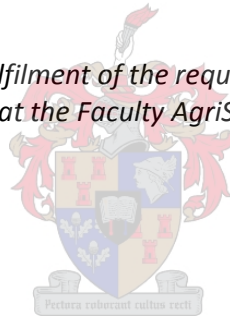


Development of potential height growth and diameter increment models for the parameterisation of an individual tree growth model for *Pinus elliottii* plantations in South Africa

by

Gerard Eckard Lindner

*Thesis presented in fulfilment of the requirements for the degree of
Master of Science in Forestry at the Faculty AgriSciences at Stellenbosch University*



Supervisor: Prof Thomas Seifert
Co-supervisor: Prof Hans Pretzsch

Department of Forest and Wood Science
Faculty of AgriSciences

April 2014

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 26 February 2014

Acknowledgements

First of all I would like to thank Prof. Thomas Seifert for his support, mentorship, guidance and valuable input into this thesis. The balance between support and research freedom allowed great personal self-development and greatly improved my skills and competencies in Forest Science research and Growth and Yield Science.

I would like to thank Prof. Hans Pretzsch for initialising the project and his valuable guidance and support, especially during my time in Germany. In this regard I would also like to thank Mr. Ralf Moshammer who managed my time in Germany, gave me practical guidance and offered his support throughout.

Particular thanks also go to postgraduate colleagues who helped me along the way and Dr. Stephan Seifert for his help with R programming applications and some valuable code, which was of great help to the study.

I would like to thank Komatiland Forests (KLF), York Timbers, Mountain to Ocean Forestry (now Cape Pine) and all of the particular people involved for their willingness to provide data. Further thanks go to the Mensuration and Modelling Research Consortium (MMRC) – particularly Mr. Heyns Kotze and Mr. Trevor Morley, for pointing me in the right direction for obtaining data for the study.

Personal thanks go to my brothers, Brian and Berndt Lindner for their constant support and discussions; Brian in particular for editing, programming and conceptual discussions. Thanks also to Esmarie Scholtz for her love and support and also for insights and support into some technical issues as well. I am grateful to my friends and family, particularly my parents, who patiently supported and encouraged me throughout the study. My appreciation is also extended to my colleagues at the Department of Forestry and Wood Science who made the experience a most enriching one.

This study was partly financed by the National Research Foundation (NRF) within the scope of the Green Landscapes Project. In particular I am grateful to the Department of Agriculture Forestry and Fisheries (DAFF) for supporting me financially from undergraduate through to postgraduate level.

Abstract

Individual tree models, as opposed to stand models, have the potential to greatly improve sensitivity of forest growth models to changing conditions such as silvicultural amendments, irregular stand structures, etc. It was the purpose of this study to extend two sub-components of a European individual tree growth model to introduce individual tree growth modelling concepts in South Africa using *Pinus elliottii* as a study species. Two main objectives were established: Modelling the potential height of stands across different site qualities and modelling diameter increment using a potential modifier approach with a combination of competition indices that change in importance according to the edaphic conditions of the site.

Potential height modelling used three steps in order to achieve this objective. The first was to compare site index models based on different model fitting techniques, namely nonlinear least squares, generalised nonlinear least squares and nonlinear mixed effects models. The nonlinear mixed effects model proved to be superior in terms of achieving the principles of regression assumptions and model fit for the data range observed. The second step was to fit potential height using nonlinear quantile regression on observed spacing trial height measurements. This proved to be a robust technique able to capture potentials according to the defined Chapman-Richards model structure. The final step was to use the predicted site index as a site classification variable in order to predict potential height. While some small deviation occurred, potential height seems to be well correlated to site index and validation on selected sites suggested that site index can be used to model potential height until a more sophisticated site classification model is used for future improvement of the model.

Diameter increment modelling followed six major steps in order to apply the full parameterisation methodology of an age-independent diameter increment model dependent on tree diameter and competition. Diameter increment potentials were fit using site index as a predictor of the potential height curves. Multiple competition indices were tested on two sites to obtain a combination of two indices, which can capture overtopping and local crowding effects. Principle components analysis and variance inflation factors calculation were applied to test for collinearity between indices. Suitable combinations were tested resulting in a combination of the KKL and Local Basal Area competition indices. Changing importance of the two indices were observed on the two sites tested indicating a shift in the mode of competition according to a water gradient.

These were combined in a deterministic potential modifier model, which mimicked competitive stages over age; however the validation showed a skewed distribution, which was not sensitive to stand density gradients. A stochastic model was constructed to model variance from observed residual plots using linear quantile regression to determine bounds for a truncated normal distribution which generates random deviates for a predicted increment. The stochastic element significantly improved the performance and sensitivity of the model, however the model was still not sensitive enough at very high and very low spacing densities. All in all two key models for an adaptation of an individual tree growth simulator to South African conditions were successfully demonstrated. The two main objectives were achieved; however some indicated improvements could be made, especially for the competition indices where the sensitivity of competition to changing resource limitation according to site and temporal scales needs to be further investigated.

Furthermore, the full set of models for simulating individual tree growth still needs to be applied. Overall, as a methodological approach, the study outlined problems and future improvements, introduced new concepts and can serve as a guideline for future parameterisation of an individual tree growth model.

Opsomming

In vergelyking met vakgroeimodelle, het individuele-boomgroeimodelle die potensiaal om die sensitiwiteit van plantasiegroeimodelle vir veranderende omstandighede soos aanpassings in boskultuur, onreëlmatige vakstrukture, ensovoorts, drasties te verbeter. Die doel van hierdie studie was om twee subkomponente van 'n Europese individuele-boomgroeimodel uit te brei om sodoende individuele-boomgroeimodelleringskonsepte in Suid-Afrika bekend te stel. *Pinus elliottii* is gebruik as studiespesie. Twee hoofdoelstellings is bepaal. Eerstens, die modellering van hoogtegroeipotensiaal van opstande oor verskeie vlakke van groeiplek kwaliteit. Tweedens, die modellering van deursnee-aanwas deur gebruik te maak van 'n potensiaal matigingsbenadering "potential modifier approach" met 'n kombinasie van kompetisie-indekse waarvan die belangrikheid verander volgens die edafiese toestande van die groeiplek.

Die hoogtepotensiaalmodellering bestaan uit drie stappe. Tydens die eerste stap word groeiplek bonniteitsmodelle vergelyk op grond van verskillende modelpassingstegnieke, naamlik nie-lineêre minimum kwadrate, algemene nie-lineêre minimum kwadrate en nie-lineêre gemengde effek modelle. Laasgenoemde het die beste gevaar in terme van die beginsels van regressiemodelle asook die mate waarin die model die waargeneemde data pas. Tweedens is hoogtegroeipotensiaal gemodelleer deur nie-lineêre kwantielregressie op waargeneemde hoogtes van spasiëringseksperimente toe te pas. Die metode is robuust en in staat om potensiale volgens die gedefinieerde Chapman Richards modelstruktuur vas te vang. Laastens is die voorspelde bonniteits indeks as 'n groeiplek klassifisering veranderlike gebruik om sodoende die hoogtegroeipotensiaal te voorspel. Alhoewel klein afwykings voorgekom het, blyk hoogtegroeipotensiaal goed gekorreleer te wees met bonniteits indeks. Uit validasie op geselekteerde groeiplekke blyk dit dat bonniteits indeks gebruik kan word om hoogtegroeipotensiaal te modelleer totdat 'n meer gesofistikeerde groeiplek klassifikasie-model beskikbaar is wat die model verder sal kan verbeter.

Die volledige parametriseringsmetodiek van 'n ouderdoms-onafhanklike deursnee-aanwas model wat afhanklik is van boomdeursnee en kompetisie bestaan uit ses hoofprosesse. Nie-lineêre kwantielregressie is gebruik om deursnee-aanwaspotensiale te pas vir verskeie groeiplekke. Dié is gekombineer met 'n bonniteits indeks om 'n nuwe model te vorm waarmee hoogtegroeipotensiaal kurwes voorspel kon word. Daar is met veelvuldige kompetisie-indekse op twee groeiplekke geëksperimenteer om 'n kombinasie van slegs twee indekse te vind wat die effekte van oorskaduwing en plaaslike verdringing kan vasvang, te vind. Hoofkomponent analise "Principle components analysis" en variansie inflasie faktore berekening "variance inflation factors calculation" is gebruik om vir kollineariteit tussen die indekse te toets. Gepaste indeksoombinasies is getoets. 'n Kombinasie van die KKL en plaaslike basale oppervlakte "Local Basal Area" kompetisie-indekse het die beste resultate gelewer. Die twee indekse is as volg geselekteer. Veranderinge in die belangrikheid van elk van die indekse is waargeneem op die twee toetspersele. Dit dui op 'n verskuiwing in die modus van kompetisie afhangend van 'n watergradiënt.

Die twee indekse is gekombineer in 'n deterministiese potensiaal matigings model wat die kompeterende stadiums oor ouderdom naboots. Validasie het egter 'n skewe verdeling wat nie sensitief vir opstandsdigtheidsgradiënte is nie, gewys. 'n Stogastiese model is ontwikkel om variansie in die residuele grafieke te modelleer. Lineêre kwantielregressie is gebruik om grense vir 'n

afgestompte normaalverdeling wat ewekansige afwykings vir 'n voorspelde aanwas te bepaal. Die stogastiese element het die prestasie van die deterministiese model merkbaar verbeter. Selfs met die stogastiese element, is die model egter steeds nie sensitief genoeg vir baie hoë en baie lae opstandsdigthede nie.

Ter opsomming is twee modelle vir 'n aanpassing van 'n individuele-boomgroeisimuleerder vir Suid-Afrikaanse toestande suksesvol gedemonstreer. Die twee hoofdoelstellings is bereik. Daar is egter steeds 'n paar aangeduide verbeterings wat aangebring kan word. Die sensitiwiteit van die kompetisie-indekse op hulpbronbeperkings wat verander op grond van die ruimtelike en temporale skale moet veral verder bestudeer word. Verder moet die volle stel modelle wat benodig word om individuele-boomgroei te modelleer nog toegepas word. As 'n metodologiese benadering, het die studie probleme uitgewys en toekomstige verbeterings aangedui, nuwe konsepte bekendgestel en kan dus dien as 'n riglyn vir toekomstige parametrisering van individuele-boomgroei modelle.

Table of Contents

Chapter 1: Introduction and background.....	1-1
1.1 Problem statement.....	1-1
1.2 Background and literature review.....	1-2
1.2.1 <i>Pinus elliottii</i> plantations in South Africa.....	1-2
1.2.2 Forest growth modelling.....	1-2
1.2.3 Competition Indices.....	1-4
1.2.4 Competition symmetry.....	1-4
1.2.5 Vulnerability of forestry to climate change.....	1-7
1.2.6 Potential modifier approach.....	1-7
1.2.7 Silva.....	1-12
1.3 Objectives.....	1-13
1.3.1 Potential height model:.....	1-14
1.3.2 Diameter increment model:.....	1-14
1.4 The use of R in the study.....	1-14
Chapter 2: Dataset description.....	2-16
2.1 The correlated curve trend (CCT) data.....	2-16
2.2 Triple S-CCT.....	2-20
2.3 Nelder spacing trial.....	2-20
2.4 Permanent sample plots (PSP) data.....	2-23
Chapter 3: Site index and potential height modelling.....	3-25
3.1 Introduction.....	3-25
3.1.1 Site index.....	3-25
3.1.2 Potential height growth.....	3-25
3.1.3 Nonlinear regression assumptions.....	3-25
3.1.4 Chapter objectives.....	3-26
3.2 Step 1: Site index modelling.....	3-26
3.2.1 Identification of dominant trees.....	3-27
3.2.2 Model selection.....	3-28
3.2.3 Dataset and trial Age.....	3-29
3.2.4 Starting values:.....	3-29

3.2.5 Nonlinear least squares	3-30
3.2.6 Generalised nonlinear least squares	3-32
3.2.7 Nonlinear mixed effects model	3-32
3.2.8 Model comparisons	3-35
3.2.9 NLME final results (model selection).....	3-37
3.3 Step 2: Potential height modelling.....	41
3.3.1 Description of the NLRQ procedure	41
3.4 Step 3: Prediction of potential height from Site Index.....	3-45
3.4.1 Effect of stand density	3-45
3.4.2 Relationship between potential height and dominant height.....	3-46
3.4.3 Predictive equation.....	3-49
3.4.4 Validation	3-50
3.5 Chapter conclusion	3-55
Chapter 4: Modelling diameter increment in response to resource limitations and site classification	4-56
4.1 Introduction	4-56
4.1.1 Dataset	4-57
4.1.2 Chapter outline.....	4-57
4.2 Step 1: Site classification according water availability and site index	4-59
4.3 Step 2: Determine the potential based on site conditions.....	4-60
4.3.1 Comparison of sites	4-60
4.4 Step 3: Fit various competition models.....	4-63
4.4.1 Distance dependent competition indices	4-63
4.4.2 Edge effects.....	4-66
4.4.3 Competition search radius (influence zone)	4-66
4.4.4 Performance of competition indices	4-69
4.5 Step 4: Selecting competition indices	4-71
4.5.1 Variable selection	4-73
4.5.2 Principle component analysis.....	4-75
4.5.3 Relative importance.....	4-79
4.6 Step 5: Use CI's in a deterministic potential modifier equation.....	4-82
4.6.1 Objective 1: Incorporating the LBA competition index	4-83

4.6.2 Objective 2: Incorporating a water index	4-84
4.6.3 Model behaviour	4-85
4.6.4 Model validation.....	4-87
4.7 Step 6: Create a stochastic model incorporating natural variability	4-88
4.7.1 Incorporating natural variation	4-88
4.8 Chapter conclusion	4-99
Chapter 5: Conclusion and recommendations	5-102
5.1 Potential height modelling.....	5-102
5.2 Diameter increment modelling	5-103
5.3 Additions for model completion	5-103
5.4 Overall thoughts.....	5-104
Chapter 6: References.....	6-105
Chapter 7: Appendices	7-111
Appendix A: NLME random effects anova table.....	7-111
Appendix B: Voronoi polygon increment relationship for the Nelder and Tweefontein sites respectively	7-113
Appendix C: Observed and predicted DBH for the deterministic model.....	7-115
Appendix D: Simulation steps for the deterministic model	7-116
Appendix E: Simulation steps for the stochastic model.....	7-120
Appendix F: Select examples of R- Code	7-125

Table of Figures

Figure 1-1: Hypothesis on the relationship of plant size on growth rate in a given stand with respect to the size symmetry of competition. Stands are usually in a continuum between these extremes where light limitation result in size asymmetric relationships and size symmetric relationships where underground resource limitations are prevalent. Symmetric competition occurs when competition is not related to tree size (figure from Pretzsch and Biber, 2010).	1-5
Figure 1-2: Figure illustrating two different types of competition indices theoretically more suited to overtopping and local crowding respectively.	1-6
Figure 1-3: Illustration of the site factor model used in SILVA, used as the precursor to modelling potential height growth (Pretzsch et al. 2002).	1-9
Figure 1-4: Effect of competition on the modifier values of dbh and height. From Ek and Dukek 1980.	1-11
Figure 1-5: Silva flow diagram and the potential height-age relation in SILVA predicted from site quality. This curve is used for the simulation initialisation	1-13
Figure 1-6: The diameter increment model function in SILVA determined by diameter (DBH). Predicted increment is then based on a modifier which takes competition into account.	1-13
Figure 2-1: Height measurements on the four CCT spacing trials showing a clear stratification between sites	2-18
Figure 2-2: Height measurements on the four CCT spacing trial indicating different growth patterns and growth trends for different planting densities	2-18

Figure 2-3: 3D plot of height growth trends in the CCT spacing trials with a linear average response curve indicating little change between maximum height growth between different stand densities	2-19
Figure 2-4: 3D plot of diameter growth trends in the CCT spacing trials with a linear average response curve indicating significant change between height growth between different stand densities	2-20
Figure 2-5: Nelder trial designs based on different spacing geometries. Nelder (1962).	2-21
Figure 3-1: Flowchart of site index modelling section of this chapter (Step2).....	3-27
Figure 3-2: Dominant height modelling example for the Mac Mac 2965spha plot with red dots representing the dominant trees, the red and black dashed lines representing the dominant and mean height curves respectively	3-28
Figure 3-3: Example of an NLS fit on the Weza 2965 plot	3-30
Figure 3-4: Regression diagnostics for the Weza 2965 plot	3-31
Figure 3-5: Scatter plot matrix of the relationship between three parameters of the Chapman Richards equation for a model fit, showing significant correlation between the parameters.....	3-33
Figure 3-6: Residual plots of an example dataset using different modelling techniques	3-36
Figure 3-7: Fitted dominant height curves for the CCT trials	40
Figure 3-8: Fitted potential height curves for the CCT dataset for different stems per hectare.	42

Figure 3-9: Potential height of the highest CCT trial value (Mac, spha = 124) plotted over PSP data of a wide range of sites, the blue line is the potential of the CCT plot, the red line represents the mean height of the PSP dataset	43
Figure 3-10: Site Index values on different stand densities for the CCT dataset.....	3-46
Figure 3-11: Example of potential and dominant height-age curves for the Weza trial	3-47
Figure 3-12: Potential height plotted over dominant height for all of the the CCT dataset plots.....	3-48
Figure 3-13: Final parameterised model of potential height-age using SI as a predictor	3-49
Figure 3-14: Deviation from the observed potential height-age compared to the potential height predicted equation	3-51
Figure 3-15: Median deviation in percentage of the observed and predicted potential height curves	3-52
Figure 3-16: Potential height curves plotted over PSP data of different classes. The red lines represent the predicted potential height for the upper bound of the SI classes presented above. The top left image presents all of the data with 15, 20, 25 and 30 SI predicted potentials	3-53
Figure 3-17: Fitted predicted potential height curves fitted on selected independent PSP data	3-54
Figure 4-1: DBH Increment - DBH scatterplot on the Mac Mac CCT trial, showing decreasing linear gradient over age	4-56
Figure 4-2: Flowchart of the chapter outline showing working steps of the methodological approach used in the study	4-58

