.851	-1.321
148	BS	[2.7]	4837	0.9508	0.549	-0.067	1.42	-0.64	1.861	-2.51
148	BS	[2.8]	4837	0.937	0.622	-0.035	1.62	-0.33	2.224	-1.096
148	JP	[2.1]	2390	0.9201	0.561	-0.042	1.73	-0.32	2.269	-0.911
148	JP	[2.2]	2390	0.9201	0.561	-0.042	1.73	-0.32	2.268	-0.903
148	JP	[2.3]	2390	0.9197	0.562	-0.045	1.74	-0.35	2.276	-0.977
148	JP	[2.4]	2390	0.9196	0.563	-0.04	1.74	-0.3	2.27	-0.851
148	JP	[2.5]	2390	0.895	0.643	0.108	1.92	0.82	2.393	2.21
148	JP	[2.6]	2390	0.9067	0.606	0.054	1.86	0.41	2.349	1.134
148	JP	[2.7]	2390	0.9139	0.582	-0.003	1.79	-0.02	2.293	-0.059
148	JP	[2.8]	2390	0.9051	0.611	-0.039	1.89	-0.29	2.525	-0.746

Note: SPP – species; n – total number of observations (height-dbh pairs); \bar{e} - average bias; MAD - absolute mean deviation; bias_per - bias %; S_e - standard deviation of the prediction errors; t - paired t test; RMSE_p - root mean square error of prediction; R_p^2 - prediction coefficient of determination .

2.5.3 Testing height-diameter relationships among and between ecoregions

2.5.3.1 Testing height-diameter relationships among the five ecoregions

Results of F-test are given in Table 2.13. Based on the testing results, height-diameter relationships differed among the five ecoregions for black spruce and jack pine.

Table 2.13. F-test of the differences for the height-diameter among five forest ecoregions.

Spp	Ecoregion pair	Full Model		Reduced model			F value	p-value
		SSE(F)	df(F)	SSE(R)	df(R)	N		
BS	Overall	16333.17	59511	16739.6	59523	59526	493.59*	<0.001
JP	Overall	13171.08	38444	13602	38456	38459	419.24*	<0.001

Note: F-value was calculated as $\{[SSE(R)-SSE(F)]/[df(R)-df(F)]\} / [SSE(F)/df(F)]$, where SSE(R) and SSE(F) are sum-of-squares of residual for reduced and full models, $df(R)$ and $df(F)$ are degree-of-freedoms of residual for reduced and full models. F-value with asterisk (*) indicate a significance at $\alpha=0.05$.

2.5.3.2 Testing height-diameter relationships between each paired ecoregions

Because the differences among five forest ecoregions may be caused by only two or all of the ecoregions involved, it is often desirable to apply the indicator variable approach to each possible pair of ecoregions so that the source of the differences can be identify and data from similar ecoregions can be combined.

Results of F-tests are given in table 2.14, which revealed that there were significant differences in the height-diameter models between any two ecoregions for black spruce and jack pine. These findings indicated that the height-diameter relationships were different and each ecoregion appeared to need a unique height-diameter model.

Figure 2.24 shows the height-diameter curve overlay by the “best” fitted models in each ecoregion for black spruce and jack pine. For black spruce, we found that there

were almost no significant differences when the diameter is less than 15 cm; when the diameter greater than 15 cm, then the curves start to be divergent, the bigger the diameter, the bigger the difference. For jack pine, we found that there were almost no significant differences when the diameter is less than 22 cm; when the diameter greater than 22 cm, then the curves start to be divergent, the bigger the diameter, the bigger the difference.

We also found, for black spruce, that the overlay curves of ecoregion 88 and ecoregion 148 are quite similar, just tiny difference, but based on the F test, they were significant difference; similarly, for jack pine, the overlay curves of ecoregion 88, ecoregion 90 and ecoregion 148 are quite similar, but based on the F test, they were significant difference. It may be that the F test is too sensitive or too “strict”.

Table 2.14. F-test of the differences for the height-diameter between each paired ecoregions in Manitoba.

Spp	Ecoregion pair	Full model		Reduced model				p-value
		SSE(F)	df(F)	SSE (R)	df(R)	N	F value	
BS	Eco91-eco148	7598.22	25151	7705.06	25154	25157	117.88*	<0.001
JP	Eco91-eco148	5170.01	14924	5369.97	14927	14930	192.41*	<0.001
BS	Eco88-eco89	3718.41	17493	3782.84	17496	17499	101.03*	<0.001
JP	Eco88-eco89	2586.11	10189	2633.51	10192	10195	62.25*	<0.001
BS	Eco88-eco90	7660.41	27851	7677.1	27854	27857	20.23*	<0.001
JP	Eco88-eco90	7426.22	20869	7447.71	20872	20875	20.13*	<0.001
BS	Eco88-eco91	4163.33	16792	4229.68	16795	16798	89.2*	<0.001
JP	Eco88-eco91	4024.73	12908	4130.87	12911	12914	113.47*	<0.001
BS	Eco88-eco148	8722.63	30327	8987.71	30330	30333	307.21*	<0.001
JP	Eco88-eco148	5167.8	17092	5243.51	17095	17098	83.47*	<0.001
BS	Eco89-eco90	6091.08	23376	6130.88	23379	23382	50.92*	<0.001
JP	Eco89-eco90	5989.81	15982	6035.94	15985	15988	41.03*	<0.001
BS	Eco89-eco91	2594	12317	2705.62	12320	12323	176.67*	<0.001
JP	Eco89-eco91	2588.32	8021	2701.5	8024	8027	116.92*	<0.001
BS	Eco89-eco148	7153.3	25852	7302.46	25855	25858	179.68*	<0.001
JP	Eco89-eco148	3731.4	12205	3849.75	12208	12211	129.04*	<0.001
BS	Eco90-eco91	6536	22675	6599.84	22678	22681	73.82*	<0.001
JP	Eco90-eco91	7428.43	18701	7634.18	18704	18707	172.65*	<0.001
BS	Eco90-eco148	11095.3	36210	11175	36213	36216	86.73*	<0.001
JP	Eco90-eco148	8571.51	22885	8610.23	22888	22891	34.46*	<0.001

Note: F-value was calculated as $\{[SSE(R)-SSE(F)]/[df(R)-df(F)]\} / [SSE(F)/df(F)]$, where SSE(R) and SSE(F) are sum-of-squares of residual for reduced and full models, $df(R)$ and $df(F)$ are degree-of-freedoms of residual for reduced and full models. F-value with asterisk (*) indicate a significance at $\alpha = 0.05$.

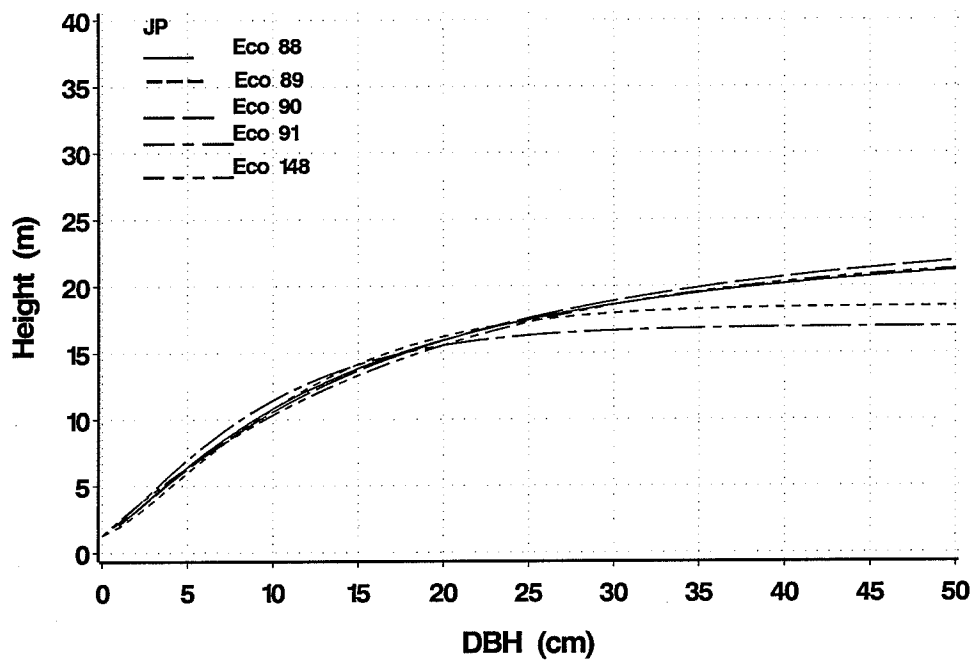
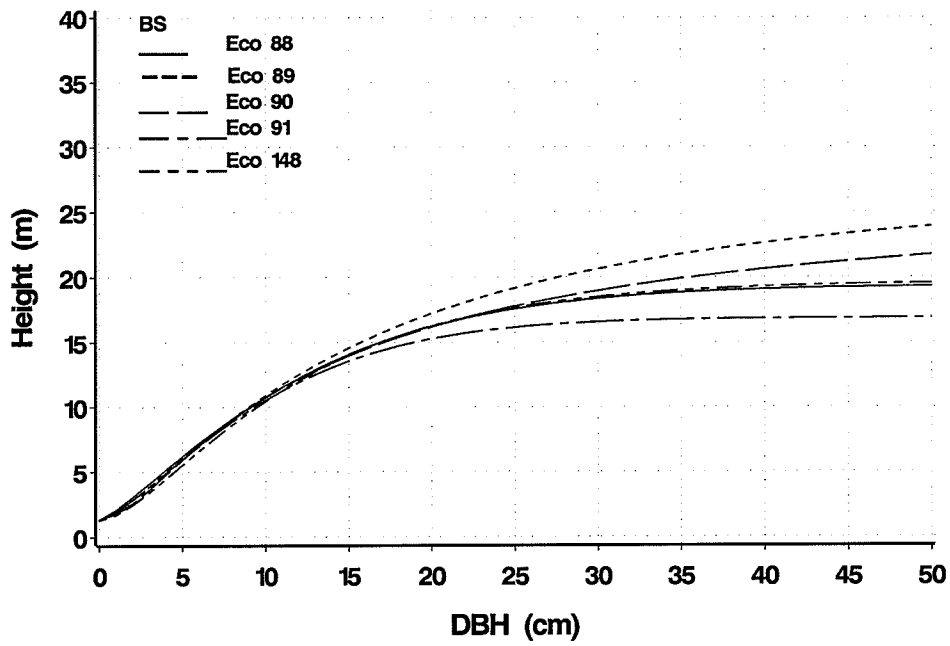


Figure 2.24. Height-diameter curves overlay by the best fitted models in five ecoregions. Above is for black spruce and below is for jack pine.

2.5.4 Prediction errors of applying the combined model to each ecoregion

Based on the tests above, a unique height-diameter model should be applied to each ecoregion. Applied the same height-diameter model to every ecoregion would result in prediction biases. To further justify the need for ecoregion-specific model, a combined height-diameter model was fitted to the data from all five ecoregions. This combined model was then applied to predict tree height from diameter for each ecoregion. Table 2.15 gives the average of measured tree height (HGT_m), average of predicted tree height (Pred_m), average bias ($\bar{\epsilon}$), mean absolute deviation (MAD), bias % (bias_per), standard deviation of the prediction errors (S_e) and the paired t-test for testing the null hypothesis that the mean prediction error equals to zero.

For black spruce, the combined model over-estimated (i.e., negative Bias_per) the tree heights about 2.71 % in ecoregion 91, 1.26 % in ecoregion 148; under-estimated (i.e., positive Bias_per) the tree heights about 1.74 % in ecoregion 88, 3.28 % in ecoregion 89, 0.47 % in ecoregion 90. The paired t test indicated the mean prediction bias significantly different from zero at $\alpha=0.05$ level.

For jack pine, the combined model over-estimated the tree heights by about 0.85 % in ecoregion 89 and 1.44 % in ecoregion 148 and under-estimated the tree heights by about 0.95 % in ecoregion 88, 0.64 % in ecoregion 90, and 0.37 % in ecoregion 91. The paired t test indicated that the mean prediction bias significantly different from zero at $\alpha=0.05$ level except in ecoregion 91.

Table 2.15. Predicting errors of applying the combined model to each ecoregion.

Ecoregion	Spp	HGT	mPred	m	n	\bar{e}	MAD	Bias_per	Se	t
88	BS	9.33	9.17	10987	0.16	1.08	1.74	1.45	11.6964*	
88	JP	8.52	8.44	7541	0.08	1.1	0.95	1.54	4.5563*	
89	BS	6.97	6.74	6512	0.23	0.77	3.28	1.17	15.7922*	
89	JP	7.34	7.4	2654	-0.06	0.98	-0.85	1.41	- 2.2929*	
90	BS	12.29	12.24	16870	0.06	1.48	0.47	1.92	3.8810*	
90	JP	14.21	14.12	13334	0.09	2.07	0.64	2.63	3.9916*	
91	BS	10.47	10.75	5811	-0.28	1.32	-2.71	1.74	-12.4143*	
91	JP	12.64	12.59	5373	0.05	1.88	0.37	2.42	1.4269	
148	BS	10.72	10.86	19346	-0.13	1.46	-1.26	1.89	- 9.9326*	
148	JP	13.35	13.54	9557	-0.19	1.82	-1.44	2.34	- 8.0563*	

Note: SPP – species; HGT_m - average of measured tree height; pred_m - average of predicted tree; n – total number of observations (height-dbh pairs); \bar{e} - average bias; MAD - absolute mean deviation; bias_per - bias %; S_e - standard deviation of the prediction errors; t - paired t test. The asterisk(*) of paired t test indicates the mean prediction bias significantly different from zero at $\alpha=0.05$ level.

2.5.5 Prediction errors of applying Alberta provincial height-diameter model

Alberta provincial height- diameter model are currently used in Manitoba for tree height estimation and volume calculations. To assess errors associated with this application, the Alberta provincial height- diameter model was applied to predict tree height for each ecoregion. Table 2.16 gives the average of measured tree height (HGT_m), average of predicted tree height (Pred_m), average bias (\bar{e}), mean absolute deviation (MAD), bias % (bias_per), standard deviation of the prediction errors (S_e) and the paired t-test for testing the null hypothesis that the mean prediction error equals to zero.

For black spruce, the Alberta provincial height-dbh model underestimated (i.e., positive Bias_per) the tree heights about 6.36 % in ecoregion 88, 7.05 % in ecoregion 89, 4.72 % in ecoregion 90, 1.97 % in ecoregion 91, and 3.53 % in ecoregion 148. The paired t test indicated that the mean prediction bias significantly differed from zero at $\alpha=0.05$ level.

For jack pine, the Alberta provincial height-diameter model overestimated (i.e., negative Bias_per) the tree heights about 5.14 % in ecoregion 90, 4.44 % in ecoregion 91 and 6.84 % in ecoregion 148. It underestimated the tree heights about 1.13 % in ecoregion 88 and 0.42 % in ecoregion 89. The paired t test indicated that the mean prediction bias significantly differed from zero at $\alpha=0.05$ level except in ecoregion 89.

Table 2.16. Predicting errors of applying Alberta provincial height-dbh model to each ecoregion.

Ecoregion	Spp	Hgt m	Pre m	n	$\bar{\epsilon}$	MAD	Bias per	Se	t
88	BS	9.33	8.74	10987	0.59	1.2	6.36	1.5	40.865*
88	JP	8.52	8.43	7541	0.1	1.23	1.13	1.8	4.726*
89	BS	6.97	6.47	6512	0.49	0.87	7.05	1.2	32.170*
89	JP	7.34	7.31	2654	0.03	1.01	0.42	1.5	1.066
90	BS	12.29	11.71	16870	0.58	1.59	4.72	2	38.078*
90	JP	14.21	14.94	13334	-0.7	2.29	-5.14	2.9	-29.470*
91	BS	10.47	10.26	5811	0.21	1.38	1.97	1.8	8.617*
91	JP	12.64	13.2	5373	-0.6	2.47	-4.44	3.2	-12.850*
148	BS	10.72	10.34	19346	0.38	1.49	3.53	1.9	27.171*
148	JP	13.35	14.26	9557	-0.9	2.08	-6.84	2.6	-33.982*

Note: SPP – species; HGT_m - average of measured tree height; pred_m - average of predicted tree; n – total number of observations (height-dbh pairs); $\bar{\epsilon}$ - average bias; MAD - absolute mean deviation; bias_per - bias %; S_e - standard deviation of the prediction errors; t - paired t test. The asterisk(*) of paired t test indicates the mean prediction bias significantly different from zero at $\alpha=0.05$ level.