Figure 4-3: Fitted potential increment curves for the four sites considered for increment potential estimation	4-61
Figure 4-4: Potential increment over DBH curves parameterised from the CCT trial data; the different curves represent different SI values	4-63
Figure 4-5: Voronoi polygons calculated for the Nelder spacing trial as an example .	4-66
Figure 4-6: Negative Correlation of the LBA competition index with diameter increment at different measurement years (ages) using different competition search radii, the correlation was multiplied by -1 for illustrative purposes.....	4-67
Figure 4-7: Linear regression of the optimum search radius over dominant height...	4-69
Figure 4-8: Relationship of diameter increment and the competition indices fitted in the two spacing trials used for this study.....	4-70
Figure 4-9: Pairwise scatterplot matrix of the various competition indices against each other in the Nelder Trial	4-72
Figure 4-10: Pairwise scatterplot matrix of the various competition indices against each other in the Tweefontein Trial.....	4-73
Figure 4-11: PCA biplot of the variable plotted in the transformed space of the first two principle components in the Nelder trial.....	4-77
Figure 4-12: PCA biplot of the variable plotted in the transformed space of the first two principle components in the Tweefontein trial	4-77
Figure 4-13: Relative importance graphs using different methods for the Tweefontein spacing trial	4-81

Figure 4-14: Relative importance graphs using different methods for the Nelder spacing trial showing improved KKL importance as compared to the Tweefontein site	4-82
Figure 4-15: Illustration of the potential modifier approach on the two spacing trials, with the top right hand diagram representing the change of the modifier over the size of competition	4-86
Figure 4-16: Comparison between the observed (red bars) and the simulated diameter (DBH) represented as a distribution of 2.5cm diameter classes. The mean DBH for the site is represented in the notched boxplot. This showed a skewed prediction grouped around a narrower DBH band	4-88
Figure 4-17: Residual plot of the predicted values of the spacing trials	4-89
Figure 4-18: 5% and 95% linear quantile regression lines which represents the upper and lower bounds for the truncated normal distribution	4-90
Figure 4-19: Predicted residual generated from a random deviate between two bounds (Figure 18) based on a normal distribution.....	4-91
Figure 4-20: Sequential process of generating random deviation, from plotting residuals (top left), predicting bounds (top right), generating random deviation (bottom left) superimposed on the residual plot (bottom right)	4-92
Figure 4-21: Observed and predicted scatterplots for the Tweefontein SS-CCT spacing trial, with black points representing the observed and green representing the predicted values. The left hand image represents the scatter of the average model – with the linear quantile in the top right, the bottom left and right images represent the scatter of the model with added modelled random variance.....	4-93

Figure 4-22: Observed and predicted scatterplots for the Lottering Nelder spacing trial, with black points representing the observed and green representing the predicted values. The top right image shows the deterministic model, the bottom left and right images represent the scatter of the model with added modelled random variance..... 4-94

Figure 4-23: Residual scatter of the added random variance model for the parameterised sites 4-95

Figure 4-24: Residual scatter overlain on the observed increment scatter of the added random variance model for the parameterised sites. The black points represent the observed and the green points represent the predicted values from the stochastic model. 4-96

Figure 4-25: Observed (red bars) vs. simulated (blue bars) for the Nelder trial with random deviation added at each point. The model shows a clear overall overprediction of DBH, however with an improved distribution compared to the model without deviation (Figure 4-16)..... 4-97

Figure 4-26: Deviation of predicted average diameter increment for a more realistic of planted stand densities for a plantation industry setup in the Nelder trial..... 4-99

Tables

Table 2-1: CCT unthinned series trial summary	2-16
Table 2-2: Thinning in advance of coemption for weed control in the CCT trial data	2-17
Table 2-3: Unthinned SSS-CCT spacing trial desing	2-20
Table 2-4: Measurement intervals for the Lottering Nelder trial	2-22
Table 2-5: Nelder spacing trial plot numbers (rings) with their corresponding stems per hectare, based on the radii and arc distances.....	2-23
Table 3-1: Random effect tests for the nonlinear mixed effects model. The column “parameter chosen” indicates which parameters used as an additional random effect in the model.	3-34
Table 3-2: Count of best fit random effects models	3-34
Table 3-3: Anova and RMSE comparisons of different modelling fitting methods on the Weza dataset.....	3-37
Table 3-4: Fitted SI-age models for the CCT trials.....	39
Table 3-5: Potential height fits by plot for the four spacing trials	44
Table 3-6: Potential height - dominant height gradient.....	3-48
Table 3-7: Final model parameterised on the pooled CCT trial datasets from Equation 3-4	3-50
Table 4-1: Classification of the sites according to the FAO-UNEP classification and SI. The Tweefontein and Mac Mac trials used the same weather station.....	4-59
Table 4-2: Potential increment model coefficients according to Equation 4-1.....	4-62
Table 4-3: Coefficients of the potential increment model (Equation 2) using site index and dbh as predictor variables.....	4-62

Table 4-4: Illustration of the different competition indices used in this study (Seifert et al. in review), <i>i</i> refers to the central tree, <i>j</i> refers to the competitor trees. Models shown below are discussed in more detail in text below.	4-64
Table 4-5: Correlation of the LBA competition index at different ages of the Nelder trial with the resulting search radius size included.....	4-68
Table 4-6: Linear regression coefficients of the optimum search radius based on dominant height.....	4-69
Table 4-7: Linear models coefficients of the square root transformation of diameter increment using the respective competition indices	4-74
Table 4-8: Full linear model of the square root diameter increment transformation using all of the CI's as regressor variables	4-75
Table 4-9: Importance measures of the principle components	4-76
Table 4-10: Principle component loadings of the respective principle components ..	4-76
Table 4-11: Linear model coefficients for the combinations of the competition indices	4-79
Table 4-12: Relative importance proportions (explanation contribution) of the KKL and LBA indices respectively using different importance measures	4-80
Table 4-13: Coefficients of the diameter increment potential modifier formula (Equation 4-5).....	4-83
Table 4-14: Anova comparison if Equation 4-5 with KKL, LBA and both indices included in Model 1, 2 and 3 respectively	4-84
Table 4-15: Coefficients of the diameter increment potential modifier formula (Equation 4-7) incorporating the water index	4-84

Table 4-16: Anova comparison of the diameter increment model with (Model1) and without (Model 2) the water index	4-85
Table 4-17; Anova comparison of the diameter increment model with (Model1) and without (Model 2) the water index by randomly subsetting data from the Nelder trial to match the number of observations in the Tweefontein trial	4-85
Table 4-18: Linear quantile regression coefficients of 0.05 and 0.95 tau values representing the upper and lower bounds of simulated residual prediction	4-90
Table 4-19: Average of observed and predicted values for the different Nelder plots, showing clearly that the model underperforms at extreme densities	4-98

Table of Equations

Equation 1-1	1-7
Equation 1-2	1-10
Equation 1-3	1-11
Equation 3-1	3-25
Equation 3-2	3-28
Equation 3-3	3-37
Equation 3-4	3-45
Equation 4-1	4-60
Equation 4-2	4-62
Equation 4-3	4-74
Equation 4-4	4-75
Equation 4-5	4-83
Equation 4-6	4-84
Equation 4-7	4-84
Equation 4-8	4-91

Chapter 1: Introduction and background

1.1 Problem statement

Due to the fast growing, even-aged and single-species nature of the local plantation stands, forest growth modelling in South Africa has, for good reason, mostly been confined to computer implanted stand models and yield table predictions (Kotze *et al.* 2012). These models have proven to project stand volume with high accuracy and are well suited to the homogenous situation of the stands applied. However, stand models like these have limitations. Top-down approaches have to model changes and distributions, or in other words have to decompose the stand volume again to obtain individual tree volumes. Stand models further are not designed to model irregular stand structures or deviations from typical silvicultural regimes. Individual tree growth models, following a bottom-up approach that models interactions between a mosaic of individual trees (Munro 1974), offer an alternative to stand modelling. Due to their interaction and feedback loops between stand structure and growth they are more flexible to changes in the stand structure and prove very useful for scenario analysis, simulation for research purposes and academic tools (Pretzsch *et al.* 2002).

Competition indices are a key element of individual tree models since they describe the interaction between the individual trees in the stand. Different indices capture different modes of competition, which is related to the specific resource limitation influencing the interactions among the trees as shown by Seifert *et al.* (in review). This also feeds back to the interaction between trees of different dimensions. In predominantly light limiting environments, overtopping of the crowns of the trees cause the competition of the trees to become size-asymmetrical, where larger trees benefit disproportionately to their size. When edaphic resources such as nutrients and water become more limiting the size symmetry shifts to situations where trees benefit proportional to their size (Schwinning and Weiner 1998). Competition models should be able to capture the variability and gradient between these two extremes.

In this study an individual tree modelling approach was undertaken for *Pinus elliottii* plantations in South Africa as an initial attempt to parameterise a growth model following the structure of an existing European growth simulator called SILVA (Pretzsch *et al.* 2002). SILVA was parameterised for central European growth conditions, which are predominantly light limited. That made some model changes necessary in order to adapt the simulation approach to predominantly edaphic limited sites in South Africa. Two aspects of the model were addressed: the potential height modelling required for the simulation initialisation, as well as diameter increment modelling. With application to South African forest inventory

standards in mind, potential height predicted from site index was modelled. For the development of diameter increment models, different competition indices were combined in order to capture the variable nature and shift of resource limitation and competition mode.

1.2 Background and literature review

1.2.1 *Pinus elliottii* plantations in South Africa

Pinus elliottii (Engelmann), commonly known as slash pine, is an introduced species grown typically in even-aged commercial plantation forests in South Africa. It is indigenous to the South Eastern United States; found predominantly in the coastal plains of North and Central Florida, although its range extends into neighbouring states as well (Poynton 1979). In South Africa, it has a relatively long history of use in the commercial forestry sector, with seeds first imported in 1916 and extensive expansion occurring since.

In South Africa, *Pinus elliottii* has a very wide planted range in both the summer and all-year rainfall regions, including a very wide altitudinal gradient. It is known as a hardy, relatively slow growing species that is adaptable to many different site conditions (du Toit 2012). The thick outer bark and its specific bark structure also makes it one of the most fire resistant species (Odhiambo *et al.* 2014), allowing it to be ameliorable to preventative under canopy burning and use as a buffer in fire prone areas.

This species was chosen for this thesis due to its representation in almost all of the growing regions in South Africa.

1.2.2 Forest growth modelling

Although many modelling approaches exist, in this chapter three main modelling approaches are mentioned: statistically based stand models and individual tree models and process based models.

Stand models represent the whole stand as a unit and do not consider individual tree interactions for model construction or parameterisation. These growth models usually make use of variables such as stems per hectare (SPHA), basal area (BA) per hectare and dominant height (Vanclay 1995, Kotze *et al.* 2012). Stand models can model individual tree or individual tree classes of diameter at breast height (DBH) for example by modelling distributions, e.g. Kassier (1993), in a typical top down approach. By combining stand table projection methods with the stand level model approach, projection from an observed inventory into the future can provide accurate stand table details (Corral-Rivas *et al.* 2009)

Individual tree models use information of individual trees in the simulation and model prediction of tree growth, and require at least the size of every tree in a stand and model interactions between the trees in a mosaic in the stand (Pretzsch *et al.* 2002, Vanclay 1995).

The spatial position of the individual trees is an optional parameter included depending on the structure of the model. Competition indices are often noted as distance dependent or distance independent competition indices (Pretzsch 2009). While statistical individual tree models have the advantage to be sensitive to irregular tree distributions for example due to insect damage, wind or snow breakage or inappropriate silvicultural practice (Ackerman *et al.* 2013) they share a major trait with statistical stand models: They are strictly speaking only valid in the range of their parameterisation. Extrapolations must be done with care, which limits their application in projections during transitional stages such as the experienced climate change.

Ecophysiological process based models do also follow the individual tree (e.g. Rötzer *et al.* 2012) or stand (Landsberg and Waring 1997) paradigm. Big-leaf stand models such as 3PG (Landsberg and Waring 1997), that make use of mechanistic processes, and allometric relationships to model resource allocation on a stand level have been successfully parameterised for some South African tree species (Dye *et al.* 2004). Process based models offer superior climate sensitivity, since they predict growth from the basis of ecophysiological processes (“white-box” approach). Unfortunately, developing mechanistic process models tends to be an extremely complicated and time-consuming task due to the complex nature of these processes, and these models are often parameterised from intense measurements of a few sites (Pretzsch 2009, Rötzer *et al.* 2009). Further limitations of the mechanistic process models, in comparison to currently applied statistical models, are that they lack the accuracy and practical output options such as log classes, harvesting costs and wood quality considerations, necessary for decision systems to support forest management on a commercial scale. However, this makes them currently more suited as research tools.

Hybrid models represent a compromise between empirical models, such as the individual tree and stand models mentioned previously, and process based models, by estimating productivity in relation to primary factors (Dzierzon and Mason 2006, Pretzsch 2009) and combining a solid forestry output with increased climate sensitivity.

Currently, growth and yield modelling in the South African plantation industry is characterised by “black box” statistical stand models, which are reasonably accurate, but unable to cater for changing site conditions, and are in danger of becoming less useful due to the environmental changes brought about by climate change (Pretzsch 2009). The stand models currently in use are also not designed for the complexities of intra-specific competition in even-aged, single species plantations under changing conditions or stress. This is a major concern in the climate change context, since statistical models are able to model growth only within a range or close to the conditions under which the measurements for the model were made as pointed out before. Extrapolation beyond this parameterisation range could lead to false conclusions.

Individual tree growth models are able to account for changes to stand structure, which occur due to mortality, mechanised row removal, etc. Changes to silvicultural regimes, such

as changing planting or thinning density, age mixture, etc., are easier to model by virtue of the bottom up approach taken to individual tree interactions and competition analysis (Pretzsch 2009). Preferably they should have certain climate sensitivity in a hybrid approach.

1.2.3 Competition Indices

There are different competition indices (CI's) applied in individual tree modelling, position dependent and position independent being the two major distinctions. While position independent CIs only use the relative size of a tree in comparison to other trees in the stand; position dependent competition indices calculate competition based on the explicit positions of the trees in relation to each other. Firstly the neighbouring trees that compete with the central tree are determined based on their position and whether they would influence the tree in question. Once the relevant neighbours have been found, the strength of the competition for each of the trees and their effect on that tree is determined.

When stands are homogeneously structured, position independent indices do not differ from position dependent indices significantly; however they are not deemed flexible and accurate enough to capture irregularity in the stand, thus position dependent models are better suited to extrapolating across a broad range of conditions and stand structures (Pretzsch 2009).

Many methods exist for the determination of the competitor trees. Fixed radii can, for example, be defined around the central tree. All trees with a distance less than the radius away from the central tree are identified as competitors, as in the case of Hegyi (1974) who used a radius of 3.048m (10ft). Fixed radii have severe disadvantages though as they are only adequate for certain sizes of trees under a specific competitive circumstance (Pretzsch 2009). The search radius would have to adjust for tree size or tree age for instance.

The next step in calculating a position dependent CI would be to determine the strength of the competition determined from specific attributes of the trees such as diameter, height, crown size, or a search crown.

1.2.4 Competition symmetry

Competition symmetry refers to a plant population where the size of a plant in competition to other plants determines its relative competitive advantage over neighbouring plants (Schwinning and Weiner 1998, Stoll *et al.* 2002, Wichmann, 2001). Size symmetry of competition can simplistically be grouped into size asymmetric, size symmetric and symmetric competition. This point is illustrated in **Figure 1-1**.

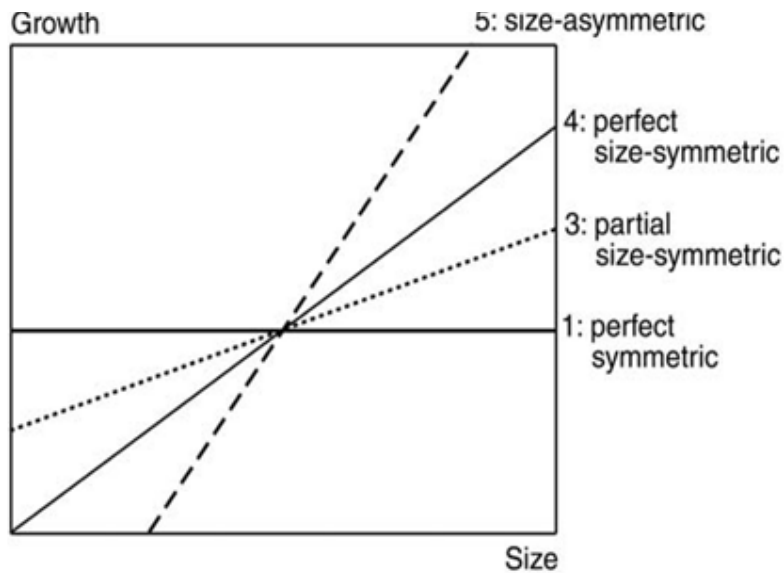


Figure 1-1: Hypothesis on the relationship of plant size on growth rate in a given stand with respect to the size symmetry of competition. Stands are usually in a continuum between these extremes where light limitation result in size asymmetric relationships and size symmetric relationships where underground resource limitations are prevalent. Symmetric competition occurs when competition is not related to tree size (figure from Pretzsch and Biber, 2010).

In forest communities growth is either size asymmetric when dominant or larger plants benefit disproportionately to their size or size symmetric where plants benefit proportional to their size (Schwinning and Weiner 1998, Wichmann 2001). The theory being that in light limited circumstances – asymmetric competition is more prevalent because light is a directional resource, where dominant trees not only gain because their canopies are larger and able to capture more sunlight, but also shade out their competitors giving those trees a distinct advantage disproportionate to their size. In more size symmetric circumstances, edaphic factors (water or nutrients) are usually limiting and therefore, since these components are not as strictly spatially dependent, the competition is symmetric with size, meaning larger trees benefit proportional to their size as they are able to obtain more resources.

This phenomenon can act on two scales. One is where a site inherently has certain characteristics, for example an ample water supply and is dominated by size asymmetric for instance. The mode of competition can also vary with time, either seasonally or between different years, for instance in drought years or years of above average rainfall. Wichmann (2001) found that in the same sites for different years, trees in their respective stand dynamics can actually shift between these two types of competition mode.

Silvicultural regimes can be catered to the specific mode of competition in productive forest stands. In predominantly light limited stands, “thinning from above” is applied, where certain trees with large canopies are removed to allow for more light to reach neighbouring trees, which can result in overall higher productivity. In more size symmetric and water

limiting circumstances, “thinning from below” is mostly practiced, whereby underperforming suppressed trees or smaller trees are removed to allow space and boost growth for larger, better performing neighbouring trees which results in optimum size and volume growth (Pretzsch 2009)

Individual tree forest growth models rarely cater for these complexities of the mode of competition (Pretzsch and Biber 2010). One option would be to combine different competition indices under such different sites (**Figure 1-2**), with the relative strength of each index changing for the different nature of the sites – i.e. one competition index would account more for competition between light, while another would account for below ground competition for resources (e.g. soil water) .

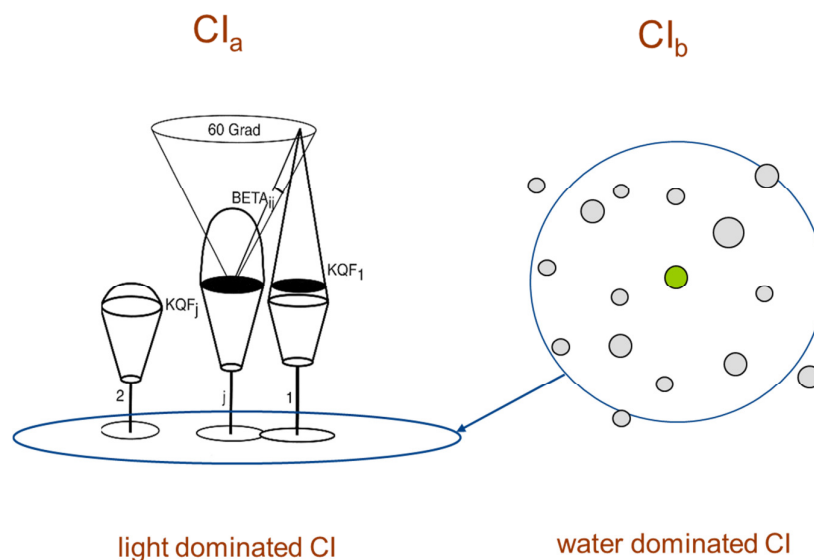


Figure 1-2: Figure illustrating two different types of competition indices theoretically more suited to overtopping and local crowding respectively.

As will be shown in this thesis, parameterisation of such models require an adequate number of sites of different average resource conditions, and longitudinal data to capture temporal changes as well, making these types of modelling strategies difficult to parameterise. However, the value of these types of models would be in their ability to cater for changing climatic conditions and their ability to adjust silvicultural regimes specifically to sites based on their inherent resource limitation.

In South Africa the commercial plantation forest environment is predominantly water limited with adequate light supply, so water or edaphic limitations are usually prevalent, leading to the hypothesis that growth is more size symmetric. However, it is not known at which point water is not limit anymore and light becomes a limitation (if such sites exist), and furthermore whether this changes seasonally or annually during different climatic weather conditions.

1.2.5 Vulnerability of forestry to climate change

Currently, the outcome of climate change in South Africa, and specifically the area under afforestation, is still uncertain (Warburton and Schulze 2006). The predicted effect on precipitation and temperature is diverse (Warburton and Schulze 2006); some regions may experience increased precipitation, while others experience drier conditions. Temperature in general is predicted to increase – but with varying levels in different regions.

The impact of climate change is thought to have substantial effects on the pest and disease occurrences and outbreaks, due to shifts in the climatic gradient and the expansion of pests and diseases into new environments (van Staden *et al.* 2004). New regions for forest growth may also become available due to the reduction of frost occurrences and a shorter frost season (Warburton and Schulze 2006); in particular the high-altitude high rainfall sites of the Drakensberg and the Highveld regions of South Africa.

Generally, the effect on plantation forestry is thought to be severe, due to the long planning horizons that are characteristic of forestry and vulnerability of plantations in the current sites, many of which are on the marginal scale of production (Fairbanks and Scholes 1999, Warburton and Schulze 2006).

Improved predictive models and decision support systems will be necessary to be able to adapt to such changes, by allowing proactive management and anticipating expected climate outcomes. For this to be realised, a high level understanding of forest growth and development under a range of conditions is necessary (Seifert *et al.* in press). Due to the significant potential impact of climate change on forestry and the relative speed at which it is developing, it is clear that the time period for response, especially for an industry dependent on long rotations, is extremely short. Before any major adaptive and reactive approaches can be made, a significant advance in the understanding of the growth of South African forests under a wide range of site conditions is necessary.

1.2.6 Potential modifier approach

The potential modifier method is an approach to controlling individual tree growth in a simulation model. In this approach, an assumed potential increment is obtained, usually from potential height-age or diameter-age curves of the upper boundary lines of a given site (Pretzsch 2009), which then represents the growth of a tree in the absence of competition. The real or predicted tree growth (i_{pred}) is obtained by multiplying this potential growth (i_{pot}) by a modifier (*mod.*), which reduces the potential growth (**Equation 1-1**).

$$i_{pred} = i_{pot} \times mod. \quad \text{Equation 1-1}$$

The modifier is a representation of competition state of the individual and has a value of between 0 (for extreme competition) and 1 (for no competition). This competition is

expressed by CI, which quantifies the competitive status of trees. The competition status is then included into a modifier function which scales the competition appropriately.

Potential Growth

Potential growth is defined as the maximum level of production possible for a tree or stand under a given state of optimal growing conditions (Reed *et al.* 2003). This is an important part of many model components, as it represents the starting point for growth modelling before the real growth is determined by the modifying factors. Determining the potential growth is commonly seen as a problematic component, because it is often difficult to observe (Bragg 2001), and presents further problems when deciding on a definition of potential growth.

Potential height is often used as a starting point to characterise the potential growth in a stand or single trees. The benefit of using the height to determine the potential growth is that dominant stand height can be used, or is closely linked to a predetermined potential, as it is not strongly affected by thinning (Ritchie and Hann 1990). Many forest inventory systems already measure dominant height and thus the information is often readily available.

The JABOWA model series (Botkin *et al.* 1972), a precursor to most North American gap models, uses inferred leaf area for its calculation of maximum height and diameter in its sub model "GROW". The individual tree-based growth simulator SILVA (Pretzsch *et al.* 2002) on the other hand, incorporates a site model, determined from a predefined list of site conditions, resulting in a site factor (**Figure 1-3**), allowing determination of the potential height growth of a single trees through the use of a Chapman Richards type equation – the basis for further growth calculations in SILVA.

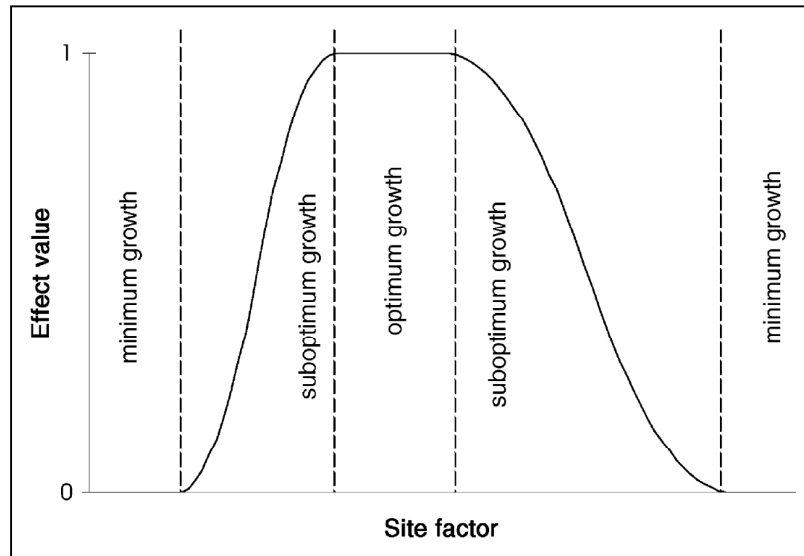


Figure 1-3: Illustration of the site factor model used in SILVA, used as the precursor to modelling potential height growth (Pretzsch et al. 2002).

The site dependent potential growth can be determined by:

- Defining and measuring a top percentage of trees under observed conditions
- Using open grown trees
- Applying treatments to stands to provide conditions for optimal growth.

The potential growth may be obtained by selecting a predefined top percentage of trees representing a maximum growth range, usually from an extensive database reference, for example in SILVA (Pretzsch *et al.* 2002). This can be useful as it represents realistic, observed values for practical forest management as these values are often obtained from large databases of repeated measurements (sample plots, etc.) and, as mentioned previously, are often already measured in many forest inventory systems. Vanclay (1995) suggested that one possible problem with this approach is that the estimates may sometimes select for measurement errors rather than real growth.

The potential growth may also be determined based on open grown trees. Botkin *et al.* (1972), including other gap models, used this definition to determine maximum growth for a forest gap model. This method is in many ways theoretically sound as it represents a competition-free state, but it has some limitations as it is often very difficult to find open grown trees in reality, especially for frost intolerant species (Bragg 2001) and in particular shade tolerant broad-leaved species will not show a maximum height growth in a solitary state but rather grow extensive lateral crowns (Uhl *et al.* 2006). So the use of solitary trees might be restricted to frost resistant pioneering species that have a clear acrotony in their growth pattern.

Edaphic limitations on their own are often used to determine the potential. The North American gap model JABOWA reduces the maximum height and diameter formed under optimal conditions with modifications made for shade tolerance, soil quality, and average climate (Botkin *et al.* 1972).

Reed *et al.* (2003) used a different approach to determine the potential growth (height and diameter) for *Eucalyptus globulus* in Portugal by opting to eliminate nutrient and water limitations by fertilisation and irrigation. The potential height or diameters were then modelled by using the cumulative air temperature heat unit (not described here) using a modified Chapman Richards type equation

$$Y_j = a * (1 - e^{-b * TDD_j})^c \quad \text{Equation 1-2}$$

Where:

Y_j = Potential height or diameter growth

TDD_j = Air temperature heat sum unit

a, b, c = Coefficients to be estimated

This approach was very successful in determining the potential growth and was unique in the sense that it tried to create and parameterise the model according to optimal conditions created by removing belowground limitations to growth, although modifiers were not developed for a feasible model and a large amount of information regarding ambient temperature was necessary. The added benefit of such an approach is that it can observe and then model responses for predefined treatments and conditions (e.g. water supply, stand nutrition, etc.). However, obtaining a full range of such conditions and separating response factors could be problematic.

The Modifying Factors

The modifier is basically a reduction factor, which represents competition or any other limitation to growth, e.g. light, water, CO₂, etc. (Porte and Bartelink 2002). Theoretically, these factors are usually based on the assumption described by Liebig's Law of the minimum which states that the resource that is in the shortest supply will most affect the growth rate of the plant.

These multipliers are typically exponential in form, especially under high competition (Ek and Dudek 1980). Diameter and height responses to competition, and their manifestation in the modifier response, are usually different (Seifert 2003). **Figure 1-4** shows the common form of this response.

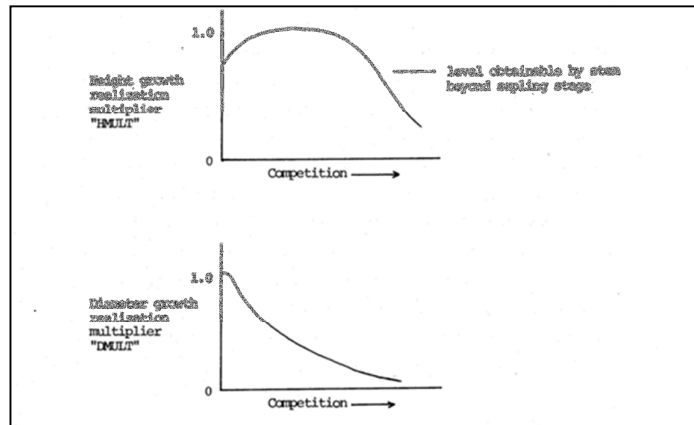


Figure 1-4: Effect of competition on the modifier values of dbh and height. From Ek and Dukek 1980

Competition is a common variable used for determining the modifier. SILVA uses a combination of competition indices and crown dimensions on the individual tree for both height growth and basal area growth (Pretzsch et al. 2002).

$$zh = zh_{pot} \cdot C_5 \cdot (1 - e^{-c_6 \cdot CSA}) \cdot e^{-c_4 \cdot (1+KMA)^{c_1} \cdot (1+NDIST)^{c_2} \cdot (KKL+c_3 \Delta KKL)} \quad \text{Equation 1-3}$$

Where: zh is potential height growth, CSA is the crown surface area, KKL is the competition index, NDIST is a measure of the centre of competition from the stem centre of the subject tree and KMA is a conifer-broadleaf tree neighbour effect.

In this way light limitation with regards to competition is taken into account, but in the site model (which determines the potential growth) SILVA takes into account a variety of site conditions such as mean and minimum temperature, soil nutrient supply, CO₂, soil water retention, etc. which gives the model a complete structure with added sensitivity to site changes such as global climate change (Porte and Bartelink 2002). Thus while not including these factors explicitly as modifying factors, the potential is determined, with competition doing the rest, resulting in a simplified model structure without the interactions between competition and site conditions.

In the growth simulator EFIMOD2, Komarov *et al.* (2003) used a slightly different approach, focussing on tree nutrition in boreal forests, whereby each tree occupies a certain space above and below ground in a “single plant ecosystem”. In this case the focus was on nutrient limitations with a complex set of soil nutrient models, primarily because the boreal forests in the study were highly dependent on nitrogen availability, with light as a secondary growth limitation, which usually comes into play in sub-dominant trees (Chertov 1990, Komarov *et al.* 2003). This soil fertility based approach is rare, especially at its rate of sensitivity (Porte and Bartelink 2002).

1.2.7 Silva

The distance dependant individual tree forest growth simulator SILVA, has been successfully implemented in Germany and other European forests, and provides the flexible, high level understanding and application to forest growth necessary to predict changes in forest growth under various site conditions and climate change scenarios (Pretzsch *et al.* 2002, Pretzsch 2009).

As a consequence of its individual and fully spatial model approach SILVA has proven to be a versatile tool for prediction of forest growth. It is for instance currently used by the Bavarian State Forest Enterprise one of the biggest forest owner in Central Europe as a standard tool for sustainability planning. SILVA has proven its worth as a tool for scenario simulations in order to optimise silvicultural thinning regimes in a variety of applications, including long term sustainability planning, wood quality prediction (Seifert and Pretzsch 2004), pest management (Seifert 2007), biomass prediction, etc. SILVA is being extensively used for practical growth modelling and application for forestry, it is used as an education tool, scenario analysis, climate change projections and visualisation for virtual forest stands for a wide range of applications.

Parameterisation of this model though, has not yet been performed in a South African forest context. This task requires substantial remodelling and re-parameterisation in order for the model to become functional in South Africa. A first NRF-financed pilot project at the Department of Forest and Wood Science of Stellenbosch University in collaboration with the working group of Prof. Hans Pretzsch at TU München Germany, showed that due to the different resource limitations SILVA cannot simply be re-parameterised but needs to be structurally changed in the sub-module for competition calculation.

To achieve this adaptation to South African forests and growth conditions two main steps are addressed in this thesis: the potential height stand initialisation model (**Figure 1-5**) and the diameter increment model based on competition indices (**Figure 1-6**).

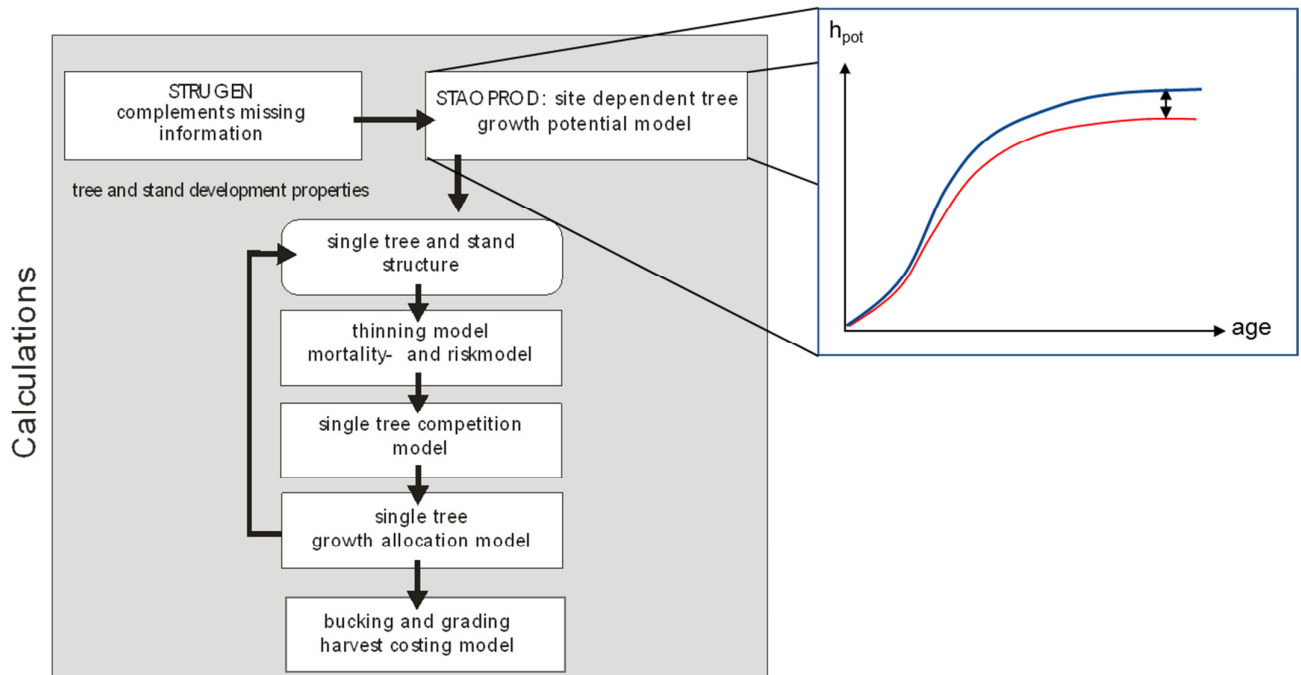


Figure 1-5: Silva flow diagram and the potential height-age relation in SILVA predicted from site quality. This curve is used for the simulation initialisation

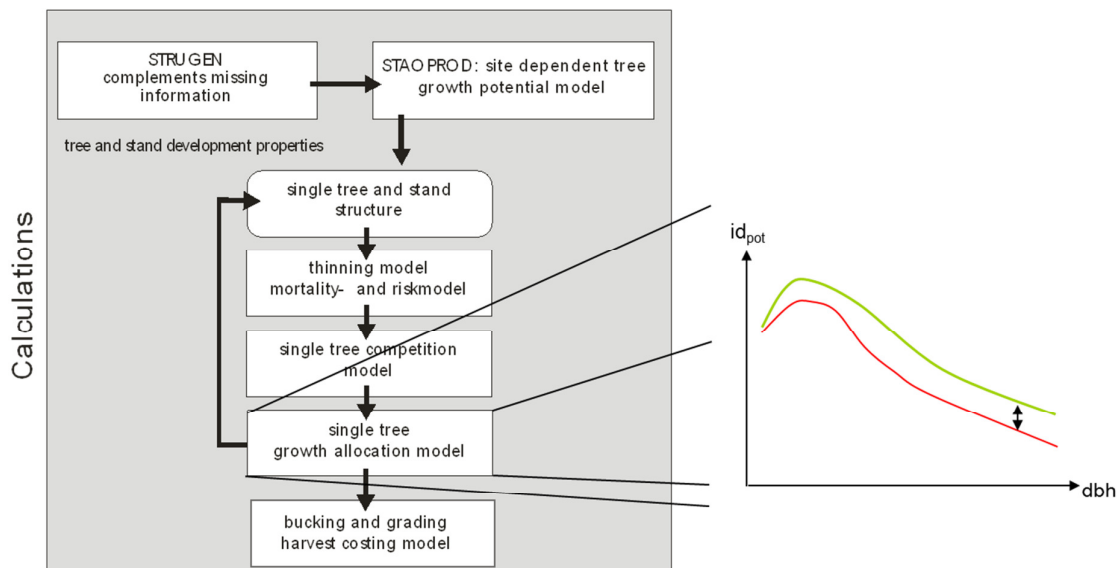


Figure 1-6: The diameter increment model function in SILVA determined by diameter (DBH). Predicted increment is then based on a modifier which takes competition into account.