2.6 Discussion and Conclusions

Based on the model fitting statistics, model [2.2] performed best in ecoregion 88, 91 for black spruce and in ecoregion 89, 91 and 148 for jack pine; model [2.3] performed the best in ecoregion 89 for black spruce and in ecoregion 88 and 90 for jack pine; model [2.1] performed the best in ecoregion 148 for black spruce and performed almost the best for black spruce in ecoregion 89, 91 and for jack pine in ecoregion 89, 90, 91 and 148.

Based on the model validation statistics, model [2.2] performed best in ecoregion 88, 91 and 148 for black spruce and in ecoregion 91 and 148 for jack pine; model [2.1] performed the best in ecoregion 89 for jack pine and in ecoregion 89, 90 for jack pine; model [2.3] performed the best in ecoregion 88 for black spruce; model [2.4] performed the best in ecoregion 90 for black spruce.

The above results demonstrated that height-diameter models were quite data related, and different data from different ecoregions displayed different height-diameter relationships. In the process of selecting the 'best' height-diameter model, residual and studentized residual plots and shapes of the height-diameter curves were examined. In addition, data-related trend and biological criteria were also considered.

Based on the fitting statistics, validation statistics and graphical examinations, model [2.2] and [2.1] are the best models to quantify height-diameter relationship for black spruce and jack pine. However, the performance of other models, such as [2.3], [2.4], is also very good. The paired t-test indicated that, in most cases, the predicted mean biases were not significantly different from zero at $\alpha=0.05$ level.

Our study indicated that the height-diameter relationship depended on ecoregions for both black spruce and jack pine. This result is not surprising given different

ecoregions have very different bio-geoclimatic conditions. Applying the combined height-diameter model to each ecoregion resulted in height predicting errors up to 3.28% for black spruce and 1.44 % for jack pine. Each ecoregion appeared to need a unique height-diameter relationship, depending on local soil and ecological conditions.

Applying Alberta provincial height-diameter model to each ecoregion in Manitoba resulted in height prediction errors up to 7.05% for black spruce and 6.84 % for jack pine. Therefore, 'borrowing' existing height-diameter models from other provinces should be avoided whenever possible.

Development of ecoregion-based growth and yield relationships, such as height-diameter models, is an essential step toward the implementation of ecosystem-based forest management. Often, growth and yield relationships are different among different ecoregions, so wherever data permit, it is preferable to develop ecoregion-specific models. These localized models allow for the uniqueness of each individual ecoregion to be captured. Consequently, ecoregion-based models are able to provide more reliable predictions.

Since the data from some ecoregions, especially ecoregion 91, were mainly from relatively younger PSPs and the ranges of dbh and height were quite narrow, it is advised that they may not necessarily apply to other prediction purposes, such as volume sampling, TSPs. Independent tests are recommended should such an application be required.

CHAPTER 3. FITTING HEIGHT-DIAMETER MODELS INCORPORATED WITH STAND VARIABLES FOR JACK PINE

3.1 Introduction

Lately, the Manitoba Growth and Yield Program has identified an urgent need to develop local tree height-diameter equations which can be used to estimate tree heights and consequently tree volumes because such information has been lacking in Manitoba. However, the height-diameter relationship of a given species may depend on local environmental and stand conditions thus varying within a geographic region (Huang et al. 2000; Zhang et al. 2002). In order to increase the height prediction accuracy and, at the same time do not increase the sampling cost, the base models using dbh (diameter at breast height, 1.3 m) as the sole predictor were modified to include stand variables, such as stand density, basal area of a stand, species compositions by using the parameter prediction approach (Clutter et al. 1983; Huang and Titus 1994; Staudhammer and LeMay 2000).

The objective of this chapter was to test if adding stand variable to height-diameter equation improves height prediction. Because height-diameter relationship varies among ecoregions (Huang et al. 2000; Zhang et al. 2002), my test was conducted within a single ecoregion, ecoregion 91.

3.2 Study Area

The study area is located in ecoregion 91 (Lake of the Woods ecoregion) in Manitoba. A brief description of ecoregion 91 is given in Chapter 2.

3.3 The Data

A total of 63 provincial permanent sample plots (PSP) were used in this study. The data were collected over the last 24 years and the PSPs were randomly located throughout the inventory areas of the ecoregion to provide representative information for a variety of subtype, site class, cutting class, crown closure, soil type, moisture class, and etc. The plot size was 12.62m in radius (500 m²). On each plot, each tree was identified to species and its dbh and height were measured. Every plot was re-measured in five-year period. A detailed description of the procedures for collecting the data can be found in Manitoba forest inventory field instruction manual (Manitoba Conservation 1998).

For introducing the stand density variables into the base models, the original PSP data were summarized to calculate: (1) basal area per hectare for all species (BASUM), (2) the number of trees per hectare for all species combined (tree_pha), and (3) percent basal area for the species being modeled (SC_SP). To avoid serial correlations among repeated measurements of the same plot, only the last measurements were selected. A total of 3022 measured jack pine trees with healthy conditions from the 63 plots were used to fit the models. The summary statistics are shown in Table 3.1. For model validation purpose, 80 % of all the data were randomly selected and used for model fitting and the remaining 20% of the data were used for model validation.

Table 3.1. PSP summary statistics for stands with jack pine in ecoregion 91.

Ecoregion 91	Mean	Minimum	Maximum	SD	n
Basal area (m ² /ha)	28.48	5.24	63.9	9.88	63
DBH (cm)	9.66	0.3	30.3	5.85	3022
Height (m)	10.86	1.4	23.5	5.06	3022
Mean dbh (cm)	10.18	2.83	18.22	3.58	63
No.of trees/ha	3709	1040	17620	3236	63
Species composition	0.35932	0.00001	1	0.43744	63

Note: SD-Standard deviation.

3.4 Selecting base model for introducing stand density variables

Based on a thorough review of many proposed models presented by Huang et al. (1992), Staudhammer et al. (2000) and the previous comparison analysis in chapter 2, the Weibull-type model (Ratkowsky 1983; Huang et al. 1992) was selected as the base model:

$$[3.1] \quad H = 1.3 + a(1 - e^{-bD^c})$$

Stand density variables were then added to the base model, following the methods given by Huang and Titus (1994) and Staudhammer and LeMay (2000). Five modified models are presented here. Model [3.2] and [3.3] directly incorporated the stand variables into the base model, while model [3.4], [3.5] and [3.6] incorporated stand variable into linear models to predict each of the parameters of the base model:

$$[3.2] \quad H = 1.3 + a_1(1 - e^{-a_2 D^{a_3} * TREE_PHA^{a_4} * BASUM^{a_5}})$$

$$[3.3] \quad H = 1.3 + a_1(1 - e^{-a_2 D^{a_3} * TREE_PHA^{a_4} * BASUM^{a_5} * SC_SP^{a_6}})$$

$$[3.4] \quad a = a_0 + a_1 TREE_PHA + a_2 BASUM + a_3 SC_SP$$

$$b = b_0 + b_1 TREE_PHA + b_2 BASUM + b_3 SC_SP$$

$$c = c_0 + c_1 TREE_PHA + c_2 BASUM + c_3 SC_SP$$

$$H = 1.3 + a(1 - e^{-bD^c})$$

$$[3.5] \quad a = a_0 + a_1 TREE_PHA + a_2 BASUM + a_3 SC_SP$$

$$b = b_0 + b_1 TREE_PHA$$

$$c = c_0 + c_1 TREE_PHA + c_2 BASUM$$

$$H = 1.3 + a(1 - e^{-bD^c})$$

$$[3.6] \quad a = a_0 + a_1TREE_PHA + a_2BASUM + a_3SC_SP$$

$$b = b_0 + b_1TREE_PHA$$

$$c = c_0 + c_1TREE_PHA$$

$$H = 1.3 + a(1 - e^{-bD^c})$$

where BASUM is the basal area per hectare for all species, TREE_PHA is the number of trees per hectare for all species combined and SC_SP is the percent basal area for the species being modeled, it is defined as:

$$[3.7] \quad SC_SP = BASUM_SP / BASUM$$

where BASUM_SP is the basal area per hectare (square meter per hectare) for the target species, a_0 to a_6 , b_0 to b_3 and c_0 to c_3 are coefficients to be estimated.

3.5 Model fitting and analysis

All models were fitted using the Gauss-Newton method of the SAS non-linear least squares procedures (SAS/STAT, SAS Institute Inc.1990). To ensure the solutions of the non-linear regressions are global rather than local least squares solutions, different starting values of the model parameters were provided for the fits. Following the model fitting, studentized residual plots were examined for outliers, lack-of-fit and unequal variance. After the residual analysis, weighted non-linear least squares were applied to refit the models. Models were compared using the R-squared values (R^2), root mean squared error (RMSE), average bias (\bar{e}), bias % (bias_per), mean absolute deviation (MAD) calculated by equations [2.12] to [2.16], respectively.

3.6 Results

3.6.1 Fitting models incorporated with stand variables

Model [3.2], [3.3], [3.4], [3.5], [3.6] were fitted and compared based on model performance measured by R^2 , RMSE, asymptotic t-statistics for the parameters of the models. Model [3.6] was found the best model and the results for the final fits of the model [3.6] are presented in Table 3.3. For comparing purposes, the base model [3.1] was fitted on the same data and presented in Table 3.2. The studentized residuals against the predicted height from weighted nonlinear least squares are shown in Figure 3.1, showing a homogenous band of the studentized residuals with zero mean, and it indicates the assumptions of the regression analysis are almost met.

Comparing model [3.1] with model [3.6] based on the same database, R^2 increased 0.9657 to 0.9783; RMSE decreased from 0.545 to 0.4141; average bias (\bar{e}) changed from -0.0026 to 0.00054 . These results indicated that model [3.6] had the higher prediction accuracy than model [3.1], which suggests that incorporating stand variables into base model enhances the performance of prediction model.

Table 3.2. Fit statistics of weighted nonlinear height-diameter model [3.1] for jack pine in ecoregion 91.

a	b	c	n	R ²	MSE	RMSE	\bar{e}	Bias_per
19.3892	0.0508	1.2086	3022	0.9657	0.297	0.545	-0.0026	-0.0242

Note: \bar{e} - average bias; MSE - mean square error; RMSE - root mean square error; bias_per - average bias percent.

Table 3.3. Fit statistics for jack pine height prediction model [3.6] in ecoregion 91.

Parameter	Estimate	Se	t	p-value	R ²	RMSE	\bar{e}	bias_per	n
a0	8.9647	0.4298	20.86	<0.001	0.9783	0.414	0.00054	0.005	3022
a1	-0.00025	0.0000	-6.25	<0.001					
a2	0.2586	0.0078	32.98	<0.001					
a3	1.0362	0.3188	3.25	<0.001					
b0	0.0514	0.0036	14.12	<0.001					
b1	0.00001	0.0000	11.93	<0.001					
c0	1.2673	0.0310	40.88	<0.001					
c1	-0.00003	0.0000	-9.77	<0.001					

Note: Se—standard error; RMSE—root mean square error; bias_per – average bias percent; R² - prediction coefficient of determination.

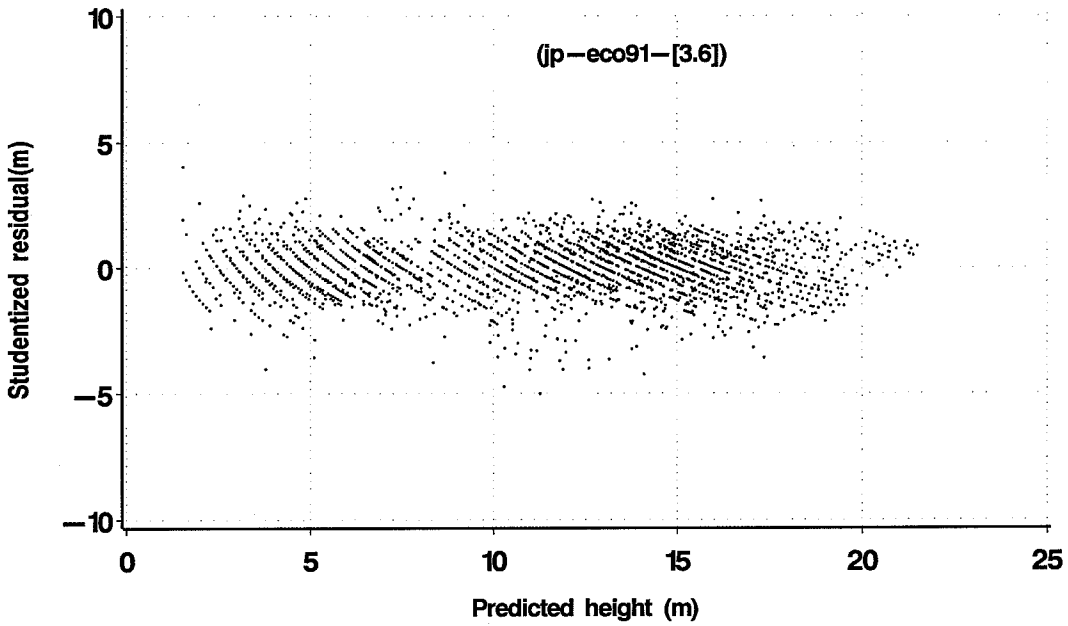


Figure 3.1. Plot of studentized residuals against the predicted height for jack pine from weighted nonlinear least squares of model [3.6].

3.6.2 Model validation

The measured height values from the validation data set were compared to values predicted by model [3.1] and [3.6] using the estimated parameters in Table 3.2 and 3.3. The bias of the prediction was obtained by subtracting the predicted height from the measured height. For model [3.6], the average bias (\bar{e}) was -0.0718 ; the standard deviation of the prediction bias (S_e) was 1.381 ; the root mean square error of prediction ($RMSE_p$) was 0.4405 ; the percent bias ($bias_per$) was -0.65% . The prediction coefficient of determination (R_p^2) is 0.9763 ; the t-test of the null hypothesis that the mean prediction bias was zero was conducted according to the method described by Rawlings (1988). The calculated $t = -1.279$ with 604 degrees of freedom, was not significant at $\alpha=0.05$ level, hence the mean prediction bias was not significantly different from zero. Similarly, the validation results of model [3.1] are: $\bar{e} = -0.0958$; $S_e = 1.659$; $RMSE_p = 0.5471$; $\bar{e} \% = -0.87\%$; $R_p^2 = 0.9631$; $t = -1.421$ with 604 degrees of freedom, was not significant at $\alpha=0.05$ level, so both model [3.1] and [3.6] are valid models.

3.7 Discussion and conclusions

Precise tree height estimations are needed because height is an important variable in volume estimation and biomass calculation. However, in Manitoba's volume sampling program, especially in its temporary sample plots (TSP) measurement, height measurements are time consuming. The prediction precision may be poor if only DBH is used as an independent, since stand variables, such as stand density, basal area, species composition, also have great impact on tree height growth. On the other hand, introducing stand variables do not cost extra effort in the volume sampling programs; they can be extracted from the existed permanent sample plots (PSP) or TSP data. So, if forest stand variables such as tree basal area, stand density and competition, site productivity, and species composition are available, such as having enough PSP data in ecoregion 91, the function [3.6] can be used to get better height estimation.

Results from this study indicated that incorporating stand variables into the base model, in the form of model [3.6], improved prediction accuracy. Compared to the base model [3.1], model [3.6] has higher R^2 and lower RMSE and average bias (\bar{e}). When tested against the validation data, model [3.6] was lower in average bias (\bar{e}), the root mean square error of prediction ($RMSE_p$) and the percent bias (bias_per) and higher in the prediction coefficient of determination (R_p^2).

In conclusion, incorporating stand variables into base height-diameter model increased the prediction accuracy, and model [3.6] was found the appropriate model. On the other hand, because the fitting statistics, such as R^2 value of the simple height-diameter model is very high already, so, if stand variables are not available, the simple height-diameter model is still good enough for prediction purpose.

CHAPTER 4. SITE INDEX MODELS DEVELOPMENT FOR BLACK SPRUCE AND JACK PINE

4.1 Introduction

Site index is usually defined as the average height of the dominant and co-dominant trees 50 years after they reached breast height (1.3m) (Carmean 1975; Goelz and Burk 1992; Huang, et al. 1994b; Philip 1998; Carmean 2001). It provides standardized comparisons of productive potential between sites, across a broad range of existing stand conditions. Dominant trees are the tallest trees and co-dominant trees are those with crowns forming the general level of the crown cover and receiving full light from above, but comparatively little from the sides, usually with medium crowns, more-or-less crowned on the sides in the stand (Manitoba Conservation. 1998). Directly estimating site index from forest trees is the most commonly used method for estimating site quality in North America (Carmean 1975; Goelz and Burk 1992; Carmean 2001; Huang et al. 1997).

Site index is an important parameter in growth and yield modelling, forest site productivity estimation and classification, and improving the accuracy and precision of forest inventory projections and provincial wood supply analysis. However, such information is lacking in Manitoba.

The main objective of this study was to develop reference-age invariant polymorphic height growth and site index models for black spruce and jack pine in Manitoba. These models can be used simultaneously to estimate site index from height and age and height from site index and age.

4.2 The data

Stem analysis or sectioning is a commonly used mensurational technique to show how a tree grows in height, in diameter and in form (Husch, Miller and Beers 1972). Past development of tree height, diameter, form and volume can be determined by ring count and by measuring the increase in diameter on each cut surface of a felled and sectioned tree (Spurr 1952). Stands sampled for stem analysis were selected according to the following criteria: the species compositions of the stands are of 80-100% jack pine or black spruce; stands should be fully stocked and mature; stands should be even-aged; stands should have little or no disturbances since initiation.

The following procedures were used for field sampling and measurement: set up a 300 m² circular (radius = 9.77 m) plot within each selected stand, keeping approximately 100 m away from stand boundaries and roads; tally all trees (minimum 7cm DBH) within the plot; start from the plot center to the north, count the trees clockwise, select and mark the 1st, 5th, 10th, 15th, 20th, 25th,trees and indicate crown class (D-dominant; C- co-dominant; I - intermediate; S- suppressed) in the tally sheets; select the 3 largest DBH trees (dominant or co-dominant), which are relatively free of visible breakages, large forks, or other growth interruption abnormality, from the plot.

Before felling each selected tree, mark the north side of each selected tree up to breast height (1.3 m) at 0.33m, 0.67m, 1.0m, and 1.3m by cuts along the north side of the stem; record the diameters at each of these heights as well as at the ground level; determine crown width. After felling each selected tree, record the height to live crown and the total height; delimit the tree. Cookies were taken at 0.33 m, 0.67 m, 1.0 m, 1.3

m, and at 1.3 m intervals above breast height until the tree reached a diameter of approximately 7 cm. After the 7.0 cm DOB (diameter outside bark) point on the tree, cookies were taken at 20 cm intervals until the tree reached a diameter of at least 4cm. The north direction was transferred onto each cookie. Tree number, cookie position and the cookie number were also recorded. All cookies collected were aged by counting number of rings along the north and east sides. Ring counts started from the cambium to the pith. Every 10th ring was marked to facilitate recounting.

Totally 352 sectioned site trees (dominant and co-dominant trees) were chosen from five ecoregions in Manitoba. These trees included 180 black spruce trees and 172 jack pine trees. Table 4.1 and Table 4.2 showed the distribution of the 352 black spruce and jack pine trees by site index and age class. For model validation purpose, 80 % of all the data were randomly selected and used for model fitting and the remaining 20% of the data were used for model validation.

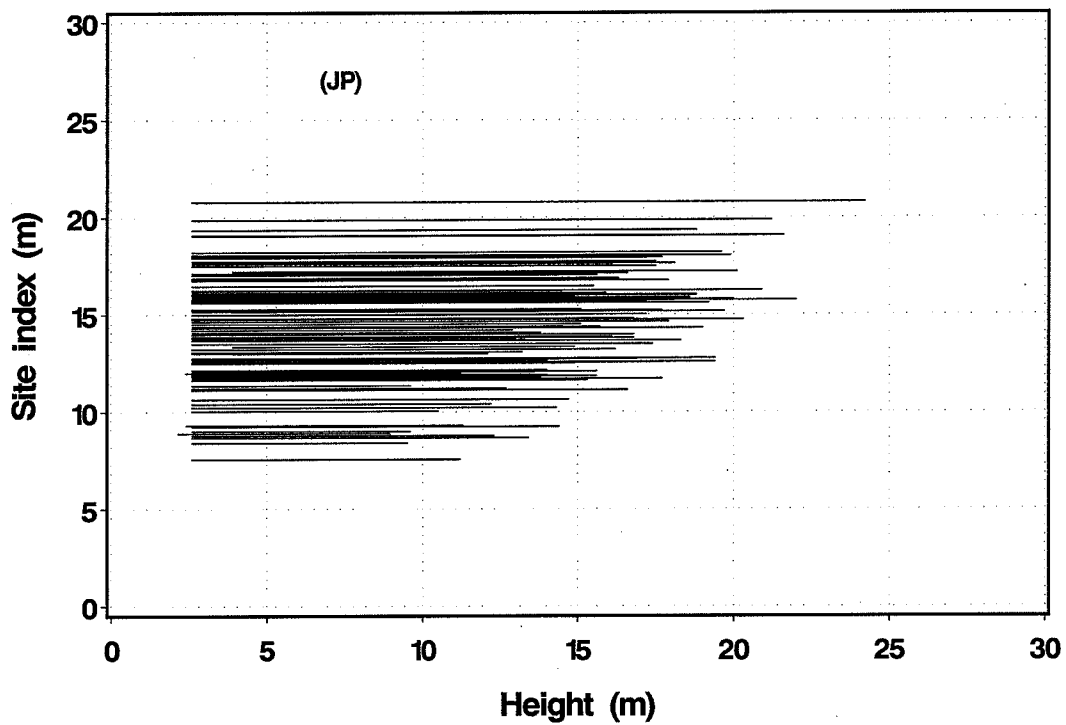
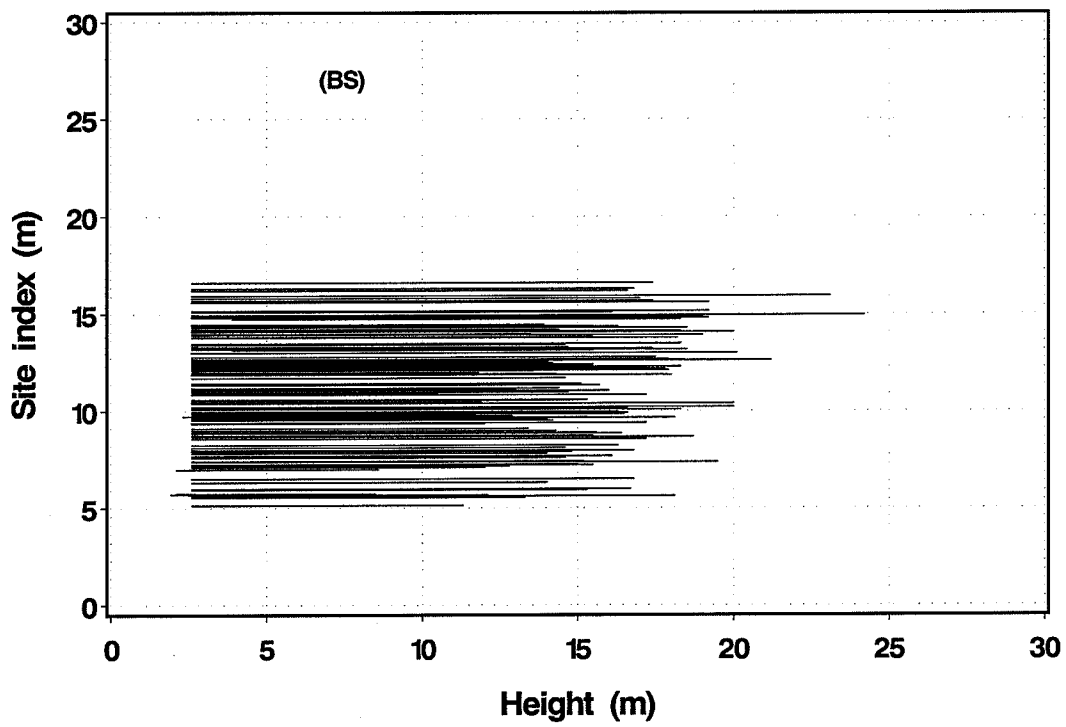


Figure 4.1. Site index and height distribution of the black spruce and jack pine trees used for model fitting. *(BS)* Calculated site index versus sectioned tree height data of black spruce. *(JP)* Calculated site index versus sectioned tree height data of jack pine.

From Table 4.1, we found that the site index of black spruce trees was quite low and the range of breast height age was wide. It was difficult to find stands that have higher site index (or higher productivity), especially at old age class. Most black spruce stands in Manitoba have low productivity, and a small proportion of high productivity black spruce stands may have been harvested.

From Table 4.2, we found that most of the site index of jack pine trees are in the range of 7 to 20 and the range of the tree breast height age are very narrow. Most of the jack pine trees are less than 110 years, which may be attributed to frequent forest fire in Manitoba.

Figure 4.1 showed the distribution of heights and site indices of the sectioned black spruce and jack pine in this study. For black spruce and jack pine, most of the tree heights of the sectioned tree were less than 25 meters and the measured site indices were about 5-15 m for black spruce and about 10-20 m for jack pine.

Table 4.1. Summary of the data used for site index model development of black spruce.

Breast height age class (years)	Site index (m)					Other	Total
	<=9	9-10.9	11-12.9	13-14.9	15-16.9		
<=49						27	27
50-59		1		2			3
60-69	1	1	3	5	6		16
70-79	2	2	6	7	3		20
80-89	2	4	6	5	1		18
90-99	3	14	8	9	2		36
100-109	5	8	4	1			18
110-119	11	3	3				17
120-129	6	1	1	1			9
130-139	5	1					6
140-149	3	1					4
150-159	2	1					3
Total	40	37	31	30	12	27	177

Table 4.2. Summary of the data used for site index model development of jack pine.

Breast height age class (years)	Site index (m)							Other	Total
	<=9	9-10.9	11-12.9	13-14.9	15-16.9	17-18.9	19-20.9		
<=49								52	52
50-59	1		3	6	8	6	1	1	26
60-69		1	1	5	3	4	2		16
70-79	1	3	5	7	11	3	2	1	33
80-89			2	2	1				5
90-99			4	3	2	1			10
100-109	3	2	7	4	7				23
110-119		1	1		1				3
120-129				1	2			1	4
Total	5	7	23	28	35	14	5	55	172

4.3 Model Selections and Analysis

4.3.1 Selected height growth and site index models

The following list is the most commonly used height and site index models found in the site index literature:

$$(4.1) \quad H = 1.3 + (SI - 1.3) \cdot \left\{ \frac{1 + b_0 (SI - 1.3) + \exp[b_1 + b_2 \ln(50 + b_3) - \ln(SI - 1.3)]}{1 + b_0 (SI - 1.3) + \exp[b_1 + b_2 \ln(Bhage + b_3) - \ln(SI - 1.3)]} \right\}$$

(Huang 1997)

$$(4.2) \quad H = 1.3 + (SI - 1.3) \cdot \left\{ \frac{1 + \exp[b_1 + b_2 \ln(50 + b_3) - \ln(SI - 1.3)]}{1 + \exp[b_1 + b_2 \ln(Bhage + b_3) - \ln(SI - 1.3)]} \right\}$$

(Monserud 1984; Huang 1997)

$$(4.3) \quad H = 1.3 + (SI - 1.3) \cdot \left\{ \frac{1 + \exp[b_1 + b_2 \ln(50 + b_3) - b_4 \ln(SI - 1.3)]}{1 + \exp[b_1 + b_2 \ln(Bhage + b_3) - b_4 \ln(SI - 1.3)]} \right\}$$

(Huang 1997)

$$(4.4) \quad H = 1.3 + b_1 (SI - 1.3)^{b_2} \left(1 - K^{\frac{Bhage}{50}}\right)^{b_3 (SI - 1.3)^{b_4}} \quad \text{where}$$

$$K = 1 - \left\{ \frac{SI - 1.3}{b_1 (SI - 1.3)^{b_2}} \right\}^{\frac{1}{b_3 (SI - 1.3)^{b_4}}} \quad \text{(Carmean et al. 2001)}$$

$$(4.5) \quad H_2 = 1.3 + (H_1 - 1.3) \cdot \left(\frac{1 - \exp(-b \cdot T_2)}{1 - \exp(-b \cdot T_1)} \right)^c \quad \text{where}$$

$$b = b_0 (H_1 - 1.3)^{b_1} b_2^{(H_1 - 1.3)}$$

$$c = b_3 (H_1 - 1.3)^{b_4} T_1^{b_5} \quad \text{(Clutter et al. 1983; Huang 1997)}$$

$$(4.6) \quad H_2 = 1.3 + (H_1 - 1.3) \cdot \left(\frac{1 - \exp(-b \cdot T_2)}{1 - \exp(-b \cdot T_1)} \right)^c \quad \text{where}$$

$$b = b_0 (H_1 - 1.3)^{b_1} b_2^{(H_1 - 1.3)/T_1}$$

$$c = b_3 (H_1 - 1.3)^{b_4} T_1^{b_5} \quad (\text{Clutter et al.1983; Huang 1997})$$

where H = total height (m); SI=site index (m); Bhage = breast height age(years); H₁=tree height (m) at time one; H₂=tree height (m) at time two; T₁=breast height age (years) at time one; T₂=breast height age (years) at time two; b₀, b₁, b₂, b₃, b₄, b₅=parameters to be estimated; exp = base of the natural logarithm (≅2.71828); 1.3=a constant commonly used because of the definition of site index.

Models [4.1] to [4.3] are modified and constrained versions of Monserud's (1984) logistic-type model. Model [4.4] is Newnham (1988) constrained version of the EK (1971) nonlinear regression model. Models [4.1] to [4.4] require site index (SI) as an independent variable, we first have to calculate the site index before the model parameter estimation. Models [4.5] to [4.6] are Chapman-Richards non-reciprocal difference equation models (Huang 1997).

4.3.2 Data structure

For fitting models [4.1] to [4.4], site index (SI) is required as an independent variable. Site index of each tree was calculated before the model parameter estimation. For fitting the difference equation model, such as models [4.5] and [4.6], which expressed in the general form of $H_2 = f(H_1, T_1, T_2)$, different data structures can be used (Borders et al 1988; Burk 1992; Cao 1993; Huang 1997). All possible growth intervals were used in this study.

4.3.3 Model fitting and analysis

All models were fitted using the Gauss-Newton method of the SAS non-linear least squares procedures (SAS/STAT, SAS Institute Inc.1990). To ensure the solutions of

the non-linear regressions are global rather than local least squares solutions, different starting values of the model parameters were tested. Following the model fitting, Studentized residual plots were examined for outliers, lack-of-fit and unequal variance. Models were compared using the R-squared values (R^2), root mean squared error (RMSE), average bias (\bar{e}), bias % (bias_per), mean absolute deviation (MAD) calculated by equation [2.12] to [2.16] respectively.

4.3.4 Model validation

Model validation is an important part of model development. It is performed to increase the credibility and gain sufficient confidence about a model. The following commonly used statistical tests in model validation were used to assess the predictive ability of fitted models.