1.3 Objectives

The main aim of the study was to introduce a methodological approach for individual tree models for South African forestry using the SILVA structure as an example. Two important

sub-components of the model were addressed: the potential height model and the diameter increment modelling. The sub-objectives of these two main components were addressed below:

1.3.1 Potential height model:

The objective was to model potential height as a function of site index, the South African standard for site quality description. Thus three working steps were identified as the sub-objectives:

1. Modelling site index using different model fitting methodologies
2. Finding a suitable method to fit potential height models on observed data
3. Predicting potential height from site index

Step 1 was used as a study into different methodologies for site index model fitting techniques. These estimations were then used for the prediction of potential height (Step 3). Step 2 involved finding a methodology to create potential height-age functions using the Chapman Richards equation, while Step 3 involves using SI to model potential height.

1.3.2 Diameter increment model:

As diameter increment is highly sensitive to competition between trees in a stand, the objective was to model diameter increment using the potential modifier methodology under changing resource limitations. The sub-objectives (steps) used to achieve this were:

1. Classify sites according to water availability
2. Determine the potential increment based on site conditions
3. Fit different competition models
4. Select a suitable combination of indices to capture overtopping and local crowding
5. Use CI's in combination in a deterministic potential modifier model
6. Create a stochastic model mimicking natural variability

The study was thus limited to two key aspects, which are deemed important for the potential application of an individual tree model, such as SILVA, in South African plantations. The two components provided insight into potential problems and future application of the model for forest growth modelling.

1.4 The use of R in the study

For all statistical purposes R (R Core Team 2013) was used in this study. R is an open source software project freely available on the internet, with multiple packages designed for multiple purposes; updated and reviewed frequently. The packages used and the software

itself proved very useful and robust, and allowed the researcher to combine statistical, spatial modelling and illustrative purposes simultaneously. Select examples of the code constructed in the study are included in Appendix F either by the researcher or with assistance from peers. It is not referred to in text, but is included for example purposes.

Chapter 2: Dataset description

Individual tree growth models and distance dependent competition indices require adequately designed spacing trials or sampled measured data in forest compartments to capture the range of possible spacing and distribution of trees. The project received support from a wide range of industry and research components, mentioned in the acknowledgements. Data from a few spacing trials were obtained representing a large number of observations and spacing designs.

In a nutshell, these are the correlated curve trend (CCT) spacing trials (O'Connor 1935, Bredenkamp 1984) initiated in 1937, the *Pinus elliottii* trials consist of four locations. In addition one ongoing SSS-CCT (Standardised Sample Size), (Bredenkamp 1990), trial initiated in 1991 was included and a Nelder spacing trial, which was measured from 1976-1998. The datasets are explained below and were used as deemed appropriate for the different purposes and objectives of this thesis. A map of the spacing trial locations is illustrated in **Figure 2-6**.

2.1 The correlated curve trend (CCT) data

The CCT trial concepts were laid down by O'Connor in 1935 as spacing trials to determine optimum planting and thinning regimes for the South African plantation industry. The CCT trials can be split into two series: an unthinned series called the basic series that were planted under a wide range of planting densities comprising of eight stand densities in stems per hectare (spha). In this study the unthinned series was used for modelling purposes with the design shown in **Table 2-1**.

Table 2-1: CCT unthinned series trial summary

Plot	Nominal Stand density (spha)	Plot size (ha)	Measurement Trees/Plot
1	2965	0.081	240
2	1483	0.081	120
3	988	0.081	80
4	741	0.081	60
5	494	0.081	40
6	371	0.081	30
7	247	0.081	20
8	124	0.081	10

To overcome suppression by competing weeds all of the plots were planted at very dense stocking levels and thinned in advance of competition (**Table 2-2**). In retrospect, this may not be the optimum solution as the high densities at these ages could include competition in any case, or lead to facilitation or other unknown effects.

For the purposes of this study this was assumed not to have a significant effect on the results and the stands were treated for their intended SPHA. The thinned plots incorporated various thinning regimes and management practices such as weed control.

Table 2-2: Thinning in advance of coemption for weed control in the CCT trial data

Plot	Age (years)							
	0	2	5.08	6.17	7.5	8.83	9.42	10.25
1	2965							
2	2965	1483						
3	2965	1483	988					
4	2965	1483	988	741				
5	2965	1483	988	741	494			
6	2965	1483	988	741	494	371		
7	2965	1483	988	741	494	371	247	
8	2965	1483	988	741	494	371	247	124

The original CCT trials proved to be a difficult dataset to use, with measurement history, explanations, etc. not available to the researcher. For instance, many trees were simply sampled and it was not possible to determine which trees succumbed to mortality or were not included in the measurement sample. Tree positions were not available through repeated re-numbering and, because the trials do not exist anymore, could not be reconstructed. For this reason the data were used only for the potential height and site index modelling section of this thesis (Chapter 3). For this purpose, the data proved to be a good tool as it covered a wide range of site conditions and, by South Africa standards, were measured to a very high age (50yrs), which proved to be of great importance for the shape and asymptotic development of the height-age growth models.

This *Pinus elliottii* dataset comprised of four sites called Weza, Kwabonambi, Dukuduku and Mac Mac. Measurements were taken at various intervals for the four trials – for example the height measurements are illustrated in **Figure 2-1** and **Figure 2-2** below illustrate the four CCT trials with regards to the height measurements over time and the height measurements of the different stand densities of each of the sites and plot densities respectively, showing differences between site and density of each stand.

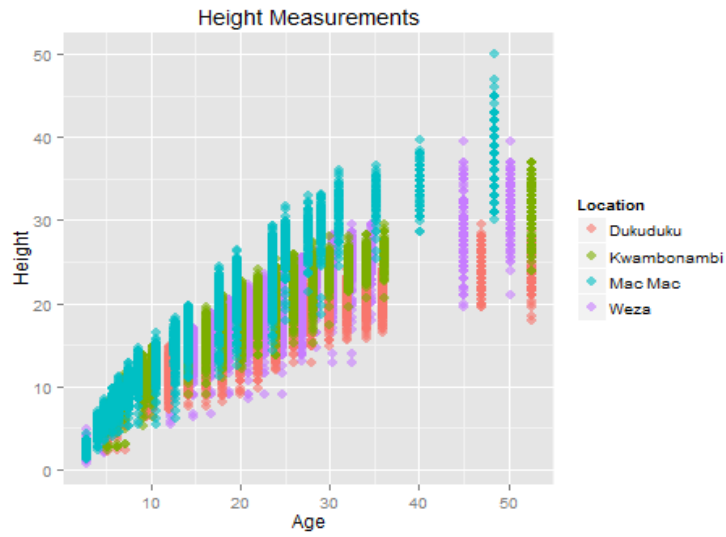


Figure 2-1: Height measurements on the four CCT spacing trials showing a clear stratification between sites



Figure 2-2: Height measurements on the four CCT spacing trial indicating different growth patterns and growth trends for different planting densities

Fig 2-2 illustrates the clear differences in height growth in sites and stand densities encountered in the spacing trials. It also illustrates the problems of re-measurement periods taken at different ages with large gaps between measurements. **Figure 2-3** and **Figure 2-4** identify trends in competition encountered in the sites, with maximum tree height responding less to competition compared to diameter; which is why the study focussed on the effect of competition on diameter increment.

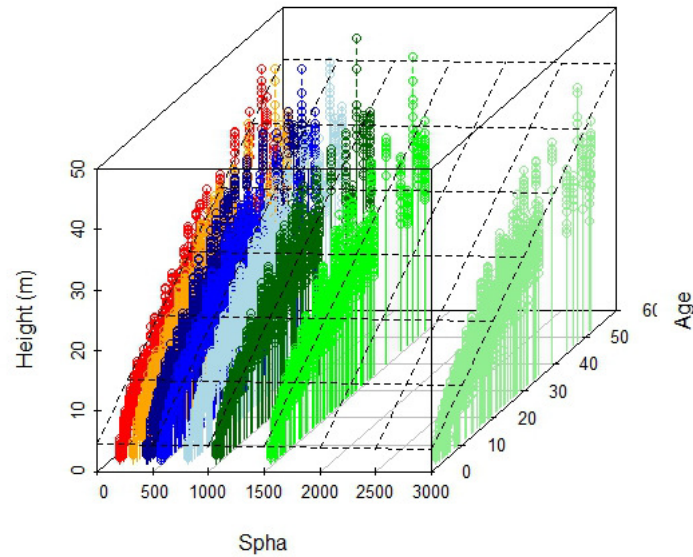


Figure 2-3: 3D plot of height growth trends in the CCT spacing trials with a linear average response curve indicating little change between maximum height growth between different stand densities

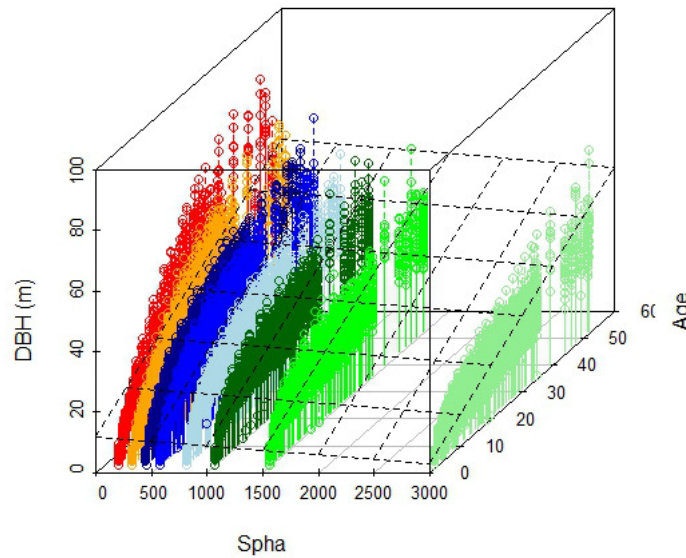


Figure 2-4: 3D plot of diameter growth trends in the CCT spacing trials with a linear average response curve indicating significant change between height growth between different stand densities

2.2 Triple S-CCT

The Triple S-CCT concept (SSS-CCT) was established and designed in the late 1980's, introduced by Bredenkamp (1990) in order to overcome the statistical shortcomings and the cost implications of the CCT trials. Six unthinned stand densities were planted at the stems per hectare indicated in **Table 2-3** below. The SSS-CCT trial used in this thesis was one of two such trials for *P. elliptii* planted in Tweefontein at an altitude of approximately 1200m.

Tree positions could be reconstructed from the planting pattern and numbering system provided by the company. For this reason this dataset was very useful as competition indices could be calculated on the site.

Table 2-3: Unthinned SSS-CCT spacing trial desing

Plot	Planting density (SPHA)	Measurement Plot size (ha)	Measurement Trees/Plot
1	245	0.102	25
2	403	0.062	25
3	665	0.038	25
4	1097	0.023	25
5	1808	0.0138	25
6	2981	0.008	25

2.3 Nelder spacing trial

The Nelder spacing trial concept is an unthinned spacing experimental design (Nelder, 1962). Different designs were created for specific objectives (**Figure 2-5**). The trial used in

this thesis was planted in 1972 in the Lotterring plantation of the coastal plateau of the Tsitsikamma region in the all-year rainfall Southern Cape Region of South Africa, based on the conventional Nelder spacing design (a in **Figure 2-5**), where the rings (or circles) represent different spacings increasing with distance from the centre. The trial burnt down in 1998, with the last measurement age at 26 years. The measurement intervals and dates are shown in **Table 2-4** below.

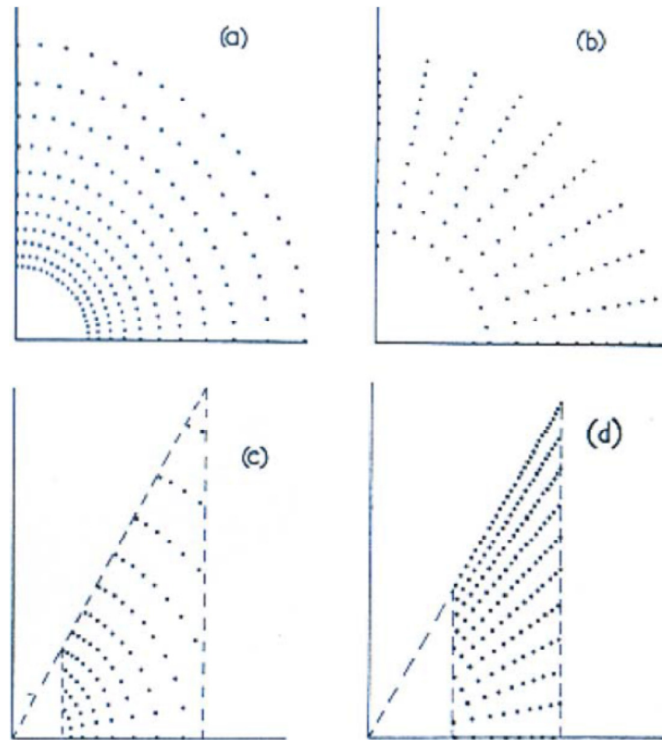


Figure 2-5: Nelder trial designs based on different spacing geometries. Nelder (1962).

Table 2-4: Measurement intervals for the Lottering Nelder trial

Remeasurement no.	Date	Age (years)	Variables measured
1	1976-12-13	4.58	Dbh, Ht
2	1978-05-08	6.00	Dbh, Ht
3	1979-05-26	7.00	Dbh, Ht, Branch thickness
4	1983-06-01	11.08	Dbh, Ht
5	1985-09-23	13.33	Dbh, Ht
6	1987-12-08	15.58	Dbh, Ht
7	1991-07-11	19.17	Dbh, Ht
8	1992-08-16	20.25	Dbh, Ht
9	1996-03-15	23.83	Dbh, Ht
10	1998-05-20	26.00	Dbh, Ht

Tree positions could be recalculated from the positions of the trees on the “spoke” and “rings” of the Nelder wheel. This means that although orientation of the spacing wheel in cardinal directions could not be determined, the relative positions of the trees with all of their neighbours could be easily calculated.

Table 2-5: Nelder spacing trial plot numbers (rings) with their corresponding stems per hectare, based on the radii and arc distances.

Arc No (Ring in Design)	Arc No (Ring no. in field)	Radii of Arcs (m)	Arc distances (m)	Distance between Arcs (m)	Area per tree (m²)	Corresponding stems per ha
3	0 Border	7.43	1.06	1.13	1.12	8900
4	1	8.56	1.22	1.31	1.49	6700
5	2	9.87	1.41	1.51	1.98	5043
6	3	11.38	1.62	1.74	2.63	3796
7	4	13.11	1.87	2.00	3.50	2857
8	5	15.11	2.16	2.31	4.65	2151
9	6	17.42	2.49	2.66	6.18	1619
10	7	20.08	2.87	3.06	8.21	1219
11	8	23.14	3.30	3.53	10.90	917
12	9	26.67	3.81	4.07	14.48	691
13	10	30.74	4.39	4.69	19.24	520
14	11	35.44	5.06	5.41	25.56	391
15	12	40.84	5.83	6.23	33.95	295
16	13	47.08	6.72	7.18	45.11	222
17	14	54.26	7.75	8.28	59.92	167
18	15	62.54	8.93	9.54	79.61	126
19	16 Border	72.09	10.29	11.00	105.76	95

2.4 Permanent sample plots (PSP) data

The permanent sample plot data was obtained with the permission of contributors and custodians of PSP inventory data from forestry companies in South Africa. These represent re-measured plots of active commercial plantations (not spacing trials). These were not used for model parameterisation as they only represented a narrow range of stand densities, but for illustrative and short validation purposes based on a few candidate sites.

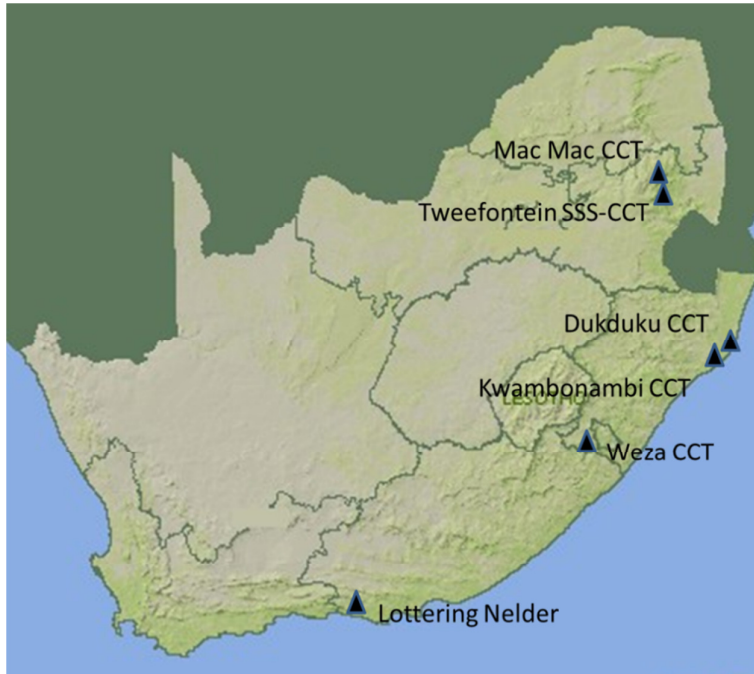


Figure 2-6: Locations of the spacing trials

Chapter 3: Site index and potential height modelling

3.1 Introduction

Silva employs a potential modifier approach for predicting height and diameter increments. Thus one of the objectives of this thesis is to obtain a potential height which could be obtained by determining potential height from site index in a model in order to obtain the potential height from a widely used and readily available site quality measure. This chapter investigates whether this is possible, how this should be done and provides an initial working model.

3.1.1 Site index

Site index is widely used in the South African forestry industry as a measure of site class and is an important predictor for many modelling and growth prediction applications (von Gadow and Bredenkamp 1992). Many definitions of top height exist, the approach most widely used in South Africa is based on the definition that top height is the height of the 20% largest trees (according to their DBH) in a stand, which is usually estimated by calculating the quadratic mean diameter of the largest 20% of trees and substituting this for D in the following formula (Bredenkamp 1993):

$$H = e^{(b_0 + \frac{b_1}{D})} \quad \text{Equation 3-1}$$

In this study, the relationship between site index and potential height curves was sought out. Site curves over age were modelled to understand the change in relationship over time.

3.1.2 Potential height growth

Potential height over age is a measure of a stand's height growth and represents an upper boundary line or curve of measured tree heights for a given site (Pretzsch 2009, Seifert 2003). In the SILVA methodology, a potential height–age curve is developed as an initialisation of the model process (Pretzsch *et al.* 2002). Therefore, a methodology had to be developed using the available data to quantify potential height development over age.

Site index values are typically obtained from plantation inventory and are not explicitly modelled in this thesis. One possible pitfall of using site index in this context is that site index on the same site can vary according to varying (stems per hectare) densities (**Figure 3-10**). In this chapter, while on the same site, each plot representing a planting density was treated as an individual with its own site index in order to determine the relationship between site index and potential height.

3.1.3 Nonlinear regression assumptions

Developing site index curves on a longitudinal, repeated measures dataset presents certain problems. Standard regressions rely on inherent assumptions to the modelling process such

as homoscedasticity of residuals and independence of the residual errors. Longitudinal, repeated measures violate the assumption of independence

Serial autocorrelation

While simple least squares estimation has been used for modelling data of this kind, a relatively simple and proven method; the longitudinal nature of the data presents a significant problem. Each of these datasets represents repeated measurements over time, which violates the assumption of independence of errors; in this case the errors are correlated with the error of the previous and later measurements of the same tree.

Heteroscedasticity

Heteroscedasticity is a common problem in model fitting, whereby the residuals increase along a predictor gradient – for example age. This was a common problem in the modelling of height in the study sites

3.1.4 Chapter objectives

The objectives of this chapter are thus to:

- Develop continuous site index curves that take into account heteroscedasticity and autocorrelated time series errors
- Develop a suitable methodology to develop potential height-age curves
- Assess the effect of density on site index for the same sites and over a comparison of sites
- Fit a model that predicts potential height from dominant height

Thus the results of this chapter are split into three steps

1. The development of dominant height-age curves
2. Nonlinear quantile regression for potential height modelling
3. Prediction of potential height from site index

3.2 Step 1: Site index modelling

This section deals with the calculation of site index. In the following sections the Weza Dataset was used to demonstrate the modelling methodology. It was deemed more suitable than the younger CCT datasets as it was assumed that only a data set with older trees would reflect asymptotic growth – which will be shown to be a very important factor. This section outline **Figure 3-1** shows the conceptual flowchart of the outline used in this section with brief descriptions.

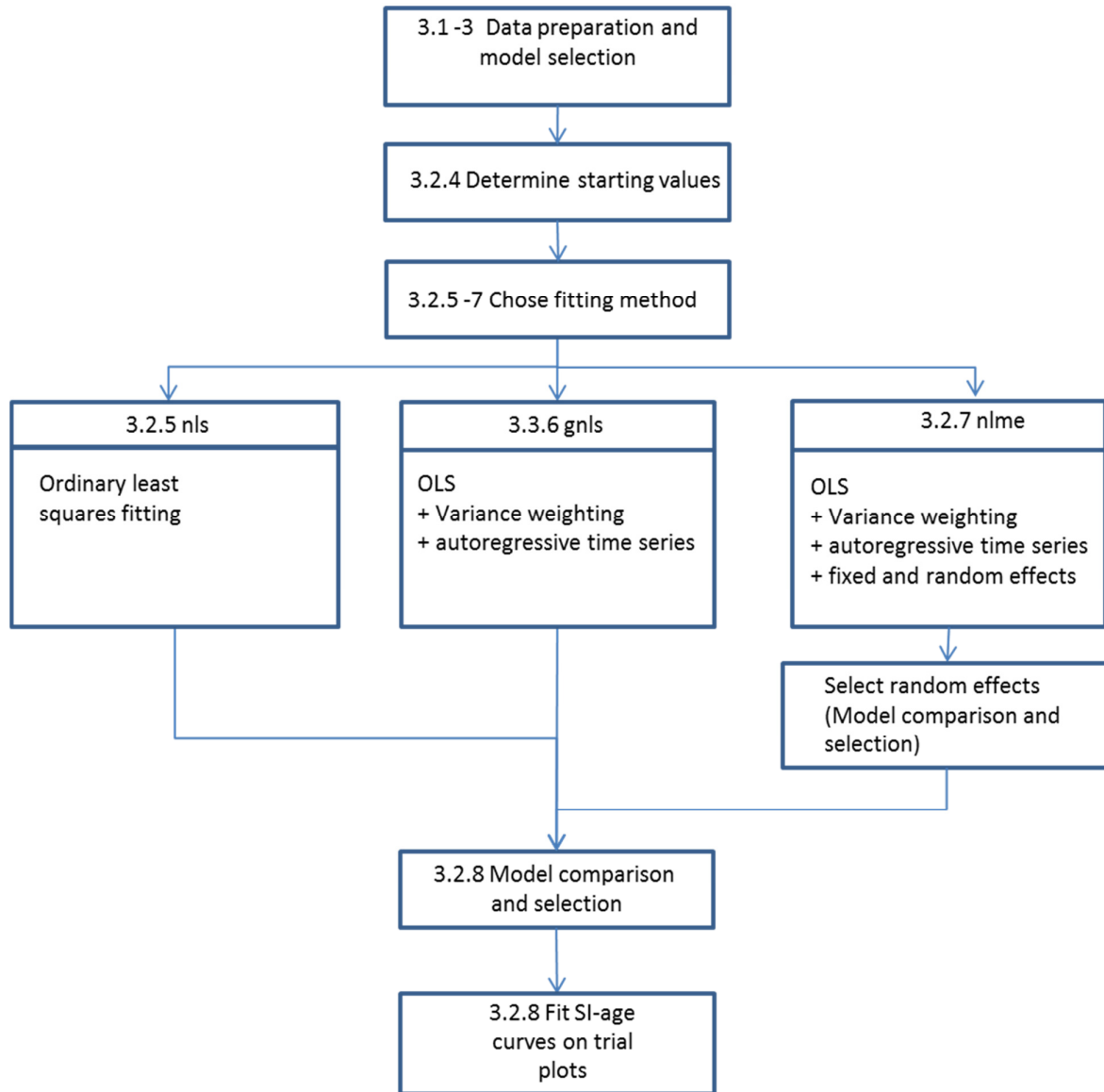


Figure 3-1: Flowchart of site index modelling section of this chapter (Step2)

3.2.1 Identification of dominant trees

Each plot from each site of the spacing trials was modelled separately. The 20% largest trees in respect to their diameter were selected by using the quantile function in R to bin the data into five 20% classes and then sub-set the data for the largest 20%. The trees selected were then used for further height model development, as illustrated in Figure 3-2. In the older CCT trials where the number of trees measured differed for the different spacing densities (lower amount of trees for low densities) – this represented a problem due to the differing amount of observations. However, it was decided to maintain the South African standard site index definition in this thesis and outline potential pitfalls with recommendations for possible improvements where necessary.

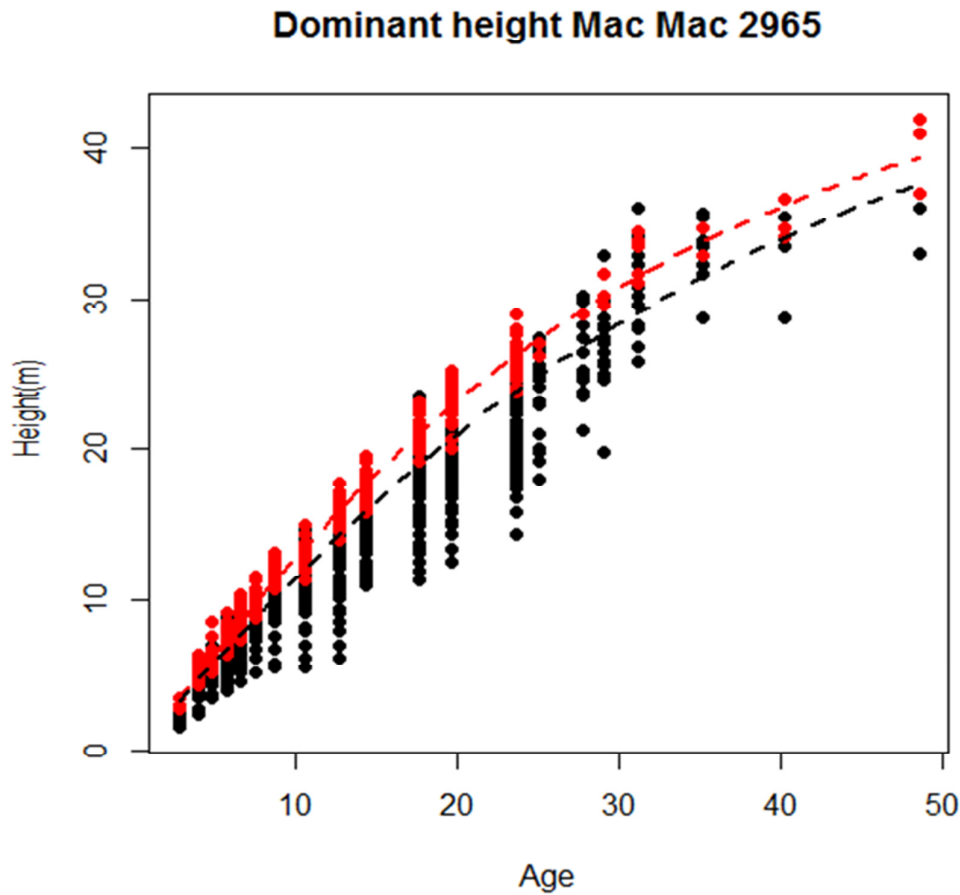


Figure 3-2: Dominant height modelling example for the Mac Mac 2965spha plot with red dots representing the dominant trees, the red and black dashed lines representing the dominant and mean height curves respectively

3.2.2 Model selection

As with the nonlinear quantile regression, the Chapman-Richards three parameter form equation was used (**Equation 3-2**) to develop anamorphic dominant height-age growth curves, which were used for site index classification for any given reference age.

$$Height = a * (1 - e^{b*age})^c \quad \text{Equation 3-2}$$

In this study, the nonlinear least squares (nls) approach was used as the standard methodology for fitting the site index function in the South African Forestry industry. The nls approach was compared to some alternative methods in order to identify the best fitting method to deal with independence violations and homoscedasticity errors. These are:

- Nonlinear Mixed Effects Model (nlme)

- Generalised Least Squares Estimation (glns)

It should be noted that parameters (slope of the curve) are often fixed in anamorphic site height-age models, often noted as **b** and **c**, and vary the asymptote (**a**) only. While this is common practice in many growth models in South Africa it was decided in this case not to fix any of the parameters in order to be able to introduce and assess random effects in the model. The purpose of this exercise is not to develop reference site index curves for management, but to model site index accurately to use as a predictor for potential height and to investigate the use of mixed effects models for site index modelling in South Africa.

3.2.3 Dataset and trial Age

As might be expected, significant differences were seen between the younger SSS-CCT spacing trials and the older CCT trials, with much more volatility in the asymptote due to the more linear growth at these ages without the realisation of the approach towards an asymptote. For this reason all of the models were fit on the original CCT datasets.

These were all tested on the Weza Dataset, whereafter the methodology was standardised for all sites. Using different random effects, weighting, autocorrelation, etc., produced markedly different results in some cases, especially at higher ages.

3.2.4 Starting values:

Starting values (sometimes called initial values) for iteration of nonlinear models can have a significant effect on the results and wrong starting values can cause significant problems in parameter estimation. The method used for the selection of starting values is outlined by Fekedulegn *et al.* (1999) and Lekwadi *et al.* (2012); the parameters for the Chapman Richards equation (**Equation 3-2**) are estimated by a simple algorithm, where:

$$a = \max(y)$$

b is defined as the rate constant at which the response variable approaches its maximum possible value b_0

$$b = \frac{\frac{w_2 - w_1}{t_2 - t_1}}{b_{0s}}$$

Or for this example rewritten as (Lekwadi *et al.*, 2012)

$$b = \frac{\frac{\max_y - \min_y}{\max_t - \min_t}}{b_{0s}}$$

c lies between zero and 1 for the Chapman-Richards as recommended by Fekedulegn *et al.* 1999. The c parameter was fixed at 0.66 or changed as seen fit to foster model conversion. This methodology worked well for the purpose of the study; improving calculation time and

producing estimates that were within the expected ranges for the parameters, resulting in model convergence.

3.2.5 Nonlinear least squares

Nonlinear Least Squares (NLS) estimation is a commonly used fitting technique in nonlinear regressions. It uses an unconstrained minimisation algorithm, which simply defines a model curve which is fit by minimizing the sum of squares that occur on the y-axis (Dalgaard 2008).

Figure 3-3 provides an example of a NLS fit in one of the CCT spacing trial plots, with some visual diagnostic plots in **Figure 3-4**.

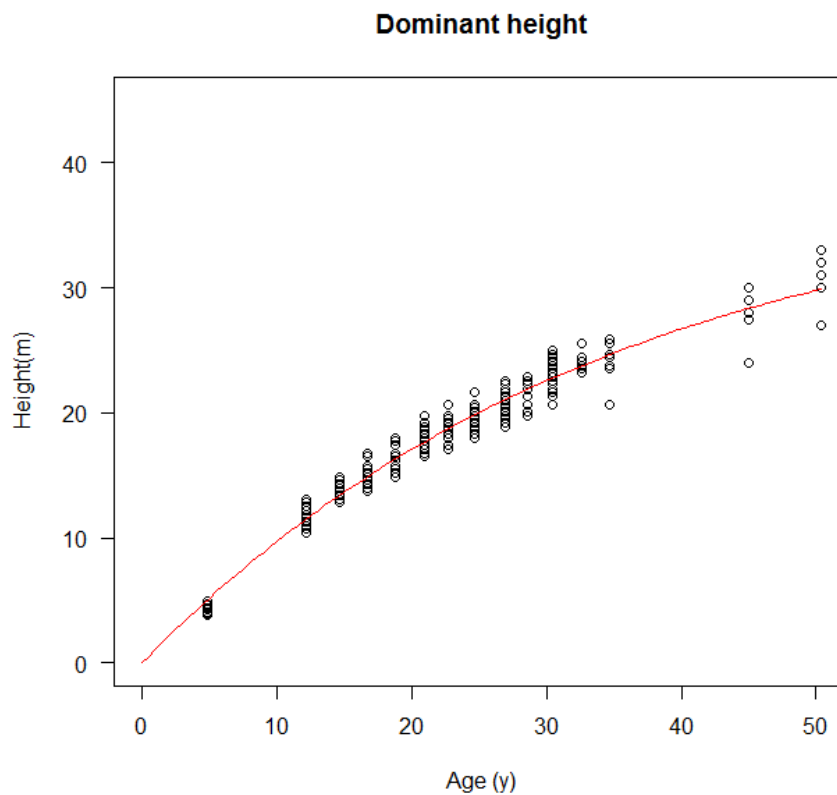


Figure 3-3: Example of an NLS fit on the Weza 2965 plot

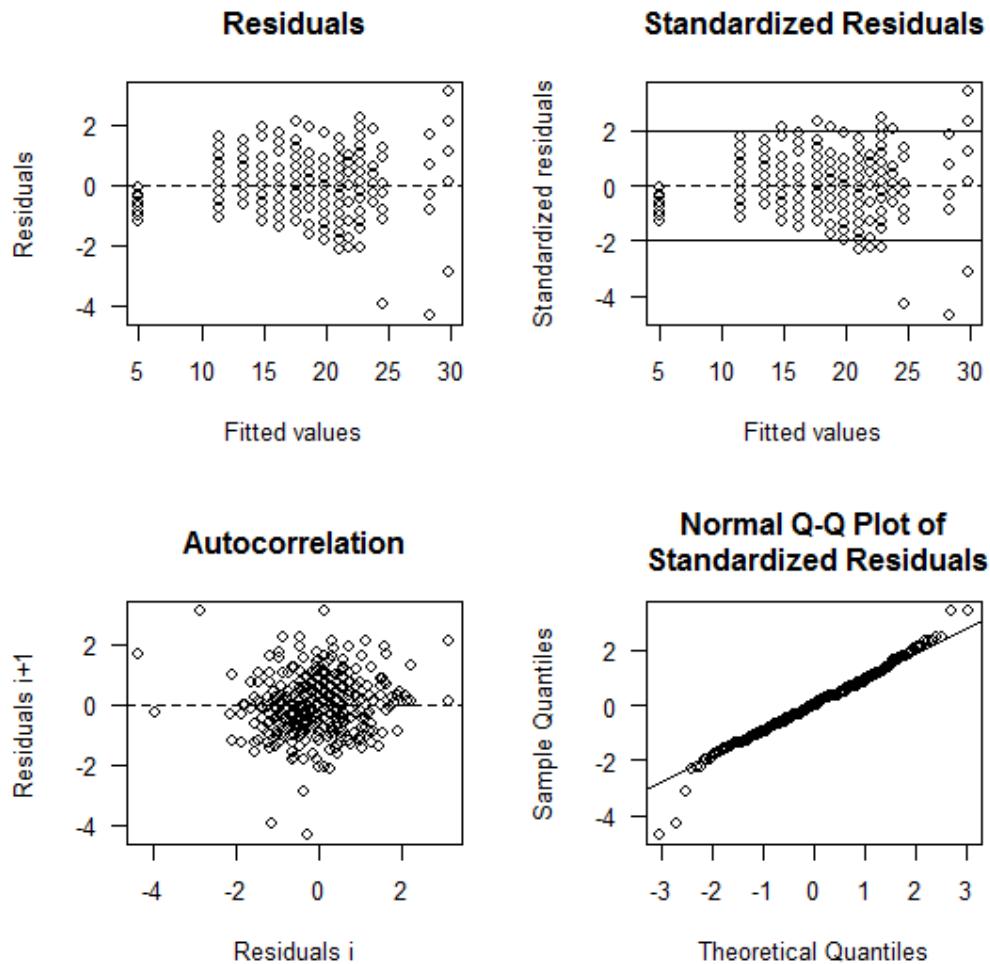


Figure 3-4: Regression diagnostics for the Weza 2965 plot

As can be seen in **Figure 3-4** above inherent problems with heteroscedasticity are prevalent, as seen in the residual plots (top right and top left images). Furthermore, while no significant trend is apparent in the autocorrelation plot in this small dataset, it is obvious that error correlation occurs in a repeated measures study such as this. For this reason, alternative methods were sought to quantify site index growth, which do not have time related autocorrelation errors. While somewhat trivial to the overall structure and contribution to the thesis, this was sought to adequately define a methodology for guide curve fitting of site index curves. The model fit shows also a model bias (overestimation of smaller values), which can be attributed to the inflexible Chapman-Richards equation. The Chapman-Richards equation is an often used model for SI modelling (Esler 2012). It was not in the scope of this thesis to compare different mathematical model formulations, however several models could present alternatives, for example the Hossfeld equation (Gea-Izquierdo, Cañellas, and Montero 2008). It is important to stress that models such as the Chapman-Richards model, which have a rigid, and sometimes inflexible structure (as

compared to polynomial models for example) are used because tree growth is based on certain assumptions and that nonlinear models such as these are applied for their interpretability, parsimony and validity (extrapolation and robustness) beyond the range of the data (Pinheiro and Bates 2000). But as seen in **Figure 3-4** this can come at a cost of a bias in the prediction of height at younger ages.