4.3.4.1 The paired t test

The null hypothesis of the paired t test is that the mean prediction error is 0 (Snedecor and Cochran 1980). It is written as follows:

$$[4.7] \quad t = \frac{\bar{e}}{S_e / \sqrt{n}}$$

Where \bar{e} is the mean prediction error, S_e is the standard deviation of the prediction errors and n is the number of observations. This test statistic has $(n - 1)$ degrees of freedom. Large t values indicate large mean prediction errors.

4.3.4.2 The Separate t tests

A widely used method of validating a model is to evaluate the simple linear regression between observed and predicted y, values $y_i = b_0 + b_1 \hat{y}_i$. For an acceptable

model, the regression will be a 45° line through the origin. The adequacy of the model can be evaluated by testing separately whether $b_0 = 0$ and $b_1 = 1$ through separate t tests (Montgomery and Peck 1992). To test $b_0 = 0$, the t value can be read directly from the fit $y_i = b_0 + b_1 \hat{y}_i$. To test $b_1 = 1$, the test statistic is calculated by:

$$[4.8] \quad t_{b_0} = (b_1 - 1) / \sqrt{\left[\sum (y_i - \tilde{y}_i)^2 / (n - 2) \right] / \sum (\hat{y}_i - \bar{\hat{y}})^2}$$

Where \tilde{y}_i is the predicted value from $y_i = b_0 + b_1 \hat{y}_i$, and $\bar{\hat{y}}$ is the mean of \hat{y}_i values.

This test statistic has $(n - 2)$ degree of freedom.

4.3.4.3 Simultaneous F test

Similar to the separate t test, the simultaneous F test is designed to evaluate the regression $y_i = b_0 + b_1 \hat{y}_i$ by testing whether $b_0 = 0$ and $b_1 = 1$ simultaneously (Montgomery and Peck 1992). The test statistic is calculated by:

$$[4.9] \quad F = \frac{n(b_0 - 0)^2 + 2 \sum \hat{y}_i (b_0 - 0)(b_1 - 1) + \sum \hat{y}_i^2 (b_1 - 1)^2}{2 \sum (y_i - \tilde{y}_i)^2 / (n - 2)}$$

where the variables are as previously defined in the separate t tests. The degrees of freedom are 2 and $(n - 2)$ for the numerator and denominator, respectively.

4.3.4.4 The novel test

The novel test regresses the prediction errors on the sum of observed and predicted values via a simple linear regression: $e_i = b_0 + b_1 (y_i + \hat{y}_i)$. The validity of the model being tested is determined by testing $b_0 = b_1 = 0$ jointly based on the extra sum of squares method (Kleijnen et al. 1998):

$$[4.10] \quad F = [(n - 2)(SSE_R - SSE_F)] / (2SSE_F)$$

where $SSE_R = \sum e_i^2$ and $SSE_F = \sum (e_i - \hat{e}_i)^2$ are the reduced and the full sums of squared errors, respectively. The degrees of freedom are 2 and $(n - 2)$ for the numerator and denominator, respectively.

The following commonly used prediction statistics such as the R-squared values (R_p^2), root mean squared error ($RMSE_p$), average bias (\bar{e}), bias % (bias_per), mean absolute deviation (MAD), standard deviation of the prediction errors (S_e) were also used to evaluate the fitted models using the validation data set. Both average bias and bias % give an average measure of the differences between the predicted and the observed tree height, with mean prediction bias in absolute terms and percent bias in relative terms. RMSE incorporates both bias and variation and is a better measure of a model performance. Besides, the examinations of plotting the site index curves were also conducted to make sure the fitted model biological meaningful.

4.4 Results

4.4.1 Fitting models

The fitted nonlinear least squares estimates of the parameters, root mean squared error (RMSE), and the coefficient of determination (R^2) for models [4.1] to [4.6] are shown in Table 4.3. The asymptotic t-statistics, average bias (\bar{e}), mean absolute deviation (MAD) are shown in Table 4.4. Because the data structures and sample sizes used to estimate the models were different, the comparison was done among model [4.1] to [4.4] and between model [4.5] and [4.6], respectively.

Table 4.3. Estimated parameters and fitted statistics of model [4.1] to [4.6] for black spruce and jack pine.

SPP	b0	b1	b2	b3	b4	b5	n	R2	RMSE	Model
BS	0.0596	11.8330	-2.0845	16.2555			1639	0.969	0.856	[4.1]
JP	-0.0103	7.0168	-1.0939	1.5081			1357	0.963	0.968	[4.1]
BS		11.8837	-2.1876	19.3051			1639	0.968	0.865	[4.2]
JP		7.1520	-1.0849	1.3789			1357	0.963	0.969	[4.2]
BS		12.2198	-2.0810	16.2289	1.3788		1639	0.969	0.855	[4.3]
JP		7.0199	-1.0880	1.4211	0.9469		1357	0.963	0.969	[4.3]
BS		9.0131	0.4560	4.8705	-0.6165		1639	0.967	0.874	[4.4]
JP		10.7415	0.3096	0.6393	0.1309		1357	0.963	0.969	[4.4]
BS	0.0195	-0.4712	1.0397	0.6591	-0.5022	0.4266	27192	0.943	1.177	[4.5]
JP	0.0293	-0.3606	1.0315	0.7865	-0.5098	0.4044	19814	0.925	1.398	[4.5]
BS	0.0146	-0.1439	0.9365	0.6348	-0.4543	0.4113	27192	0.942	1.184	[4.6]
JP	0.0156	0.0347	1.5468	0.7594	-0.3838	0.3389	19814	0.925	1.399	[4.6]

Note: RMSE - root mean square error of prediction; R^2 - prediction coefficient of determination.

Table 4.4. Prediction statistics of model [4.1] to [4.6] for black spruce and jack pine.

SPP	n	\bar{e}	MAD	Bias_per	Se	t	Model
BS	1639	0.055	0.603	0.484	0.853	2.609	[4.1]
JP	1357	0.091	0.722	0.826	0.963	3.489	[4.1]
BS	1639	0.048	0.622	0.423	0.863	2.254	[4.2]
JP	1357	0.087	0.722	0.792	0.964	3.342	[4.2]
BS	1639	0.055	0.602	0.483	0.852	2.608	[4.3]
JP	1357	0.089	0.722	0.803	0.964	3.390	[4.3]
BS	1639	0.075	0.617	0.657	0.870	3.474	[4.4]
JP	1357	0.085	0.724	0.771	0.964	3.254	[4.4]
BS	27192	0.048	0.781	0.416	1.176	6.775	[4.5]
JP	19814	0.081	0.981	0.714	1.396	8.194	[4.5]
BS	27192	0.104	0.789	0.890	1.179	14.481	[4.6]
JP	19814	0.128	0.981	1.122	1.393	12.897	[4.6]

Note: SPP – species; n – total number of observations (height-Bhage pairs) from sectioned trees; \bar{e} - average bias; bias_per - bias %; MAD - absolute mean deviation; bias_per - bias %; S_e - standard deviation of the prediction errors;

4.4.2 Model comparison by prediction statistics

For black spruce, models [4.1], [4.2], [4.3] and [4.4] performed quite well. The R^2 values were very high (> 0.95) and the RMSEs were small (< 1.00). Prediction biases of these models were very low (< 0.08 m), with model [4.2] having the lowest average bias (\bar{e}) and bias_per, 0.048 m and 0.42%, respectively. Model [4.4] showed the highest \bar{e} and bias_per, 0.075 m and 0.66%, respectively. Model [4.2] also showed the lowest value of paired t test, 2.254. When compared to model [4.6], model [4.5] had higher R^2 and lower RMSE, \bar{e} , bias_per, and MAD (Table 4). Model [4.5] also had lower value of paired t test, 6.775. These results indicated that model [4.5] performed better than model [4.6].

For jack pine, models [4.1], [4.2], [4.3] and [4.4] also performed very well. The R^2 values were very high (> 0.95) and the RMSEs were small (< 1.00). Prediction biases of these models were very low, less than 0.10 m. Model [4.4] had the lowest average bias (\bar{e}) and bias_per, 0.085 m and 0.77%, respectively. Models [4.5] and [4.6] were similar in R^2 and RMSE. However, model [4.5] had lower \bar{e} and bias_per. The mean absolute deviations (MAD) of model [4.5] and [4.6] were the same.

The above comparisons indicated that model [4.2] was the 'best' model among models [4.1] to [4.4], and model [4.5] was better than model [4.6] for both black spruce and jack pine in Manitoba.

4.4.3 Model comparison by graphical performance

Figure 4.2 and figure 4.3 show the residuals plotted against predicted height from model [4.1] to model [4.6] of black spruce and jack pine, respectively. The residual plots

of model [4.1] to [4.4] revealed an approximately horizontal band of the data points, with zero crossing the center of the band. On the other hand, residual plots of model [4.5] to [4.6] were quite poor.

Figure 4.4 displays site index prediction biases from model [4.1] to [4.6] over time for black spruce. Each line represents a tree on the graph. At the reference age of 50 years, site index prediction biases were equal zero. This is expected because these models were constrained to pass through the reference age to meet the definition of site index. Patterns displayed in all graphs were quite similar. However, the prediction errors were much larger at young age classes (Figure 4.4).

Figure 4.5 shows the average site index prediction biases of models [4.1] to [4.6] by breast height age classes for black spruce. The solid line in the middle represents the average site index prediction errors and the two dashed lines ($\bar{e} \pm$ standard deviation) connected the standard deviations associated with the average errors. Regardless the model, the average site index prediction errors were significant different from zero at the age younger than 20 years. However, models [4.5] and [4.6] were much more extreme than models [4.1] to [4.4] at young age classes. Therefore, site index predictions for younger stands were not reliable, even if the best models were developed and applied.

Figure 4.6 shows the average height prediction biases of models [4.1] to [4.6] by breast height age classes for black spruce. The solid line in the middle represents the average height prediction errors and the two dashed lines ($\bar{e} \pm$ standard deviation) connected the standard deviations associated with the average errors. Models [4.1], [4.2] and [4.3] were quite reliable for the height prediction at the age younger than 110 years given site index. Beyond that age range, the prediction errors increased sharply. Model

[4.4] displayed essentially similar results but a little bit poorer than models [4.1], [4.2] and [4.3]. On the other hand, models [4.5] and [4.6] produced very high prediction errors at the younger stand ages. For older stands, the prediction errors are quite reasonable.

Figure 4.7 to figure 4.12 show the site index curves generated from model [4.1] to [4.6], overlaid with the observed tree sectioning data of black spruce and jack pine. Site index values at 50 years were given for each site index curves. When the site index curves proceeded from low to high, the shape of the curves changed. The lower site index curves displayed flatter, linear like curves while the higher site index curves approached the asymptote more rapidly. The site index 25 curve in figure 4.10 of model [4.4], was very flat after age 100 years for black spruce. The site index curves ($si = 20$ and $si = 25$) in figure 4.11 and figure 4.12 displayed extreme trends over the time for black spruce. Compared to site index curves generated by other models, the site index curves generated by models [4.2], [4.1] and [4.3] were more reasonable.

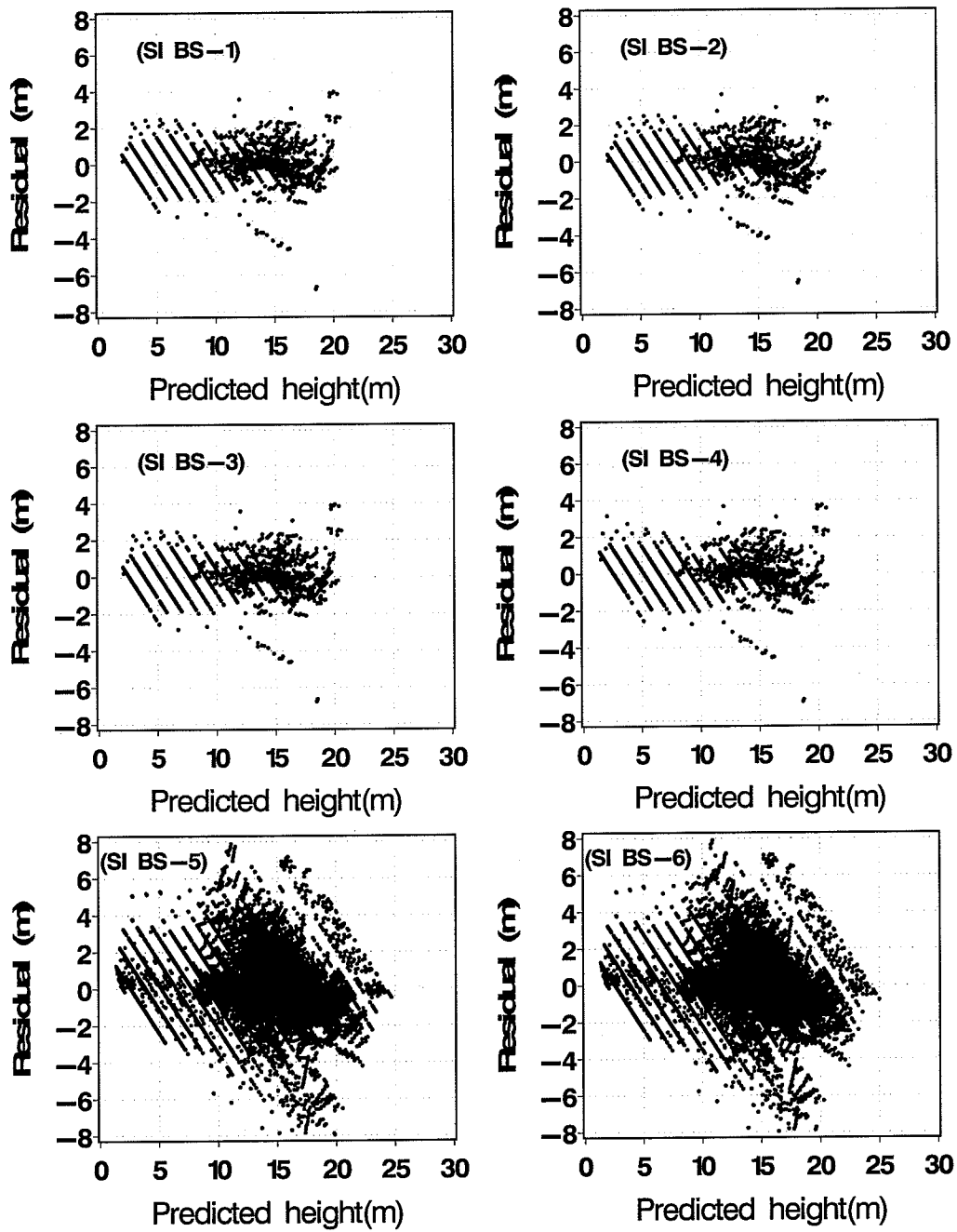


Figure 4.2. Residuals plotted against predicted height of black spruce.