3.2.6 Generalised nonlinear least squares

Generalised nonlinear least squares (GNLS) is comparable to NLS, except that when there is reason to believe that the assumptions of equal variance and uncorrelated errors (independence) are violated, GNLS is a possible option to overcome some of these issues by the incorporation of autocorrelated error functions (ACF) and variance weighting transformations.

Variance and Autocorrelation

In order to cater for unequal variance and autocorrelation, variance weighting transformations including an autocorrelation function (ACF) were applied to the GNLS and the NLME models. A power weighted variance transformation for residuals and the ACF function for the lag factors significantly improved the models and were included from the outset; these improve the heteroscedasticity and the autocorrelated time errors in the model (shown in the heteroscedasticity improvement in **Figure 3-6** and improvements of fit in **Table 3-3**).

3.2.7 Nonlinear mixed effects model

Nonlinear Mixed Effects modelling (nlme) is a modelling technique used for grouped data (Pinheiro and Bates 2000). Mixed models effectively split the variance in fixed effects that can be explained by factors and random effects that cannot be explained but are nonetheless inherently present in the dataset such as observed variability within trees and between trees and sites and can be at least attributed to those clustering entities. In this case no distinction was made according to the differences in site; however the deviation of the parameters from every tree were defined as random effects. Thus in this case the deviation of parameter estimates for every tree (the random effects) must be chosen by deciding, which parameters to include as random effects.

The methodology applied to developing site index curves with mixed effects modelling was done by following the procedures set out by Pinheiro and Bates (2000) for mixed effects model fitting, and Fang and Bailey (2001) for application to forestry related examples (incidentally, also fitted for *P. elliottii*).

In describing the steps to specify fixed and random effects Fang and Bailey (2001) suggest that the nature of the Chapman-Richards parameters (a, b, c) be specified as fixed and random effects or purely mixed effects. According to Fang and Bailey (2001) model comparison can start by examining the full model with all the parameters as random effects without considering covariates, and then picking the parameters or a combination one by

one, as described in Pinheiro and Bates (2000). While the full models (considering every coefficient as random effects) can sometimes improve model fits, they often do not converge. There is also a risk of over-parameterisation when applying such a methodology (Pinheiro and Bates 2000).

Consequently either one or a combination of two of the parameters was tested as random effects. For this reason the NLME section starts by describing how the different parameters affected the model in **Table 3-1** and **Appendix A** over a wider range of plots.

Random Effects

As mentioned earlier, a decision had to be made on which components to include as random effects. First it is necessary to see whether there is correlation between the parameter estimates; this can simply be seen visually in a scatter plot matrix (**Figure 3-5**).

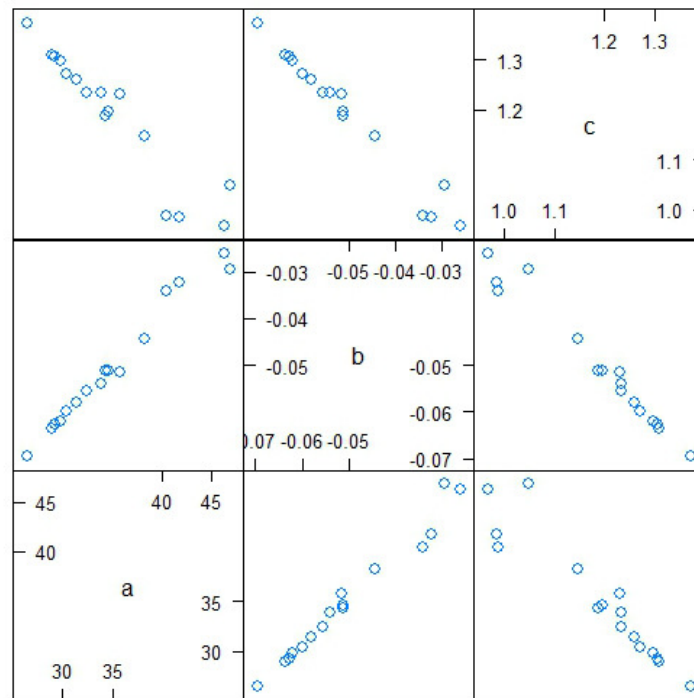


Figure 3-5: Scatter plot matrix of the relationship between three parameters of the Chapman Richards equation for a model fit, showing significant correlation between the parameters.

From **Figure 3-5** it is clear that with significant correlations between coefficients not all have to be included in the random effects – this pattern was consistent for all of the plots. However, there is no clear methodology to determine which should be included without testing the models first by alternating the coefficients chosen for random effects, as explained by Pinheiro and Bates (2000).

Appendix A summarises the results from the Weza dataset – which was used as a sample to decide on which parameters to include in the random effects, an example of one site is shown in **Table 3-1**. Model 4 in Appendix A, with only the asymptote as a random effect,

provided the most consistent results, with the resulting models providing improved Akaike Information Criterion (AIC) values, often significantly lower compared to the models using the other parameters as random effects. The counts of the number of times a model, with its corresponding random effect, was significantly the best performing model tested on the Weza dataset is summarised in **Table 3-2**.

Table 3-1: Random effect tests for the nonlinear mixed effects model. The column “parameter chosen” indicates which parameters used as an additional random effect in the model.

Parameter chosen (as random effects)	Convergence (Y,N)	df	AIC	BIC	Test	p-value
a,b,c	N					
a,b	Y	9	269.5768	295.922		
a,c	Y	9	269.5766	295.9219		
a	Y	7	265.5755	286.0662		
b,c	Y	9	272.0675	298.4128		
b	Y	7	268.0678	288.5586	6 vs 5	0.9999
c	Y	7	269.1501	289.6409		

Table 3-2: Count of best fit random effects models

Random effects models	
Parameters Included	No. of Lowest AIC values
a,b,c	0
a,b	0
a,c	1
a	7
b,c	0
b	3
c	5

As shown in Table 3-2, including the asymptote (α) only as the random effect was resulted in improved model fits over a number of sites. Thus the asymptote was used as the standard random effect for all further NLME models, which were tested against the NLS models, as the benchmark, and the GNLS models as an alternative.

3.2.8 Model comparisons

Diagnostics for nonlinear regression is a contentious issue as nonlinear regression curves are fixed to the definition of the model, and thus drawing meaningful conclusions for the quality of the fit (such as the R^2 value) is difficult. In this case a straightforward approach was taken, by simply analysing the visual structure of the model and the significance of the parameters. Residual plots for every model type and site were compared – these give a good idea on how the data is scattered around the mean curve, with priority given to homoscedastic residual plots.

Different models were then compared with the Akaike Information Criterion (AIC) described by Akaike (1972), which can be used to compare models with different covariance parameters as fixed and random effects (Fang and Bailey 2001). The root mean square error (RMSE) was also used as an indication of model fit.

A comparison of all of the main types of residuals and improvements is shown in **Figure 3-6** and **Table 3-3** below: what was noticed from the summary of the models (**Table 3-4**) is a depreciation of the asymptote compared to the NLS fitting method – although small.

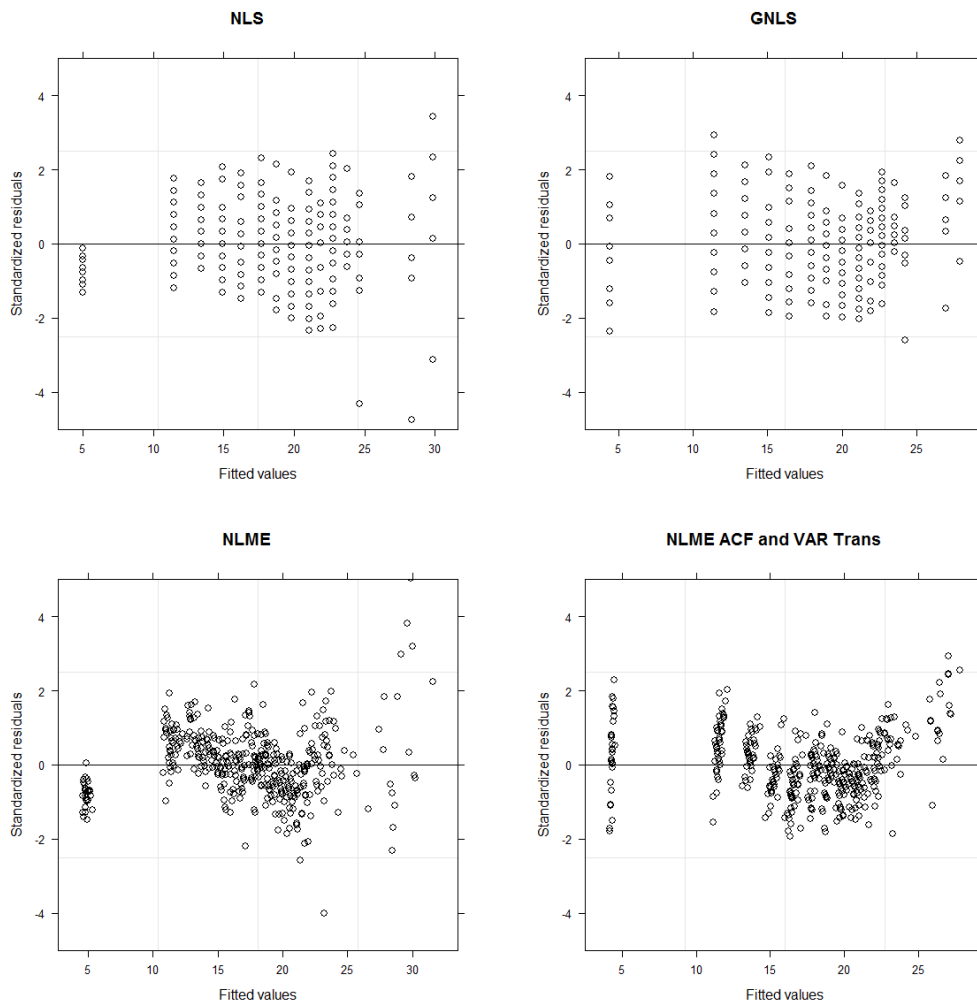


Figure 3-6: Residual plots of an example dataset using different modelling techniques

Clearly the NLME models are superior in both the results of the residual plots as the residual plots are more staggered, with less heteroscedasticity and a smaller unexplained variance. ACF and weighting significantly improve the model output, as can be seen from the conversion of NLS to GNLS, and the standard NLME model to the variance weighting and autocorrelated error inclusion to the NLME model.

Anova comparison and RMSE

Comparison between the different models was performed using ANOVA, where the AIC and RMSE were calculated (**Table 3-3**). While often Mean Squared error is used for model evaluation which includes bias and precision elements (Seifert and Seifert 2014) the Root Mean Square Error (RMSE) provides the advantage to be in the same unit as the modelled variable. It is an aggregation of the residuals obtained from a predicted model and is a good measure of how well a model fits overall, especially when compared to other models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}}$$

Equation 3-3

From the NLME, GNLS, and NLS model fits, an Anova comparison was tested, looking closely at the AIC and RMSE values as a measure of comparison.

Table 3-3: Anova and RMSE comparisons of different modelling fitting methods on the Weza dataset

Plot (spha)	Model	df	AIC	BIC	Test	p-value	RMSE
124	GNLS	6	47.65	54.46			0.9131529
	NLME	7	51.59	59.54	1 vs 2	0.1638	0.8415766
	NLS	4	69.07	73.61	2 vs 3	<.0001	0.9127091
247	GNLS	6	102.05	113.40			0.6671696
	NLME	7	99.86	113.10	1 vs 2	0.0406	0.6672069
	NLS	4	106.36	113.93	2 vs 3	0.0058	0.6567917
371	GNLS	5	188.10	199.75			0.8037401
	NLME	7	148.52	164.83	1 vs 2	<.0001	0.8151494
	NLS	4	190.21	199.53	2 vs 3	<.0001	0.8023609
494	GNLS	6	241.99	256.93			1.155355
	NLME	7	165.12	182.54	1 vs 2	<.0001	0.958164
	NLS	4	286.11	296.06	2 vs 3	<.0001	1.154249
741	GNLS	6	389.17	406.74			1.048519
	NLME	7	265.58	286.07	1 vs 2	<.0001	0.6787632
	NLS	4	412.69	424.40	2 vs 3	<.0001	1.048472
988	GNLS	5	462.56	478.15			1.066507
	NLME	7	321.67	343.50	1 vs 2	<.0001	0.7933879
	NLS	4	502.45	514.92	2 vs 3	<.0001	1.063379
1483	GNLS	6	589.71	610.36			0.9901197
	NLME	7	488.85	512.95	1 vs 2	<.0001	1.245127
	NLS	4	630.70	644.47	2 vs 3	<.0001	0.9313569
2965	GNLS	6	872.70	896.18			0.8478684
	NLME	7	574.83	602.22	1 vs 2	<.0001	0.5896667
	NLS	4	935.20	950.85	2 vs 3	<.0001	0.8470756

According to the AIC values, besides one plot, the NLME fitted model outperformed the other model types with p-values all below 0.05, suggesting a significantly improved fit on all of the plots. Furthermore, it most resulted in lower RMSE values, indicating an overall improved fit.

3.2.9 NLME final results (model selection)

The summary of the NLME models for all of the four sites is represented in *Table 3-4*; with illustrated curves in **Figure 3-7**. What is noticeable is that the site index is not strictly sensitive to the stems per hectare (as opposed to mean height), there is a general tendency

of a density effect on site index (**Figures 3-7** and **3-10**). This tendency must be taken into consideration when site index is used for prediction of growth. However, as previously stated, each plot was treated as an individual with its corresponding potential height (calculated in step two), which are explained in the third section of this chapter results.

The NLME model was finally chosen for model fitting as it provided a better fit for the relevant data range (50 years). However, with the associated problems using the asymptote as a random effect, to achieve an improved fit for the entire data range, extrapolation beyond 50 years may be compromised as the asymptote seems to be lower than the other fitting methods used. It was decided that, seeing as it is unlikely that plantations will be grown for such long periods for commercial purposes, this would be an acceptable compromise.

The overestimation of the model at young ages, which is most likely an effect of the inflexible Chapman-Richards equation was not solved by the NLME. However, to remain compatible to SILVA the compromise of applying the Chapman-Richards equation was made.

Table 3-4: Fitted SI-age models for the CCT trials

Density		Weza					Mac Mac					Dukuduku					Kwambonambi				
		Value	Std. Error	DF	t-value	p-value	Value	Std. Error	DF	t-value	p-value	Value	Std. Error	DF	t-value	p-value	Value	Std. Error	DF	t-value	p-value
124	a	117.65	74.87	17	1.57	0.1345	45.53	1.39	80	32.85	0	25.11	1.30	45	19.26	0	28.53	1.93	47	14.78	0
	b	-0.01	0.01	17	-1.02	0.3216	-0.04	0.00	80	-12.78	0	-0.06	0.01	45	-6.47	0	-0.06	0.01	47	-6.66	0
	c	0.83	0.06	17	13.62	0	1.18	0.03	80	37.19	0	1.01	0.09	45	11.69	0	1.09	0.07	47	16.34	0
247	a	42.38	3.82	41	11.08	0	37.64	1.50	91	25.07	0	28.23	1.21	80	23.28	0	48.98	5.01	74	9.78	0
	b	-0.03	0.01	41	-4.58	0	-0.07	0.01	91	-11.58	0	-0.06	0.01	80	-8.67	0	-0.02	0.00	74	-5.15	0
	c	0.96	0.08	41	11.56	0	1.43	0.05	91	26.09	0	1.12	0.07	80	16.94	0	0.89	0.04	74	24.74	0
371	a	41.10	2.61	62	15.74	0	38.37	0.89	113	43.24	0	29.46	1.23	114	24.00	0	42.50	1.96	91	21.64	0
	b	-0.03	0.01	62	-6.04	0	-0.06	0.00	113	-16.99	0	-0.04	0.00	114	-9.16	0	-0.03	0.00	91	-9.34	0
	c	1.02	0.08	62	13.24	0	1.37	0.04	113	35.87	0	0.94	0.04	114	21.16	0	0.89	0.03	91	30.06	0
494	a	34.62	1.57	73	22.06	0	41.46	1.80	127	23.05	0	28.75	1.11	136	25.99	0	39.85	1.51	117	26.40	0
	b	-0.05	0.01	73	-8.84	0	-0.05	0.00	127	-11.59	0	-0.04	0.00	136	-10.03	0	-0.03	0.00	117	-11.05	0
	c	1.15	0.07	73	17.63	0	1.23	0.04	127	32.01	0	0.95	0.04	136	23.18	0	0.92	0.03	117	31.40	0
741	a	38.08	1.49	117	25.64	0	42.26	2.13	151	19.87	0	30.66	1.15	159	26.58	0	50.39	3.31	155	15.22	0
	b	-0.04	0.00	117	-10.25	0	-0.05	0.00	151	-11.14	0	-0.03	0.00	159	-9.21	0	-0.02	0.00	155	-7.24	0
	c	1.02	0.04	117	22.95	0	1.26	0.04	151	33.17	0	0.83	0.04	159	22.37	0	0.84	0.03	155	30.49	0
988	a	45.36	3.34	143	13.58	0	45.17	2.35	184	19.23	0	29.18	0.72	228	40.67	0	47.27	2.26	181	20.90	0
	b	-0.03	0.00	143	-6.60	0	-0.05	0.00	184	-11.24	0	-0.04	0.00	228	-13.04	0	-0.02	0.00	181	-9.55	0
	c	0.93	0.04	143	24.25	0	1.24	0.04	184	34.78	0	0.96	0.04	228	23.14	0	0.85	0.02	181	36.12	0
1483	a	30.74	0.77	182	39.76	0	41.94	1.36	268	30.88	0	32.43	1.39	281	23.41	0	42.93	1.75	246	24.54	0
	b	-0.06	0.00	182	-21.78	0	-0.05	0.00	268	-16.16	0	-0.02	0.00	281	-8.85	0	-0.02	0.00	246	-10.89	0
	c	1.46	0.03	182	49.19	0	1.19	0.03	268	43.75	0	0.79	0.03	281	27.28	0	0.86	0.02	246	38.23	0
2965	a	28.437	0.58596	362	48.53	0	42.37	0.97	491	43.85	0	27.28	0.61	518	44.82	0	34.63	1.15	388	30.06	0
	b	-0.0599	0.00217	362	-27.63	0	-0.05	0.00	491	-24.81	0	-0.05	0.00	518	-17.36	0	-0.04	0.00	388	-13.82	0
	c	1.36917	0.0181	362	75.66	0	1.19	0.02	491	73.10	0	1.02	0.03	518	36.26	0	0.91	0.02	388	46.58	0