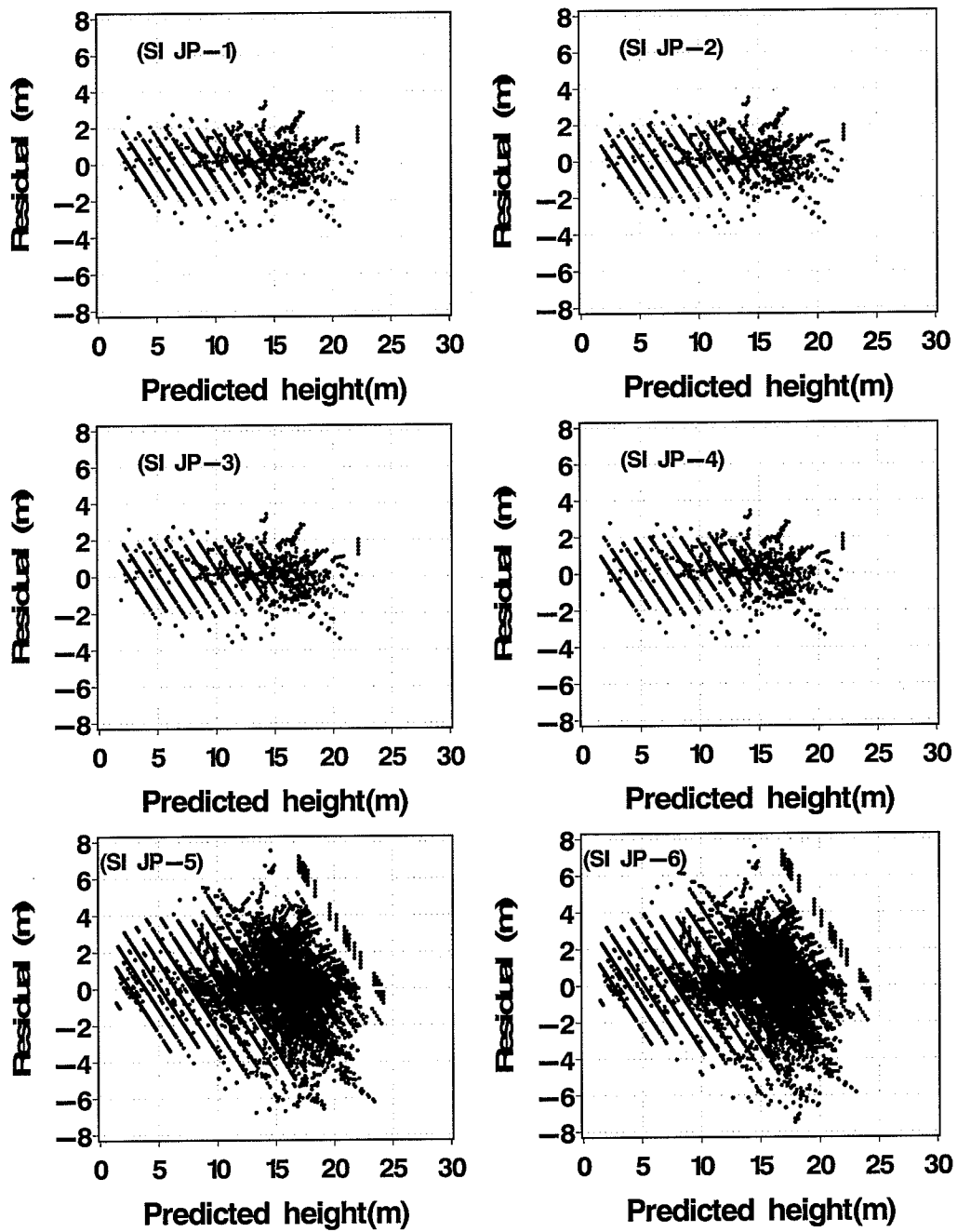


Figure 4. 3. Residuals plotted against predicted height of jack pine.

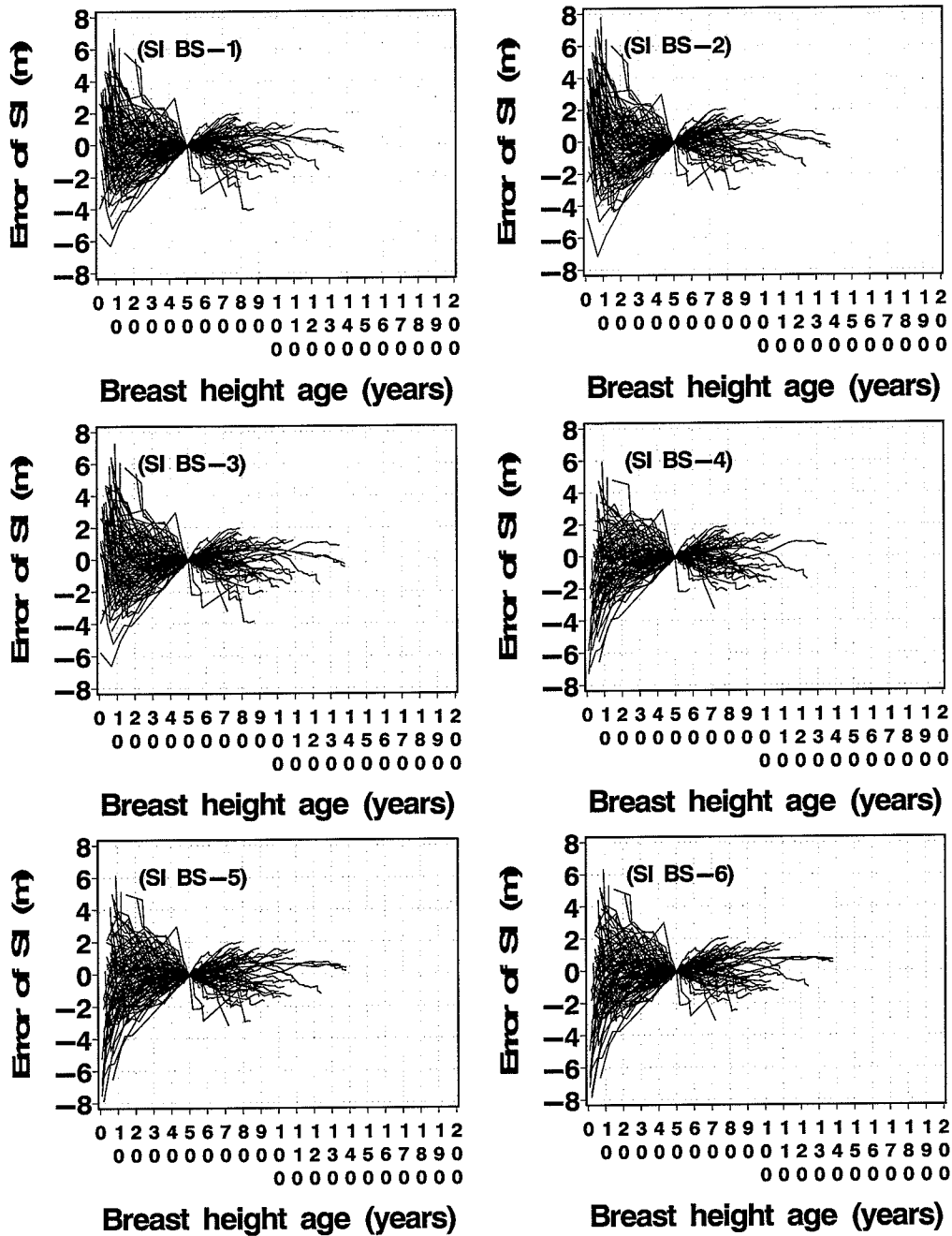


Figure 4.4. Site index prediction biases by breast height age classes of black spruce from model [4.1] to [4.6]. Each line represents a tree on the graph.

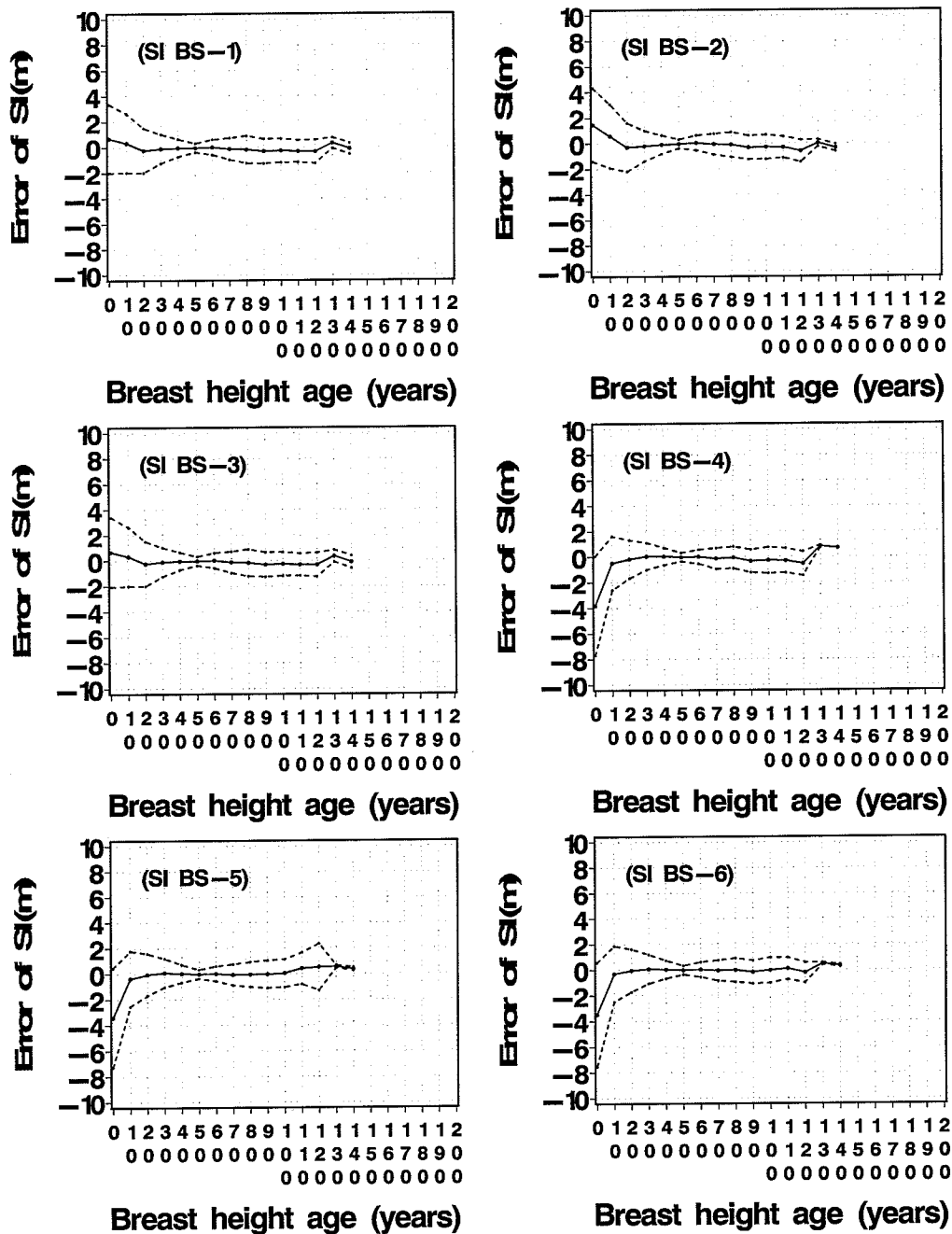


Figure 4.5. Average site index prediction biases by breast height age classes of black spruce from model [4.1] to [4.6]. The solid line in the middle represents the average height prediction errors and the two dashed lines ($\bar{e} \pm$ standard deviation) connected the standard deviations associated with the average errors.

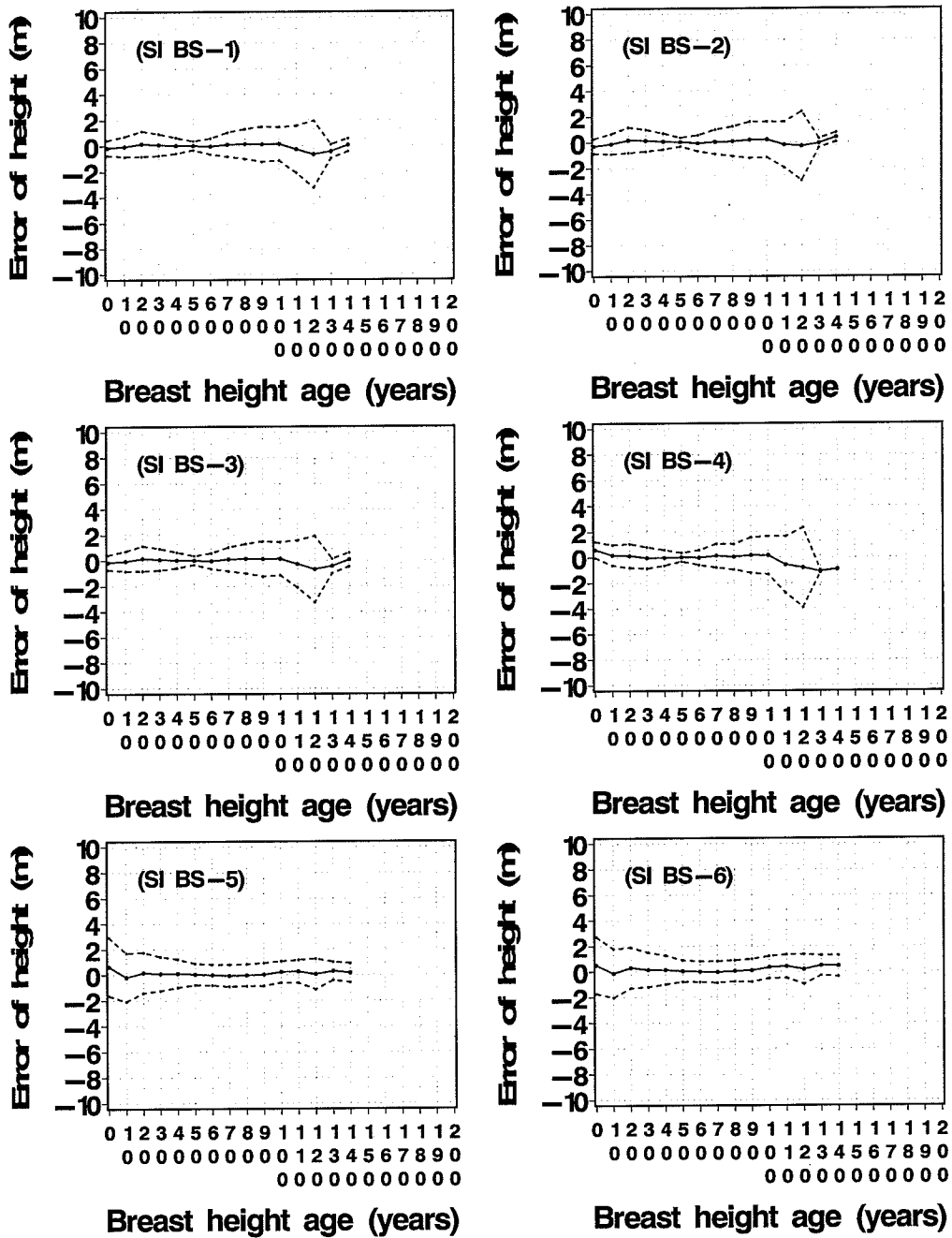


Figure 4.6. Average height prediction biases of model [4.1] to [4.6] by breast height age classes for black spruce. The solid line in the middle represents the average height prediction errors and the two dashed lines ($\bar{e} \pm$ standard deviation) connected the standard deviations associated with the average errors.