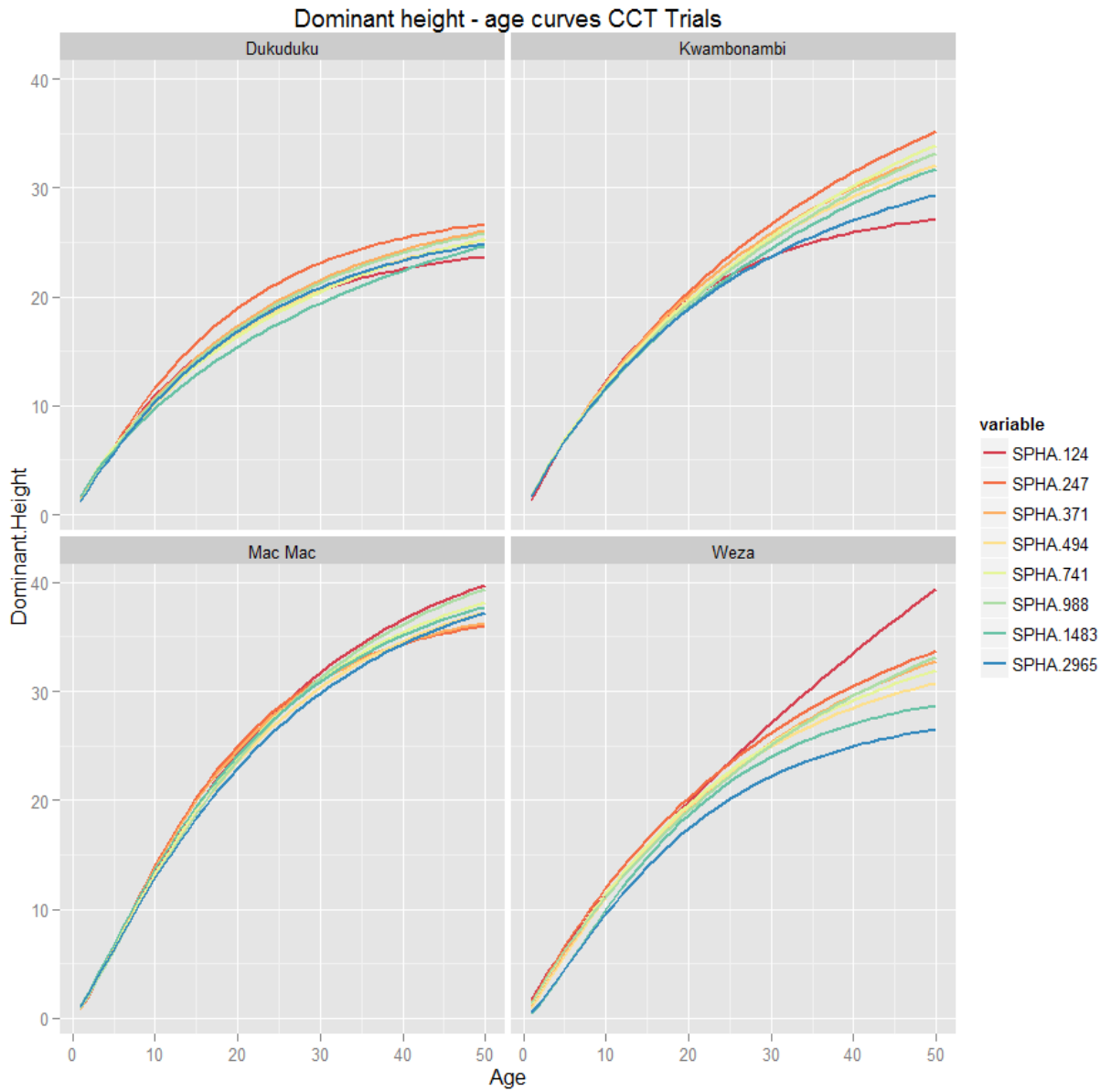


Figure 3-7: Fitted dominant height curves for the CCT trials

3.3 Step 2: Potential height modelling

Nonlinear quantile regression is one possible method to quantify potential height growth. Quantile regression was introduced by Koenker and Bassett (1978) as a statistical set of methods for defining conditional means in data analysis. Another option would be to subset the data into quantiles and perform some type of least squares estimation to achieve a nonlinear fit for that subset; however quantile regression was preferred as dividing the data into subsets and achieving a mean, instead of median response delivers differing values; more eloquently described in Koenker and Hallock (2001).

3.3.1 Description of the NLRQ procedure

Quantile regression is a generalisation of the median regression. While the latter was introduced as a robust method for dealing with outliers, which used the median instead of the arithmetic mean to fit a regression curve, quantile regression is able to use any quantile (denominated by the “tau” value) of the distribution to fit a regression curve to it (Cade and Noon 2003). Thus it is an appropriate technique to fit potentials. Nonlinear quantile regression was fit using the *nlrq* function in the *quantreg* package in R (Koenker 2006).

In this study, tau vectors of 0.9, 0.95, 0.975 and 0.99 quantiles (which represent the conditional quantile fits, e.g. 0.5 represents the 50% - or median fit) were used to determine the sensitivity of the nonlinear quantile regression on the parameterisation data. As a starting point for the potential modifier approach, the maximum possible definition must first be determined as the modifier only reduces the potential height value (Pretzsch 2009).

After careful consideration and visual inspection, it was decided to use the 0.975 quantile for the potential height definition. While it does not achieve the maximum values for height growth, it seems more flexible than the 0.99 tau vector and does not exclude too much of the maximum potential height series. The effects of these choices are discussed further on. Nonlinear quantile regression proved to be a robust and easy method to obtain potential height-age growth series. By specifying the 0.975 quantile the fits were more consistent. The potential height fits are shown in **Figure 3-8** below, fitted over their respective spacing trial series for each plot (spha). The results per plot are summarised in **Table 3-5**.

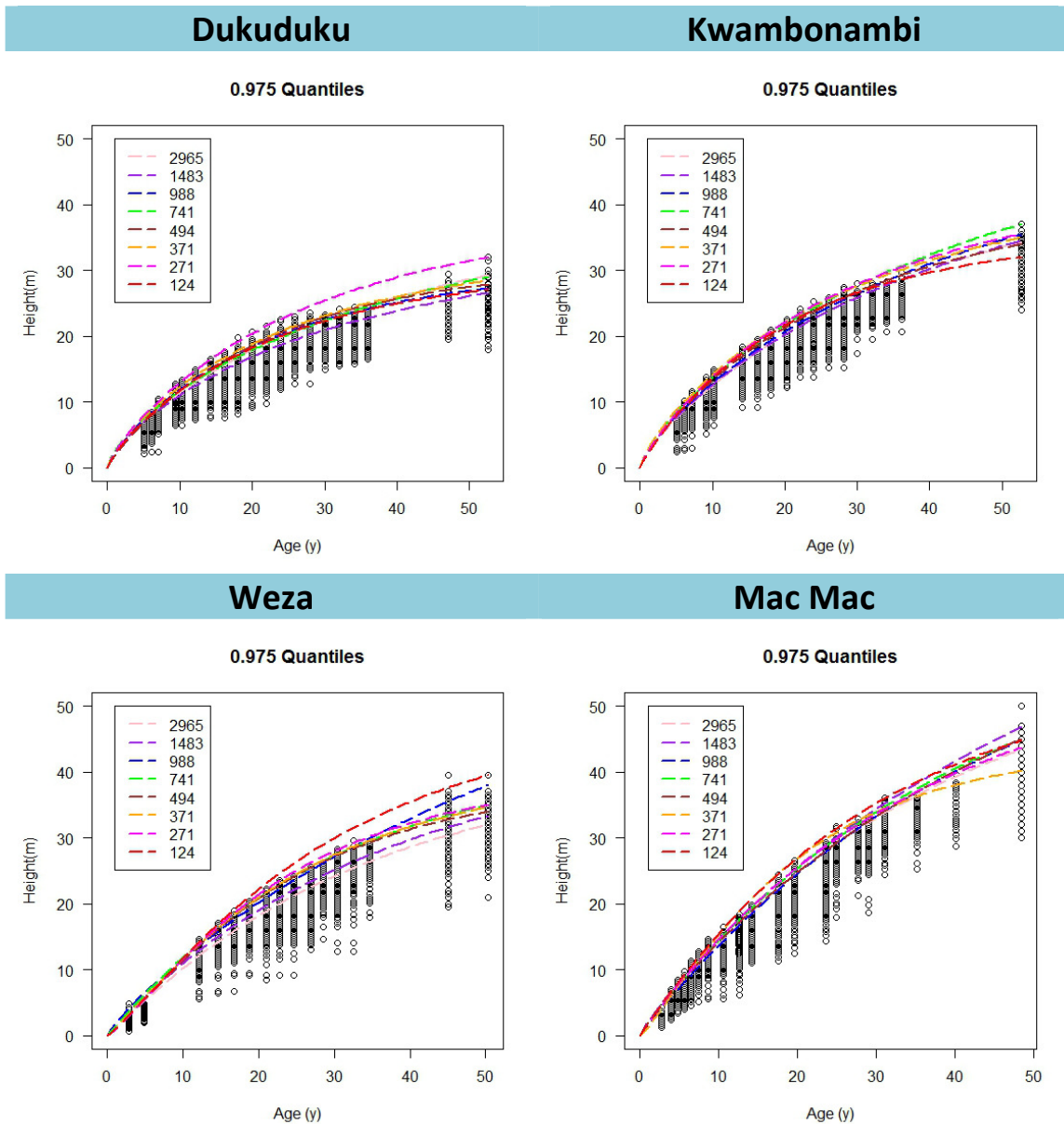


Figure 3-8: Fitted potential height curves for the CCT dataset for different stems per hectare.

From simply observing the growth curve series – a trend appears (seen within each trial series) that the 0.975 quantiles seem to be affected by density only to a small degree, although the extreme high and low ranges of stand density (e.g. 124 and 2965 spha plots) tend to dominate the upper and lower bounds respectively. However the difference between sites does not seem to overtly density dependent, but may be more strongly dependent on the site quality or location of the plots, suggesting that a site quality measure may be useful for predicting potential height.

In order to see whether these potential heights are representative, visual assessment was done on the PSP dataset. **Figure 3-9** below shows that for the best site in the CCT trials (the Mac Mac lowest density plot); the potential captures the upper bound of the PSP height series quite well, besides a few outliers, suggesting that the potential models fitted on the spacing trials represent the upper boundary of observed potentials.

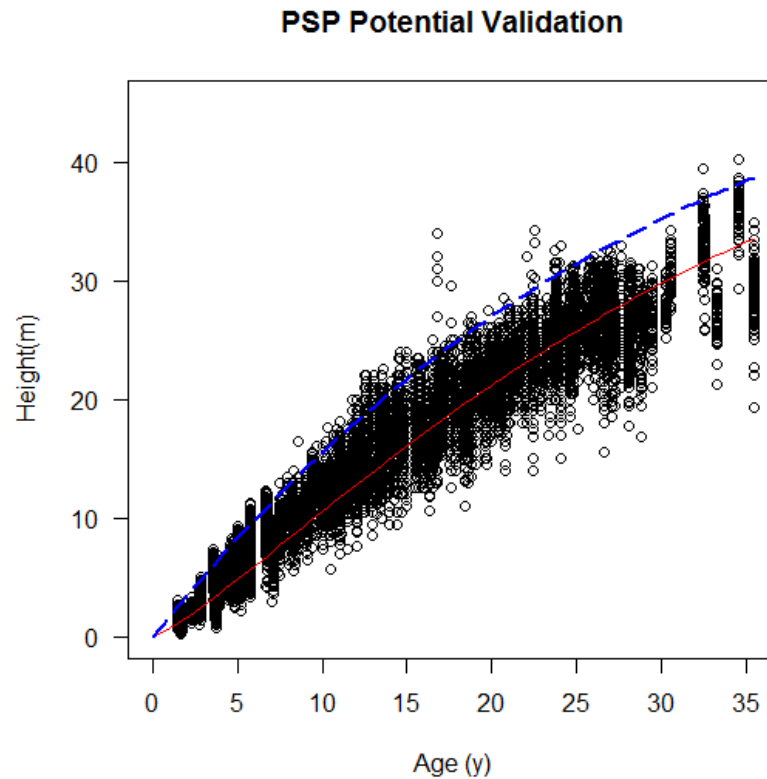


Figure 3-9: Potential height of the highest CCT trial value (Mac, spha = 124) plotted over PSP data of a wide range of sites, the blue line is the potential of the CCT plot, the red line represents the mean height of the PSP dataset

Table 3-5: Potential height fits by plot for the four spacing trials

Density		Weza				Dukuduku				Mac Mac				Kwambonambi			
		Value	Std. Error	t value	Pr(> t)	Value	Std. Error	t value	Pr(> t)	Value	Std. Error	t value	Pr(> t)	Value	Std. Error	t value	Pr(> t)
2965	a	41.26	2.81	14.67	0	39.27	2.69	14.61	0	59.23	5.61	10.55	0	46.16	3.95	11.68	0
	b	-0.03	0.00	-6.43	0	-0.02	0.00	-6.31	0	-0.03	0.00	-6.27	0	-0.02	0.00	-5.59	0
	c	1.04	0.06	17.10	0	0.74	0.02	30.87	0	1.02	0.03	32.40	0	0.82	0.03	29.51	0
1483	a	44.77	3.44	13.03	0	33.13	2.11	15.68	0	79.91	19.56	4.08	0.00005	48.15	4.79	10.05	0
	b	-0.03	0.00	-5.37	0	-0.03	0.00	-5.84	0	-0.02	0.01	-2.64	0.00848	-0.02	0.00	-4.68	0
	c	0.96	0.07	14.15	0	0.74	0.04	20.69	0	0.90	0.04	23.67	0	0.79	0.03	23.82	0
988	a	69.25	12.87	5.38	0	30.62	1.17	26.21	0	67.38	18.20	3.70	0.00022	50.97	3.20	15.90	0
	b	-0.01	0.00	-3.05	0.00232	-0.04	0.00	-8.06	0	-0.02	0.01	-2.27	0.0231	-0.02	0.00	-6.71	0
	c	0.87	0.04	20.19	0	0.83	0.05	17.32	0	0.98	0.07	14.01	0	0.77	0.03	28.58	0
741	a	42.50	3.43	12.38	0	39.74	3.33	11.95	0	64.78	7.03	9.21	0	53.54	3.76	14.26	0
	b	-0.04	0.01	-5.22	0	-0.02	0.00	-4.66	0	-0.02	0.00	-5.14	0	-0.02	0.00	-5.39	0
	c	1.04	0.08	12.60	0	0.70	0.04	18.38	0	0.97	0.04	23.99	0	0.75	0.03	22.23	0
494	a	38.33	0.81	47.46	0	30.82	1.98	15.53	0	73.13	18.22	4.01	0.00007	40.61	3.39	11.98	0
	b	-0.05	0.00	-20.46	0	-0.04	0.01	-4.78	0	-0.02	0.01	-2.42	0.01576	-0.03	0.01	-4.84	0
	c	1.24	0.03	37.49	0	0.92	0.09	10.66	0	0.92	0.05	16.99	0	0.86	0.06	15.33	0
371	a	40.09	1.25	32.15	0	32.65	1.56	20.98	0	43.65	1.41	30.89	0	42.72	2.83	15.10	0
	b	-0.04	0.00	-16.22	0	-0.03	0.01	-5.97	0	-0.06	0.01	-10.79	0	-0.03	0.01	-5.59	0
	c	1.16	0.03	33.76	0	0.79	0.06	14.07	0	1.24	0.06	20.62	0	0.79	0.05	16.81	0
247	a	39.38	1.05	37.64	0	38.08	2.26	16.84	0	58.28	6.25	9.32	0	42.72	3.13	13.66	0
	b	-0.05	0.00	-20.16	0	-0.03	0.01	-5.96	0	-0.03	0.01	-4.47	0.00001	-0.03	0.01	-5.22	0
	c	1.26	0.03	50.58	0	0.82	0.05	16.92	0	0.99	0.07	15.18	0	0.85	0.06	14.37	0
124	a	48.22	3.32	14.51	0	30.02	1.22	24.57	0	55.78	6.25	8.93	0	34.90	1.03	33.82	0
	b	-0.04	0.00	-9.58	0	-0.04	0.00	-9.05	0	-0.03	0.01	-3.19	0.00151	-0.05	0.00	-10.31	0
	c	1.21	0.03	37.05	0	0.86	0.04	20.64	0	1.03	0.11	9.30	0	0.94	0.04	21.22	0

3.4 Step 3: Prediction of potential height from Site Index

It was decided to use SI_{20} , the dominant stand height at the base age of 20 years, as a reference index. First the relationship of the potential height and site index curves was plotted and analysed after which the corresponding dominant height age 20 is used as the site index.

The equation used to describe this relationship is a Chapman Richards height-age function with site index as a factor (Payandeh 1974):

$$h_{pot} = a * (SI_{20})^b * (1 - e^{c*age})^{d*(SI_{20})^e} \quad \text{Equation 3-4}$$

h_{pot} is the potential height (m), SI_{20} the expected site index, age is measured in years, and $a1$, $a2$ and $a3$ are regression parameters. This model was chosen due to its parsimonious structure.