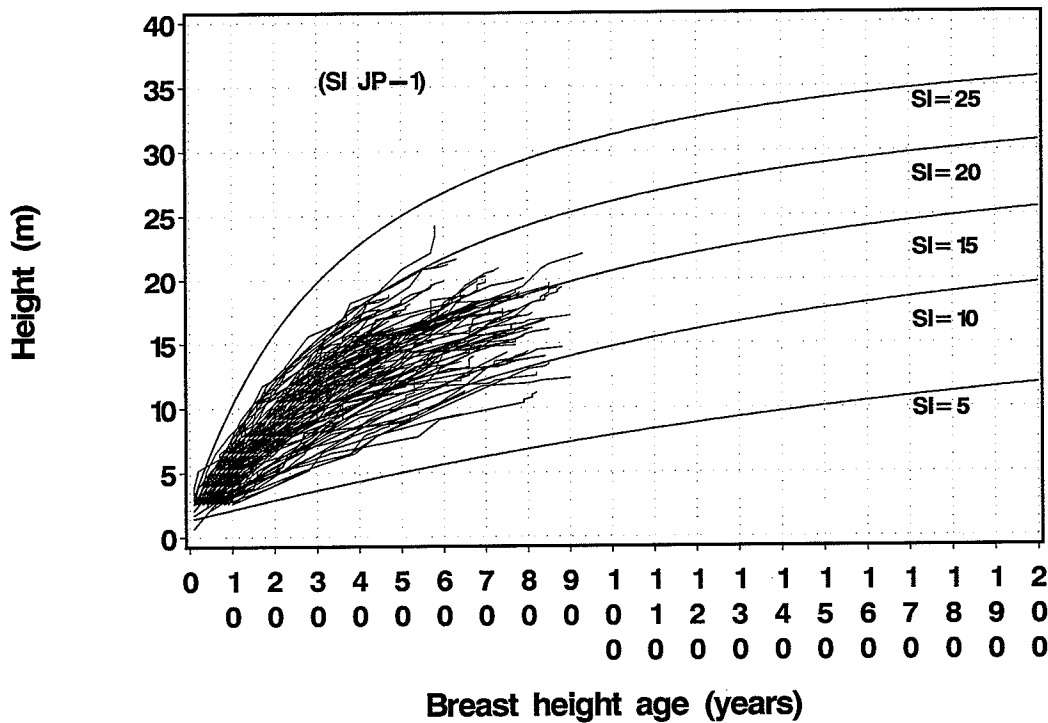
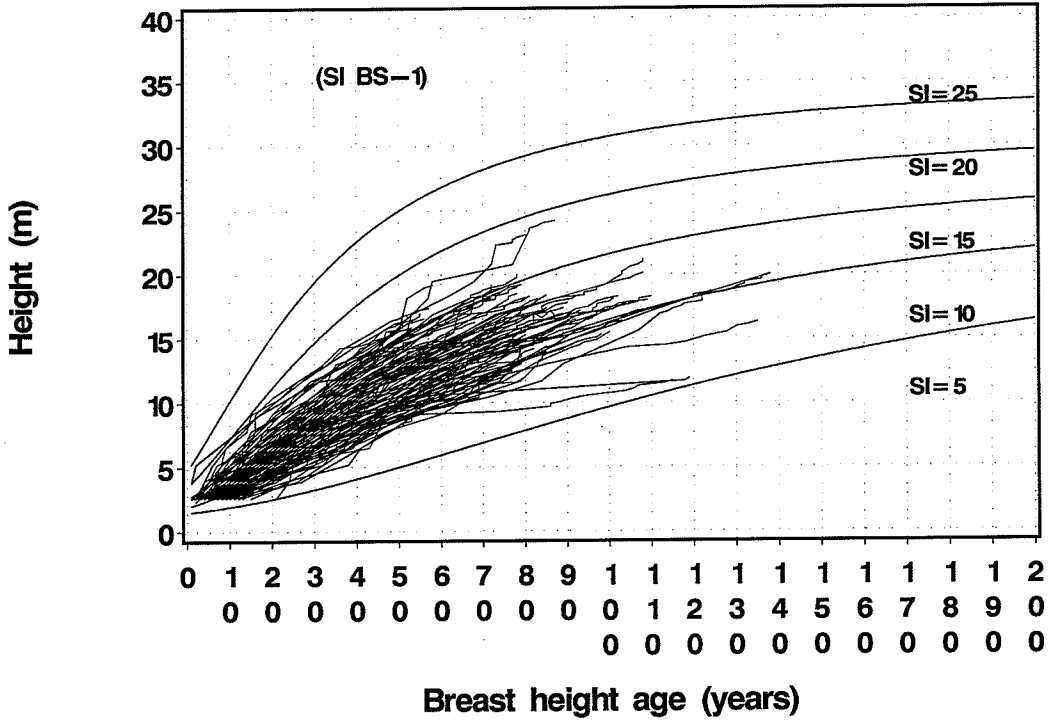


Figure 4.7. Site index curves generated from model [4.1], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves.

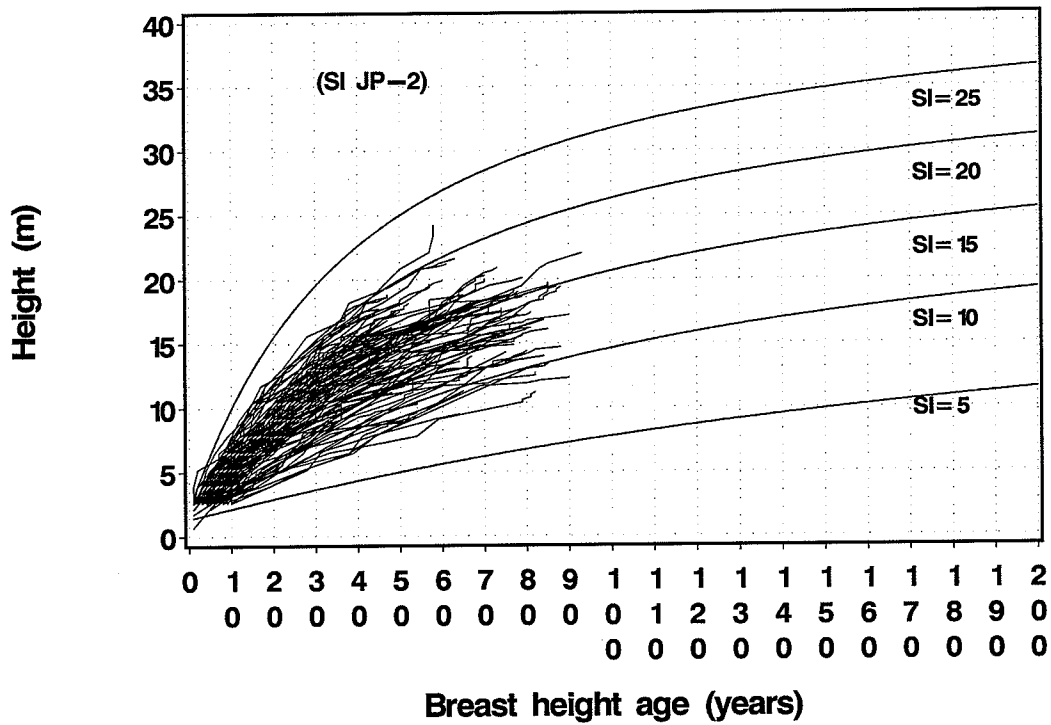
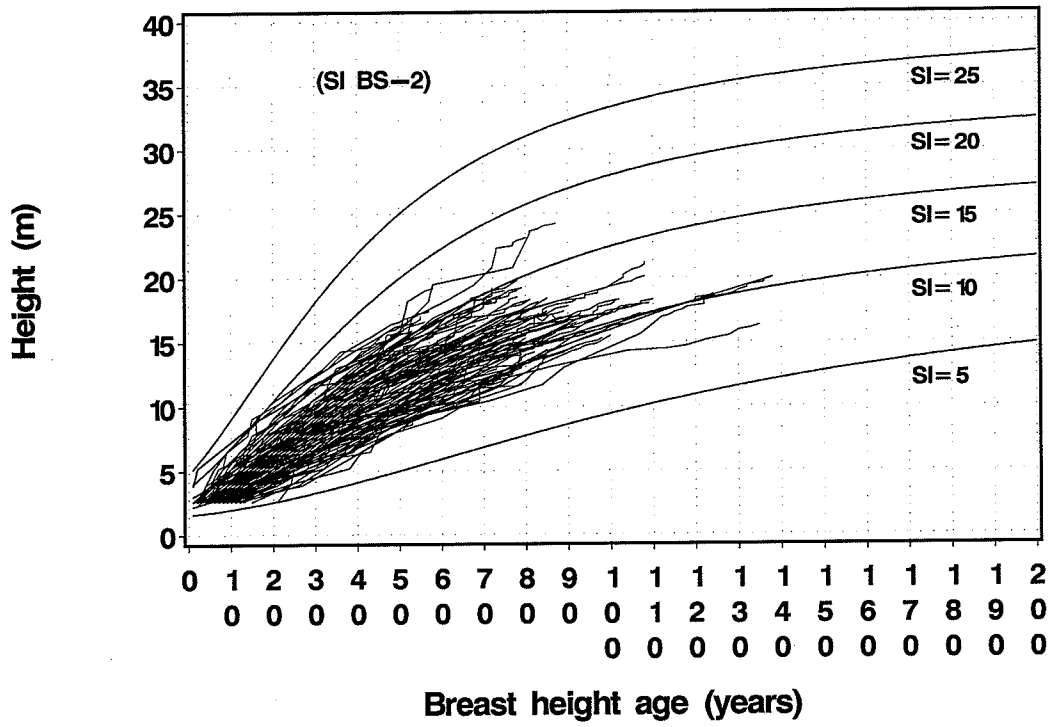


Figure 4.8. Site index curves generated from model [4.2], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves.

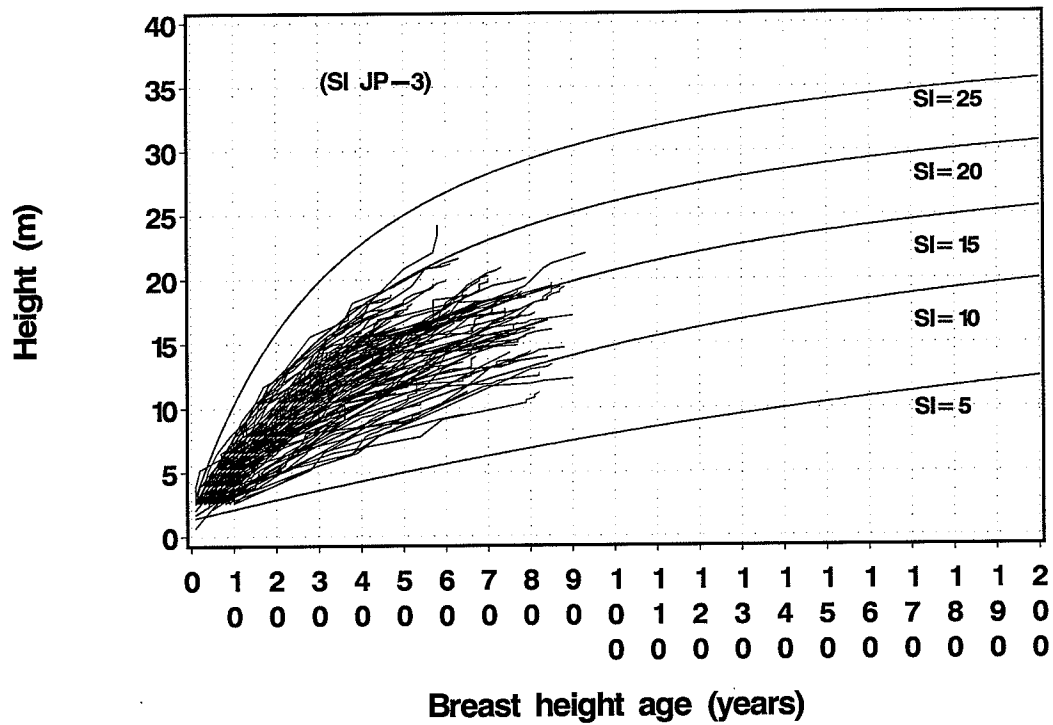
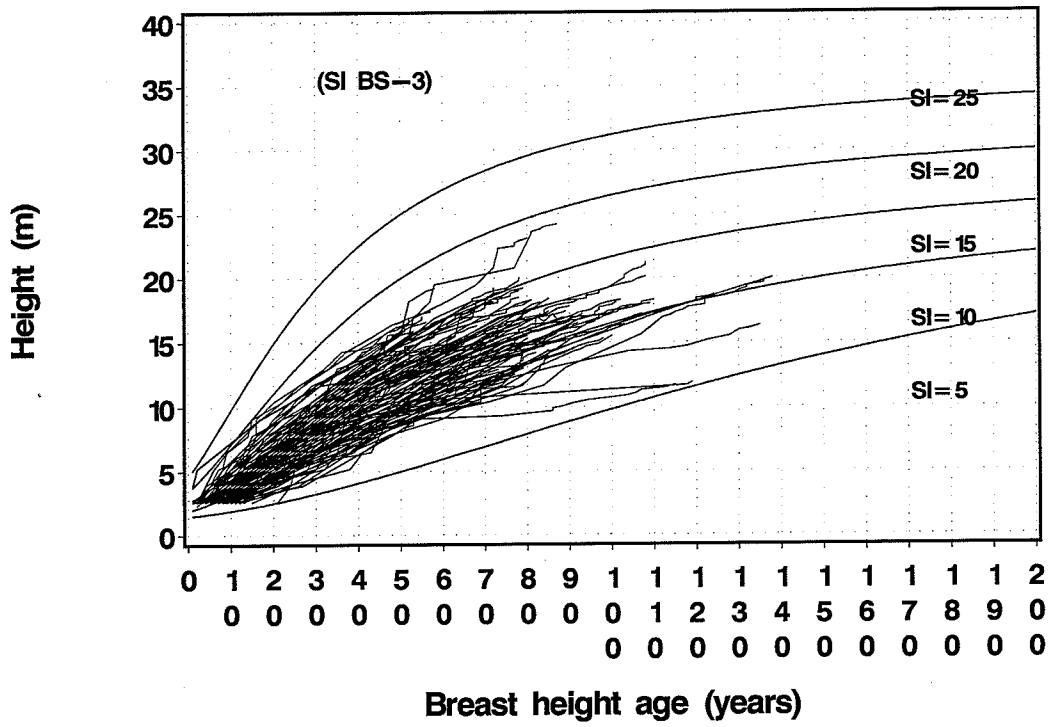


Figure 4.9. Site index curves generated from model [4.3], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves.

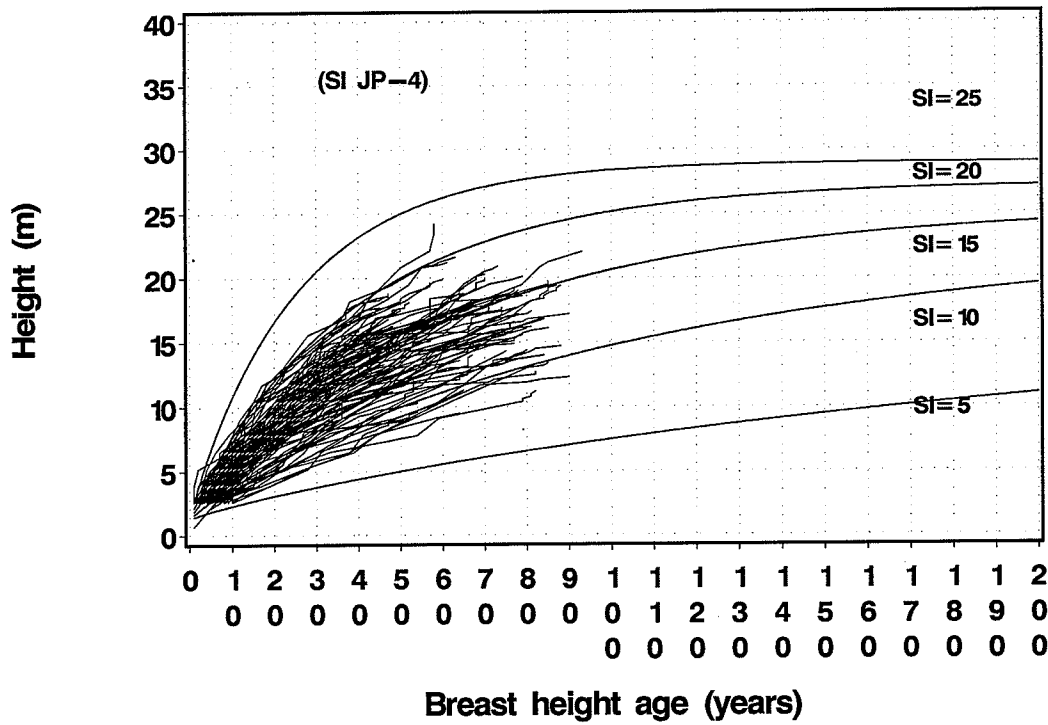
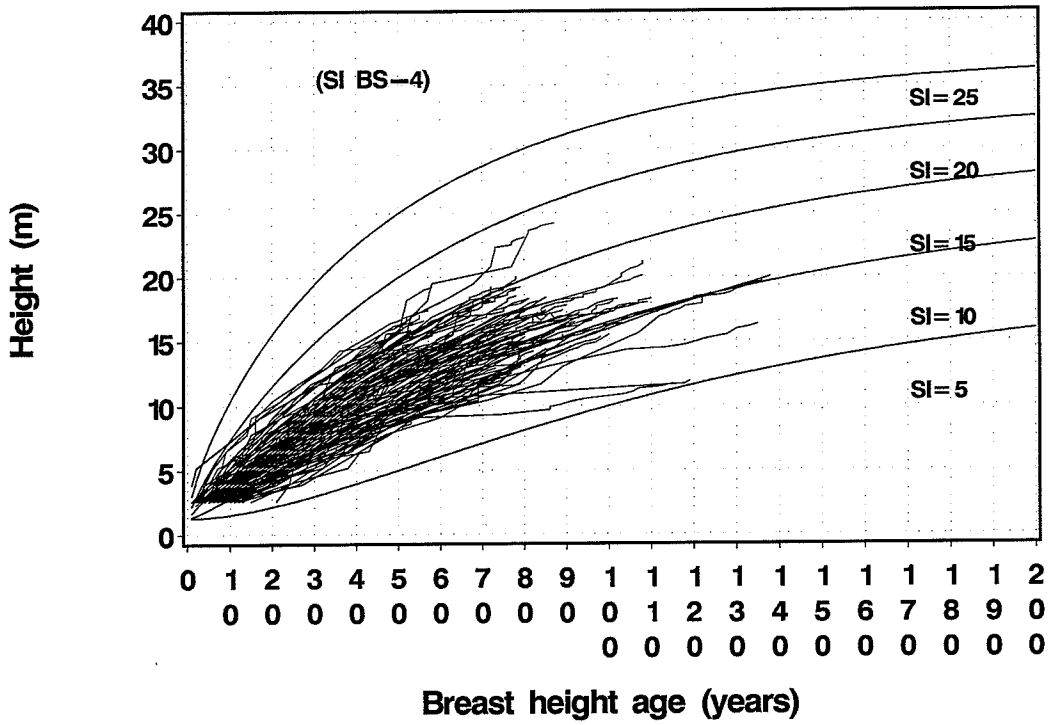


Figure 4.10. Site index curves generated from model [4.4], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves.

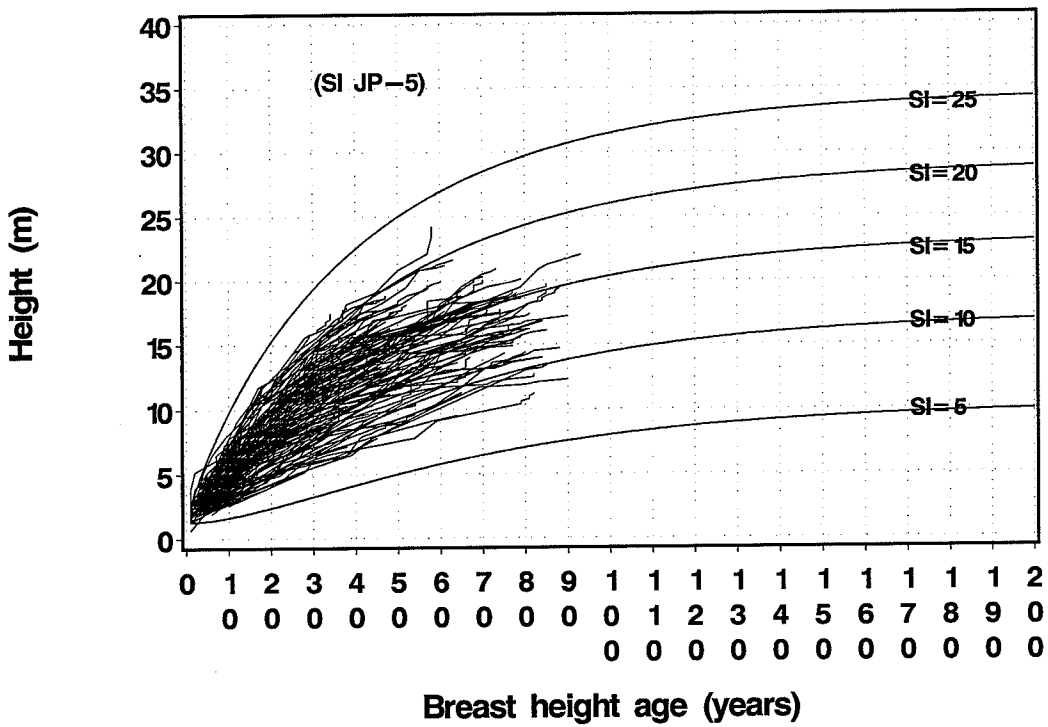
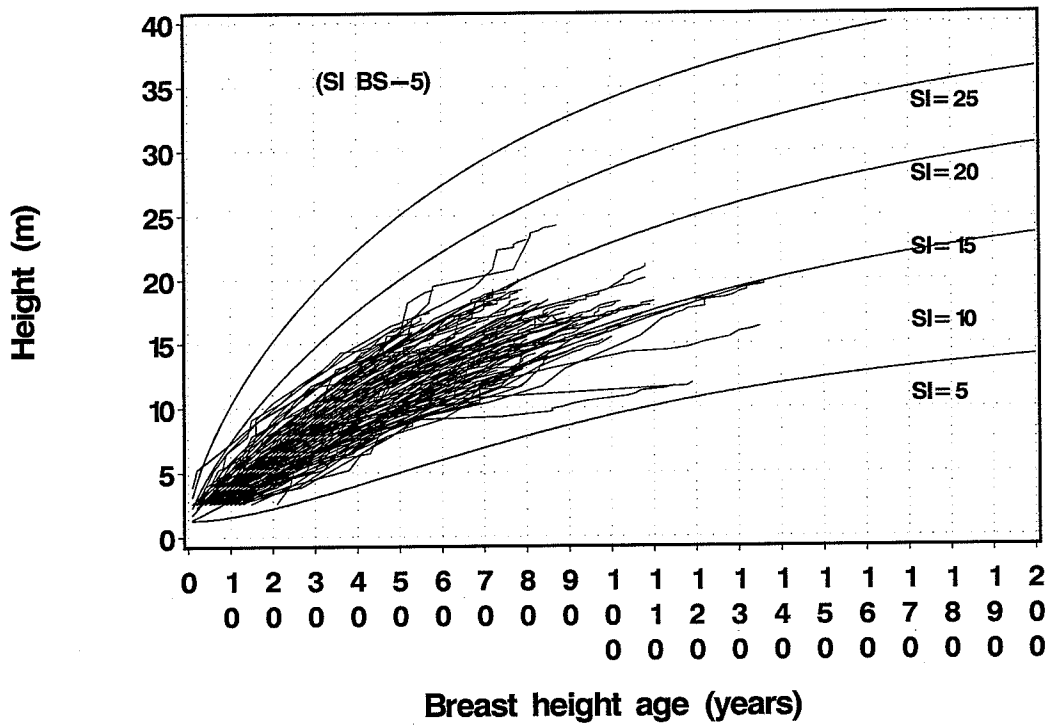


Figure 4.11. Site index curves generated from model [4.5], overlaid with the observed tree sectioning data of black black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves.

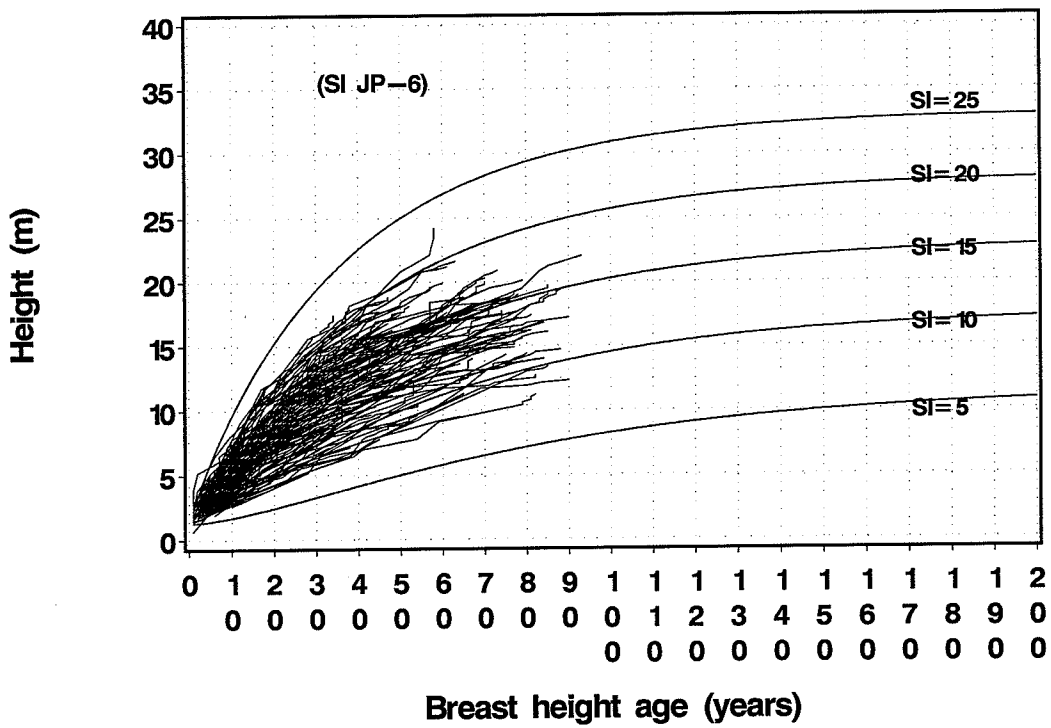
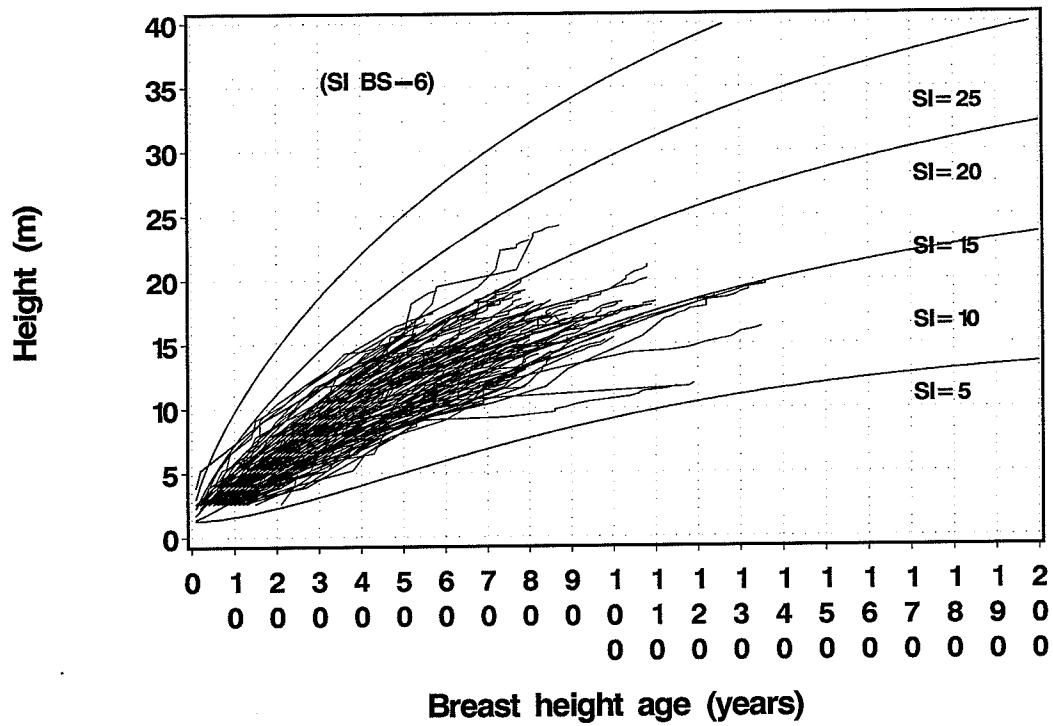


Figure 4.12. Site index curves generated from model [4.6], overlaid with the observed tree sectioning data of black spruce (top) and jack pine (bottom). Site index values at 50 years are given for each site index curves.

4.4.4 Model validation

Table 4.5 lists the results of height prediction statistics of model [4.1] to [4.6] based on the validation data for black spruce and jack pine. For black spruce, model [4.1] to [4.4] performed quite well. Model [4.2] and [4.4] had the lower mean bias ($\bar{\epsilon}$), -0.010 and 0.002, respectively, and the lower percent bias (Bias_per), -0.09% and 0.02%, respectively. But model [4.2] also had the highest MAD among models [4.1] to model [4.4]. Model [4.4] had the highest prediction coefficient of determination (R_p^2) value, 0.947, and the lowest root mean square error of prediction (RMSE_p), 1.053. Both model [4.4] and model [4.2] performed better for black spruce in the validation data.

For jack pine, models [4.1] to [4.4] also performed quite well. Model [4.1], [4.2] and [4.3] got higher prediction coefficient of determination (R_p^2), 0.963, and model [4.2] got the lowest root mean square error of prediction (RMSE_p), 0.952. Model [4.2] got the lowest mean bias ($\bar{\epsilon}$) and the lowest percent bias (Bias_per), 0.283 and 2.52% respectively. Model [4.2] also got the lowest MAD among model [4.1] to model [4.4]. Model [4.2] performed the best for jack pine in the validation data.

Compared to model [4.6], for black spruce, model [4.5] got the higher prediction coefficient of determination (R_p^2) and the lower root mean square error of prediction (RMSE_p), 0.917 and 1.314, respectively. However, model [4.6] got a lower mean bias ($\bar{\epsilon}$) and a lower percent bias (Bias_per), 0.004 and 0.03% respectively. For jack pine, model [4.5], compared to model [4.6], had the higher prediction coefficient of determination (R_p^2) and the lower root mean square error of prediction (RMSE_p), 0.939 and 1.246, respectively. The standard deviation of the prediction bias (S_e) of model [4.5] was 1.239.

Model [4.5] also got a lower mean bias (\bar{e}) and the lower percent bias (Bias_per), 0.123 and 1.07%, respectively.

Table 4.6 lists the hypothesis test results of site index prediction for model [4.1] to [4.3] and model [4.5] to [4.6]. For black spruce, there were no significant differences between observed and predicted site index values ($\alpha=0.05$) for the four tests. It appears all the five models are “good” models for black spruce site index predictions.

For jack pine, model [4.1] and [4.3] indicated significant differences between observed and predicted site index values ($\alpha=0.05$) for the separate t test and the simultaneous F test. However, these models passed the Novel test and the paired t test. Model [4.2], [4.5] and [4.6] passed the all four tests, suggesting that they are “good” models for jack pine site index predictions.

Table 4.5. Height prediction statistics of model [4.1] to [4.6] based on the validation data set for black spruce and jack pine.

SPP	Model	n	R_p^2	RMSE _p	\bar{e}	MAD	Bias_per	Se	t
BS	[4.1]	442	0.946	1.061	-0.041	0.707	-0.365	1.057	-0.810
JP	[4.1]	392	0.963	0.960	0.291	0.763	2.586	0.911	6.320
BS	[4.2]	442	0.945	1.063	-0.010	0.743	-0.085	1.061	-0.189
JP	[4.2]	392	0.963	0.952	0.283	0.760	2.519	0.906	6.188
BS	[4.3]	442	0.946	1.063	-0.042	0.706	-0.380	1.058	-0.843
JP	[4.3]	392	0.963	0.955	0.286	0.761	2.543	0.908	6.236
BS	[4.4]	442	0.947	1.053	0.002	0.707	0.016	1.049	0.036
JP	[4.4]	392	0.962	0.965	0.292	0.770	2.596	0.916	6.308
BS	[4.5]	7320	0.917	1.314	-0.049	0.833	-0.430	1.312	-3.172
JP	[4.5]	5981	0.939	1.246	0.123	0.929	1.073	1.239	7.705
BS	[4.6]	7320	0.915	1.328	0.004	0.841	0.034	1.328	0.251
JP	[4.6]	5981	0.938	1.255	0.163	0.930	1.419	1.244	10.152

Note: SPP – species; n – total number of observations (height-Bhage pairs) from sectioned trees; \bar{e} - average bias; MAD - absolute mean deviation; bias_per - bias %; S_e - standard deviation of the prediction errors; t - paired t test; RMSE_p - root mean square error of prediction; R_p^2 - prediction coefficient of determination .

Table 4.6. Results of four different tests based on the validation data for site index prediction of model [4.1] to [4.3] and [4.5] to [4.6] for black spruce and jack pine.

SPP	Test	Equation	[4.1]	[4.2]	[4.3]	[4.5]	[4.6]
BS	Novel test	[4.10]	0.904	1.545	1.018	2.033	2.581
BS	Paired t test	[4.7]	-0.596	-0.511	-0.576	0.371	0.673
BS	Seperate t test	[4.8]					
	$b_0 = 0$		0.34	-0.03	0.27	-0.24	-0.32
	$b_1 = 1$		-0.489	-0.082	-0.409	0.326	0.479
BS	Simultaneous F test	[4.9]	0.293	0.13	0.245	0.12	0.335
JP	Novel test	[4.10]	2.15	0.692	2.068	2.043	0.682
JP	Paired t test	[4.7]	1.796	1.204	1.774	1.994	1.193
JP	Seperate t test	[4.8]					
	$b_0 = 0$		2.520*	1.24	2.470*	1.93	1.28
	$b_1 = 1$		-2.124*	-1.01	-2.078*	-1.512	-1.046
JP	Simultaneous F test	[4.9]	4.126*	1.235	3.971*	3.245	1.261

Note: SPP – species; an asterisk indicates significant differences between observed and predicted site index values ($\alpha = 0.05$), which leads to model rejection.

4.4.5 Model re-fitting

After comprehensive considerations and comparisons, model [4.2] was selected as the 'best' model for both the black spruce and jack pine. To improve the model performance, both the fitting and validation data were pooled together to fit the final model.

The final fitted results of nonlinear least squares estimates of the parameters and fitted statistics are given in Table 4.7. For jack pine and black spruce, the fitted root mean squared errors (RMSE) were 0.835 and 0.962, respectively; the coefficients of determination (R^2) were 0.970 and 0.963, respectively; the average biases (\bar{e}) were 0.044 and 0.118, respectively; and the mean absolute deviations (MAD) were 0.621 and 0.727, respectively.

Table 4.7. Estimated parameters and fitted statistics of model [4.2] for black spruce and jack pine based on the pooled data.

SPP	b1	b2	b3	n	R^2	RMSE
BS	11.4236	-2.0759	17.6235	2053	0.970	0.835
JP	7.0847	-1.0712	1.2135	1760	0.963	0.962

SPP	\bar{e}	MAD	Bias_per	Se	t
BS	0.044	0.621	0.390	0.834	2.405
JP	0.118	0.727	1.069	0.954	5.204

Note: SPP – species; n – total number of observations (height-Bhage pairs) from sectioned trees; RMSE - root mean square error of prediction; R^2 - prediction coefficient of determination. \bar{e} - average bias; bias_per - bias %; MAD - absolute mean deviation; bias_per - bias %; S_e - standard deviation of the prediction errors;

The height and site index prediction accuracy and precision of model [4.2] for black spruce and jack pine by age classes are shown in Table 4.8 Table 4.9, Table 4.10 and Table 4.11. Based on the results provided in these tables, other statistics such as the standard errors of the mean prediction biases and the confidence intervals, can also be calculated.

The overall site index prediction bias and root mean squared error (RMSE) for black spruce were 0.044 and 1.0954, respectively (Table 4.8). The bias percent, 0.36%, was very low. For jack pine, the overall site index prediction bias and root mean squared error (RMSE) were 0.293 and 1.4768, respectively. The bias percent is 2.0 % (Table 4.9). On the other hand, the overall height prediction bias and root mean squared error (RMSE) for black spruce were 0.052 and 0.9089, respectively. The bias percent is 0.46% (Table 4.10). For jack pine, the overall height prediction bias and root mean squared error (RMSE) were 0.118 and 0.9615, respectively. The bias percent is 1.07 % (Table 4.11).

Although the overall height and site index prediction biases were very low, the prediction biases for different age classes were quite variable. The bias percent was higher at younger and older age classes, and lower at middle age classes.

Table 4.8. Accuracy and precision of site index predictions from model [4.2] for black spruce.

Bhage class (Years)	Ave_age (Years)	n		Observed	Predict	Bias	Bias_per	RMSE _p
0<Bhage<=2	1.6	18	Mean	13.5	12.94	0.561	4.16	1.7894
			s.d.	2.47	2.73	1.579		
2<Bhage<=5	4.3	48	Mean	12.66	12.46	0.207	1.63	1.08
			s.d.	2.08	2.2	1.036		
5<Bhage<=15	10.4	215	Mean	12.56	12.3	0.262	2.08	1.1601
			s.d.	2.28	2.22	1.124		
15<Bhage<=25	20.3	208	Mean	12.3	12.12	0.18	1.46	1.0405
			s.d.	2.19	2.1	1.02		
25<Bhage<=35	30.2	192	Mean	12.42	12.2	0.222	1.79	1.025
			s.d.	2.19	2.13	0.995		
35<Bhage<=45	40.8	233	Mean	12.41	12.31	0.102	0.82	1.0299
			s.d.	2.19	2.08	1.02		
45<Bhage<=55	50.5	290	Mean	12.72	12.51	0.207	1.63	1.124
			s.d.	2.45	2.16	1.101		
55<Bhage<=65	60.4	233	Mean	11.78	11.68	0.096	0.82	1.0674
			s.d.	1.98	1.96	1.058		
65<Bhage<=75	70.5	244	Mean	11.75	11.69	0.057	0.49	0.9372
			s.d.	2.02	2.04	0.932		
75<Bhage<=85	79.9	197	Mean	11.33	11.66	-0.331	-2.92	1.1226
			s.d.	2	2.11	1.066		
85<Bhage<=95	89.8	82	Mean	10.33	10.82	-0.484	-4.69	1.355
			s.d.	1.53	1.71	1.246		
95<Bhage<=105	99.9	62	Mean	10.49	11.35	-0.853	-8.13	1.4025
			s.d.	1.64	1.72	1.079		
105<Bhage<=115	109	21	Mean	10.13	10.54	-0.415	-4.1	1.6885
			s.d.	1.75	2.52	1.544		
115<Bhage<=125	121.2	18	Mean	9.23	9.79	-0.567	-6.14	1.7349
			s.d.	0.94	1.29	1.522		
125<Bhage<=135	130.3	15	Mean	9.61	10.01	-0.401	-4.18	0.5092
			s.d.	1.01	1.17	0.223		
135<Bhage<=145	136.8	5	Mean	10.28	10.83	-0.549	-5.34	0.9113
			s.d.	0.11	0.09	0.197		
Overall	50.1	2081	Mean	12.01	11.97	0.044	0.36	1.0954
			s.d.	2.24	2.13	1.094		

Note: Bhage class – Breast height age class; Ave.age – average age; SPP – species; n – total number of observations (height-Bhage pairs) from sectioned trees; s.d. – standard deviation; bias_per - bias %; RMSE_p - root mean square error of prediction.

Table 4.9. Accuracy and precision of site index predictions from model [4.2] for jack pine.

Bhage class (Years)	Ave_age (Years)	n		Observed	Predict	Bias	Bias per	RMSE _p
0<Bhage<=2	1.6	47	Mean	15.51	14.92	0.594	3.83	1.4429
			s.d.	2.52	2.69	1.277		
2<Bhage<=5	3.9	88	Mean	15.25	14.88	0.365	2.4	1.4561
			s.d.	2.54	2.58	1.392		
5<Bhage<=15	10.3	308	Mean	14.89	14.43	0.46	3.09	1.4596
			s.d.	2.51	2.61	1.38		
15<Bhage<=25	20.4	257	Mean	14.79	14.36	0.425	2.88	1.3657
			s.d.	2.57	2.65	1.292		
25<Bhage<=35	30.5	237	Mean	14.93	14.46	0.466	3.12	1.3834
			s.d.	2.52	2.65	1.296		
35<Bhage<=45	40.2	259	Mean	15.2	14.83	0.371	2.44	1.1528
			s.d.	2.59	2.57	1.086		
45<Bhage<=55	50.1	173	Mean	14.42	14.29	0.122	0.85	1.4545
			s.d.	2.91	2.89	1.441		
55<Bhage<=65	60.1	158	Mean	14.6	14.72	-0.113	-0.78	1.7576
			s.d.	2.67	3.08	1.743		
65<Bhage<=75	69.8	135	Mean	13.64	13.6	0.041	0.3	1.8813
			s.d.	1.97	2.43	1.867		
75<Bhage<=85	79.9	87	Mean	12.72	12.75	-0.021	-0.16	1.8595
			s.d.	2.47	2.39	1.838		
85<Bhage<=95	88.8	11	Mean	13.5	14.25	-0.749	-5.54	2.5553
			s.d.	2.42	2.29	2.146		
Overall	35.2	1760	Mean	14.68	14.38	0.293	2	1.4768
			s.d.	2.62	2.7	1.446		

Note: Bhage class – Breast height age class; Ave.age – average age; SPP – species; n – total number of observations (height-Bhage pairs) from sectioned trees; s.d. – standard deviation; bias_per - bias %; RMSE_p - root mean square error of prediction.

Table 4.10. Accuracy and precision of height predictions from model [4.2] for black spruce.

Bhage class (Years)	Ave_age (Years)	n		Observed	Predict	Bias	Bias per	RMSEp
0<Bhage<=2	1.6	18	Mean	2.8	2.81	-0.005	-0.19	0.7563
			s.d.	0.68	0.4	0.71		
2<Bhage<=5	4.3	48	Mean	2.8	3.09	-0.281	-10	0.5641
			s.d.	0.47	0.43	0.474		
5<Bhage<=15	10.4	215	Mean	4.11	4.2	-0.081	-1.96	0.8487
			s.d.	1.33	0.96	0.841		
15<Bhage<=25	20.3	208	Mean	6.27	6.11	0.154	2.45	0.9694
			s.d.	1.64	1.33	0.952		
25<Bhage<=35	30.2	192	Mean	8.53	8.34	0.193	2.27	0.8671
			s.d.	1.96	1.69	0.841		
35<Bhage<=45	40.8	233	Mean	10.66	10.59	0.071	0.66	0.6118
			s.d.	2.24	2.04	0.605		
45<Bhage<=55	50.5	290	Mean	12.86	12.81	0.046	0.36	0.3495
			s.d.	2.53	2.48	0.345		
55<Bhage<=65	60.4	233	Mean	13.47	13.54	-0.064	-0.47	0.6415
			s.d.	2.07	2.09	0.635		
65<Bhage<=75	70.5	244	Mean	15.07	15.05	0.021	0.14	0.9788
			s.d.	2.37	2.36	0.975		
75<Bhage<=85	79.9	197	Mean	15.93	15.78	0.15	0.94	1.1702
			s.d.	2.48	2.3	1.155		
85<Bhage<=95	89.8	82	Mean	15.89	15.71	0.185	1.16	1.4883
			s.d.	2.08	1.88	1.458		
95<Bhage<=105	99.9	62	Mean	17.12	16.89	0.227	1.33	1.4206
			s.d.	1.9	2.09	1.378		
105<Bhage<=115	109	21	Mean	16.82	17.22	-0.392	-2.33	2.1711
			s.d.	3.2	2.07	2.02		
115<Bhage<=125	121.2	18	Mean	16.78	16.96	-0.179	-1.06	2.731
			s.d.	2.28	1.2	2.559		
125<Bhage<=135	130.3	15	Mean	18.03	18.06	-0.024	-0.13	0.3544
			s.d.	1.59	1.28	0.327		
135<Bhage<=145	136.8	5	Mean	19.84	19.29	0.55	2.77	0.8924
			s.d.	0.17	0.16	0.142		
Overall	50.1	2081	Mean	11.32	11.27	0.052	0.46	0.9089
			s.d.	4.78	4.69	0.907		

Note: Bhage class – Breast height age class; Ave.age – average age; SPP – species; n – total number of observations (height-Bhage pairs) from sectioned trees; s.d. – standard deviation; bias_per - bias %; RMSE_p - root mean square error of prediction.

Table 4.11. Accuracy and precision of height predictions from model [4.2] for jack pine.

Bhage class (Years)	Ave_age (Years)	n		Observed	Predict	Bias	Bias_per	RMSE _p
0<Bhage<=2	1.6	47	Mean	2.75	2.43	0.324	11.76	0.7476
			s.d.	0.62	0.34	0.654		
2<Bhage<=5	3.9	88	Mean	3.29	3.32	-0.031	-0.93	0.6523
			s.d.	0.77	0.59	0.644		
5<Bhage<=15	10.3	308	Mean	5.48	5.52	-0.046	-0.85	0.9821
			s.d.	1.69	1.44	0.978		
15<Bhage<=25	20.4	257	Mean	8.58	8.55	0.021	0.25	1.0533
			s.d.	1.96	1.77	1.049		
25<Bhage<=35	30.5	237	Mean	11.47	11.19	0.282	2.45	1.0595
			s.d.	2.28	2.12	1.017		
35<Bhage<=45	40.2	259	Mean	13.67	13.47	0.2	1.46	0.6811
			s.d.	2.62	2.4	0.648		
45<Bhage<=55	50.1	173	Mean	14.45	14.41	0.045	0.31	0.3368
			s.d.	2.81	2.83	0.332		
55<Bhage<=65	60.1	158	Mean	16.23	16.05	0.171	1.05	0.8293
			s.d.	2.93	2.8	0.806		
65<Bhage<=75	69.8	135	Mean	16.45	16.21	0.239	1.45	1.3759
			s.d.	2.24	2.2	1.344		
75<Bhage<=85	79.9	87	Mean	16.3	16.2	0.095	0.58	1.5047
			s.d.	2.58	2.85	1.484		
85<Bhage<=95	88.8	11	Mean	18.91	17.92	0.985	5.21	2.2414
			s.d.	3.1	2.91	1.718		
Overall	35.2	1760	Mean	11.07	10.95	0.118	1.07	0.9615
			s.d.	4.99	4.87	0.954		

Note: Bhage class – Breast height age class; Ave.age – average age; SPP – species; n – total number of observations (height-Bhage pairs) from sectioned trees; s.d. – standard deviation; bias_per - bias %; RMSE_p - root mean square error of prediction.

4.5 Discussion and Conclusions

Over the past decades various approaches to site index estimation have been proposed, and many height and site index models have been developed. My study was to find the most appropriate models that are suitable for Manitoba.

The height-site index-age models [4.1]-[4.3] require site index as an independent variable, and therefore, the trees used to estimate the models must have site index values. As a result, each sampled trees must be older than 50 years. Based on the fitting statistics, validation statistics, validation diagnostic tests, the examinations of different kind of graphs and comparisons with other site index curves, we concluded that model [4.2] was best suited for application in Manitoba for black spruce and jack pine.

The two difference equations (models [4.5] and [4.6]) take the form of $H_2 = f(H_1, T_1, T_2)$, and are not expressed as functions of site index. H_1 and T_1 are height and age at time one, and H_2 and T_2 are height and age at time two. They can be formulated from height-site index-age models (e.g., by substituting SI and 50 by H_1 and T_1 , respectively). The estimation of the non-reciprocal difference equations (models, [4.5] and [4.6]) does not require site index as an independent variable. Sampled trees that do not have site index values (e.g., breast height age less than 50 years) can also be used for fitting the models. However, these two models did not performe as good as model [4.2], [4.1] and [4.3].

The comparison of the fitting statistics of the models, residual plots and site index curve shapes showed that the performance of model [4.1], [4.2], [4.3] and [4.4] were very good, with model [4.2] being the best. Model [4.5] also performed better than model [4.6].