The model incorporates an effect of the site quality in SI_{20} , which was obtained from the predicted site index model of the different sites (Step 1). Thus the potential height development of the stand is modelled according to their respective SI_{20} values. This was done for each site first and finally for all of the sites combined. In order to validate the final model, first the potential height series predicted from dominant height was compared to the actual potential height in the CCT plots.

3.4.1 Effect of stand density

In the dataset used, stand density affected site index. **Figure 3-10** clearly illustrates this point with linear trendlines applied for each of the four sites used in the study, although some variations around the trend existed. This was possibly due to microsite and management differences of the plots. This was not significant to the overall performance of the predictive model, as each site index plot was used as an input for the parameterisation. It is included here as a possible pitfall of using measured site index that does not take into account changing the possible changing density in a following rotation; although the gradient is quite low for the range of commonly planted stand densities.

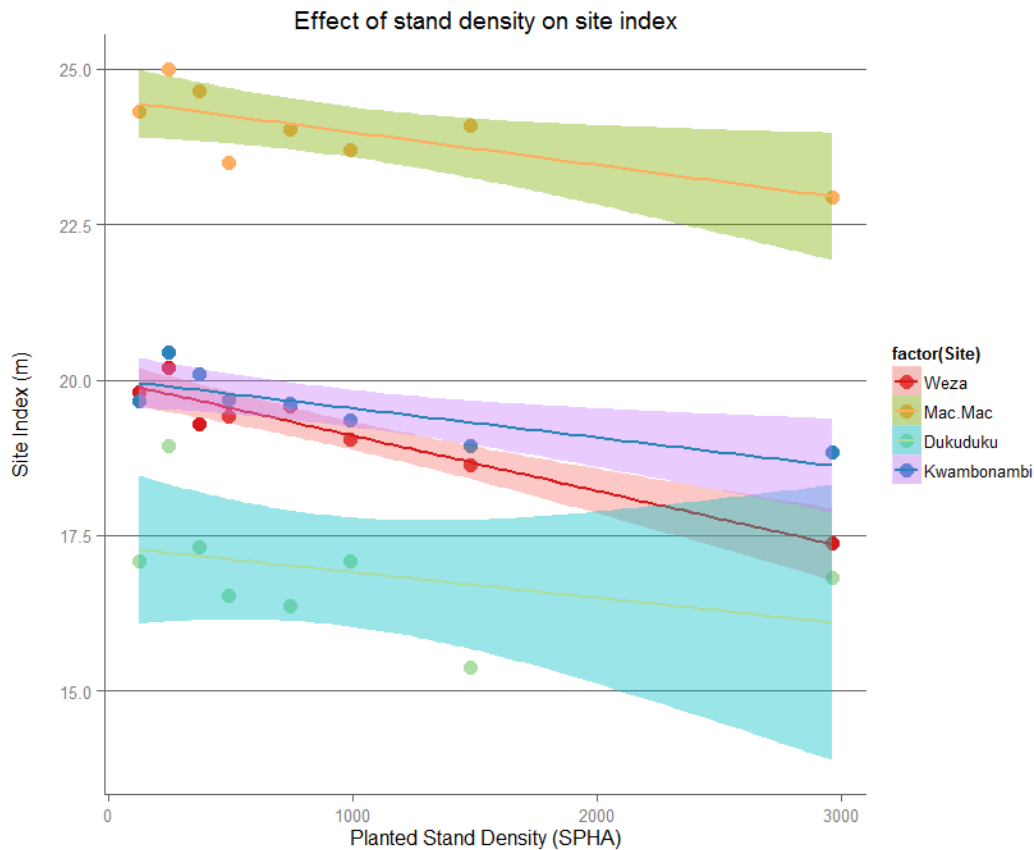


Figure 3-10: Site Index values on different stand densities for the CCT dataset

The 95% confidence bands are expectedly wide due to the low number of points. What is noticeable from Fig 3 – 10 is that each site has a different SI-density gradient and that the lines shift upwards or downwards dependent on the inherent site quality. This further emphasises the need for a site quality predictor based on edaphic conditions as proposed by (Esler 2012, Louw and Scholes 2002).

3.4.2 Relationship between potential height and dominant height

It is important to see what the relationship between dominant height and the fitted potential height on the same site over age is. As an example, **Figure 3-11** illustrates this for the Weza dataset for each of the plots.

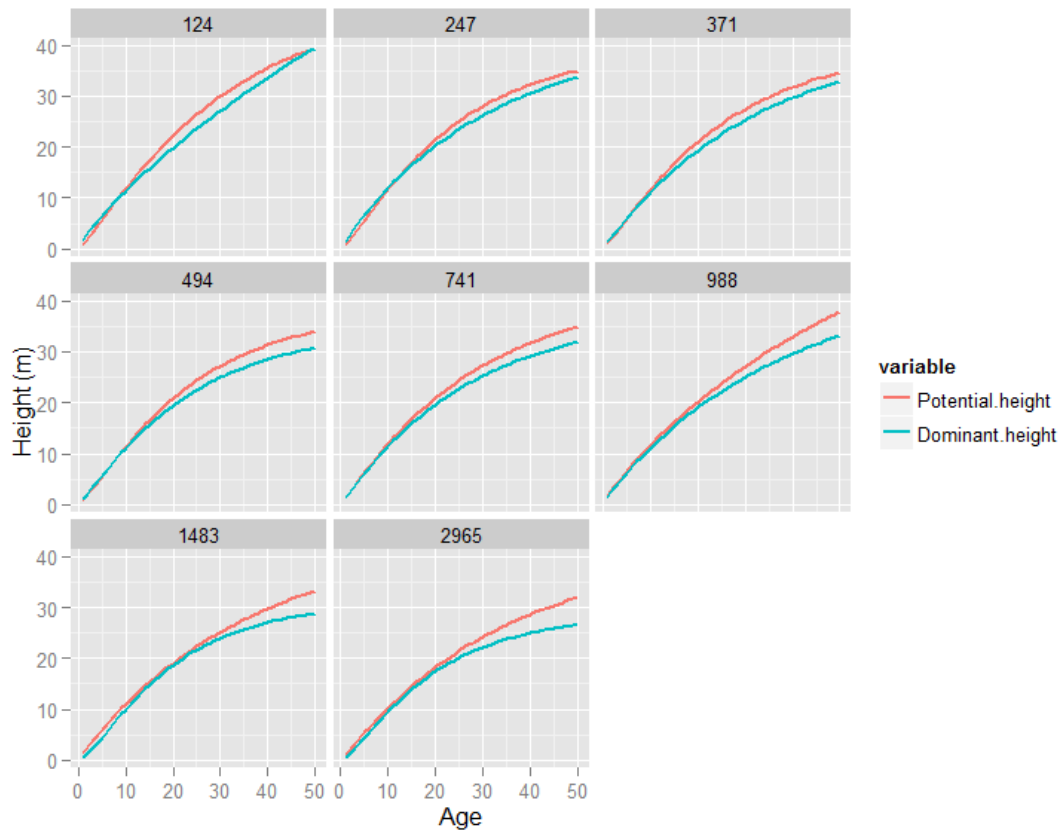


Figure 3-11: Example of potential and dominant height-age curves for the Weza trial

This shows how the difference between potential and dominant height increases over age. In order to assess this relationship potential height is plotted over dominant height for all of the four CCT trials in **Figure 3-12** below.

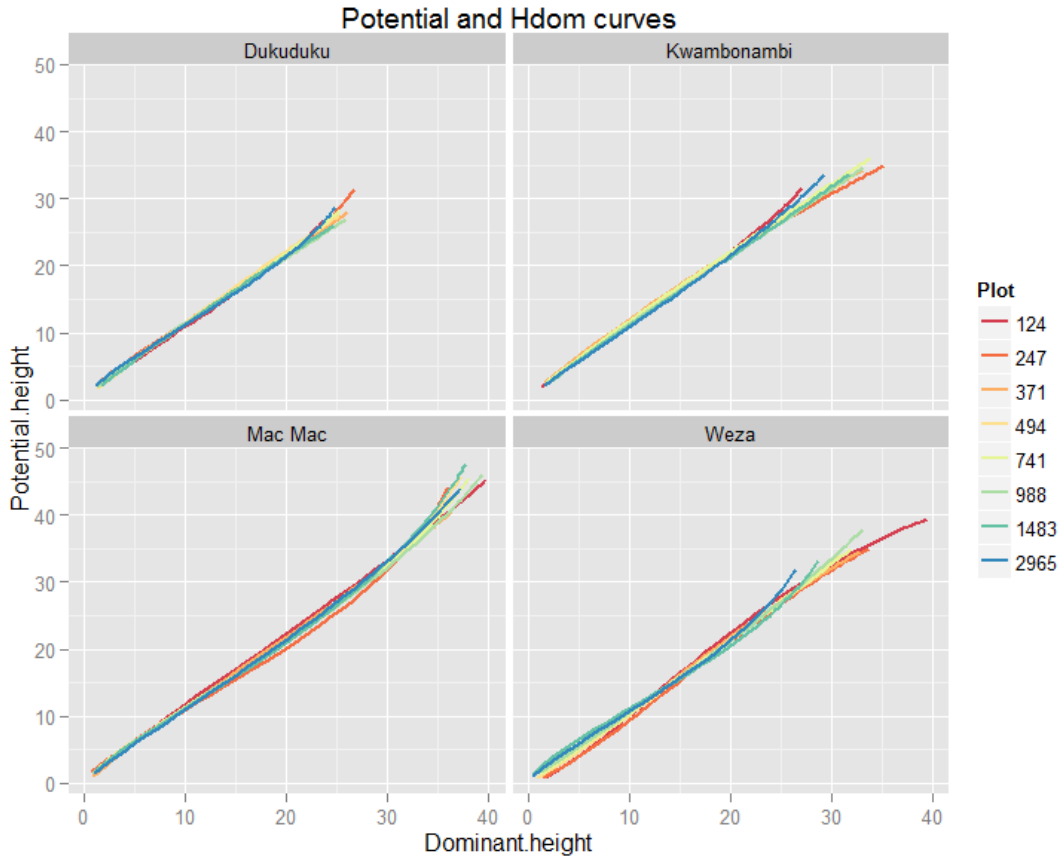


Figure 3-12: Potential height plotted over dominant height for all of the the CCT dataset plots

Figure 3-12 shows a generally linear relationship for the dominant and potential heights over age, although the values begin to vary as age, and consequently dominant height, increase. This gradient was modelled by a least squares linear fit, forcing the intercept through zero, which suggests about a 10% difference in potential and dominant height values over age (Table 3-6), i.e. the difference between potential and dominant height increases over age, as expected.

Table 3-6: Potential height - dominant height gradient

Site	Gradient	Std. Error	t value	Pr(> t)	R ²
Dukuduku	1.094623	0.001635	669.3	<2e-16	0.9991
Kwambonambi	1.124993	0.002466	456.2	<2e-16	0.9981
Weza	1.080868	0.001919	563.1	<2e-16	0.9987
Mac Mac	1.124993	0.002466	456.2	<2e-16	0.9981

This linear relationship and the relatively similar gradients suggest that a reference SI_{20} could be introduced as an added effect, as specified in Equation 3-4.

3.4.3 Predictive equation

Site index will change over two gradients: the site and, to a smaller degree, the density (a potential source of error in the model). The objective here is to obtain potential height models for a given site index (in this case, SI_{20}). The potential height over site index relationship is defined for all sites combined. It must be noted that the effect of density on site index, which was calculated on the South African top height definition, is not ideal, and a great improvement could be made by basing the site index on site factors mentioned in the previous section.

Final Combined Model

The four CCT spacing trials were pooled into the final model for potential height-age prediction from site index using **Equation 3-4**. Pooling all of the sites together will prevent the occurrence of extreme cases and provide a balanced model that is able to fit most cases sufficiently. **Figure 3-13** represents the final proposed model for prediction of potential height from site index.

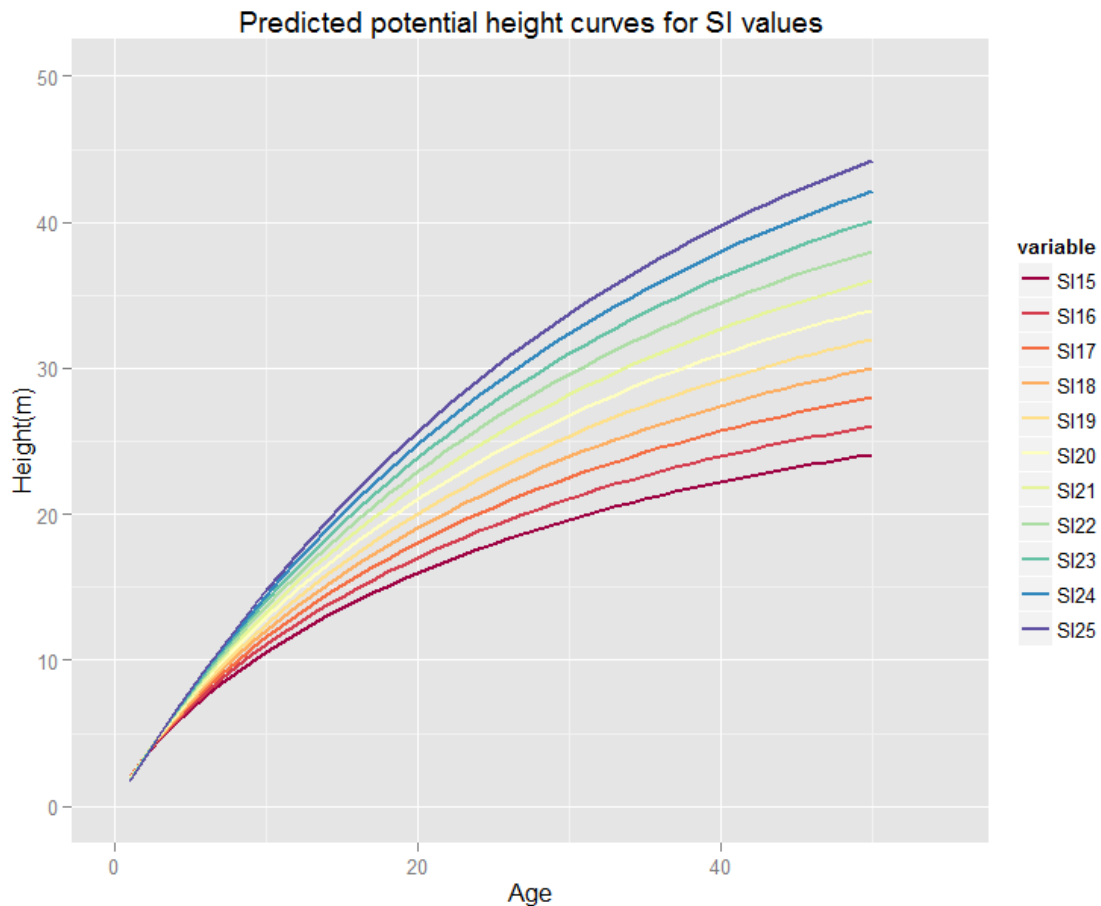


Figure 3-13: Final parameterised model of potential height-age using SI as a predictor

Table 3-7: Final model parameterised on the pooled CCT trial datasets from Equation 3-4

Parameter	Value	Std. Error	t value	Pr(> t)
a	0.8338	0.04237	19.67727	0
b	1.31306	0.01982	66.26476	0
c	-0.02965	0.00101	-29.3729	0
d	0.16613	0.00948	17.52046	0
e	0.55563	0.01795	30.94866	0

3.4.4 Validation

Observed vs. Predicted Potential height

It was decided to see how the potential height compared with the potential height predicted from the site index (in the pooled model). This was done by calculating the difference in predicted potential height compared to the observed potential height fitted for each of the spacing trial plots (**Fig 3-8, Table 3-5**). The models fits tested on the individual sites a relatively high error margin (over and under prediction) before age 10 (**Fig 3-15**). They begin to converge age 20 and then fan out again representing an increased error predicted for higher ages (**Fig 3-14**), although the values are small considering the actual height of the trees at those ages.

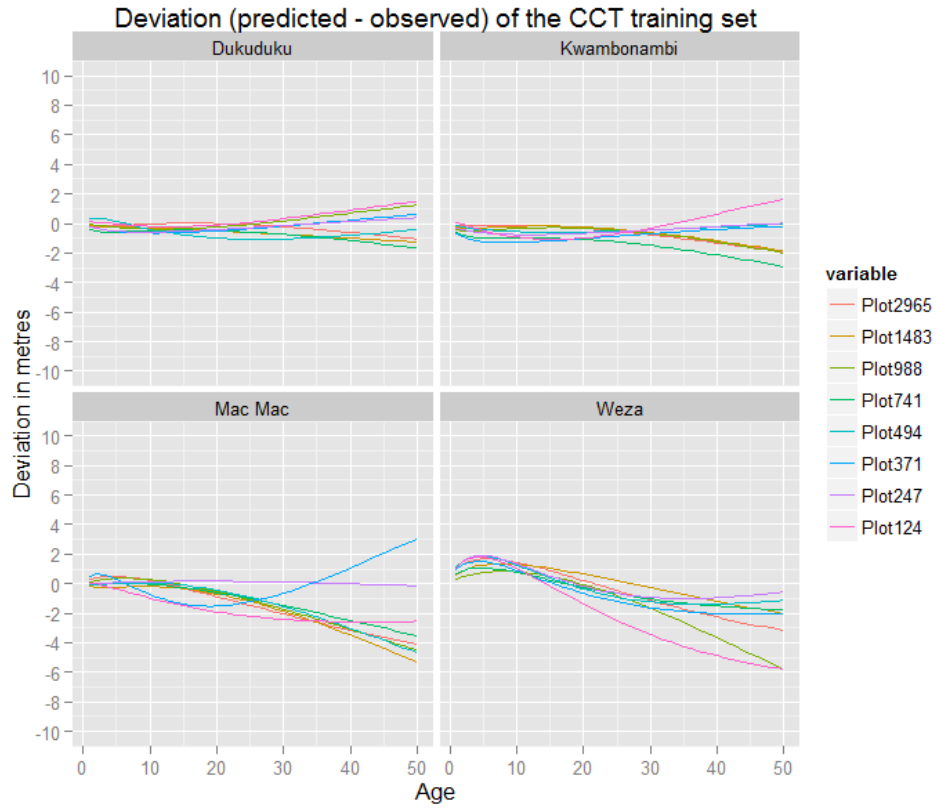


Figure 3-14: Deviation from the observed potential height-age compared to the potential height predicted equation

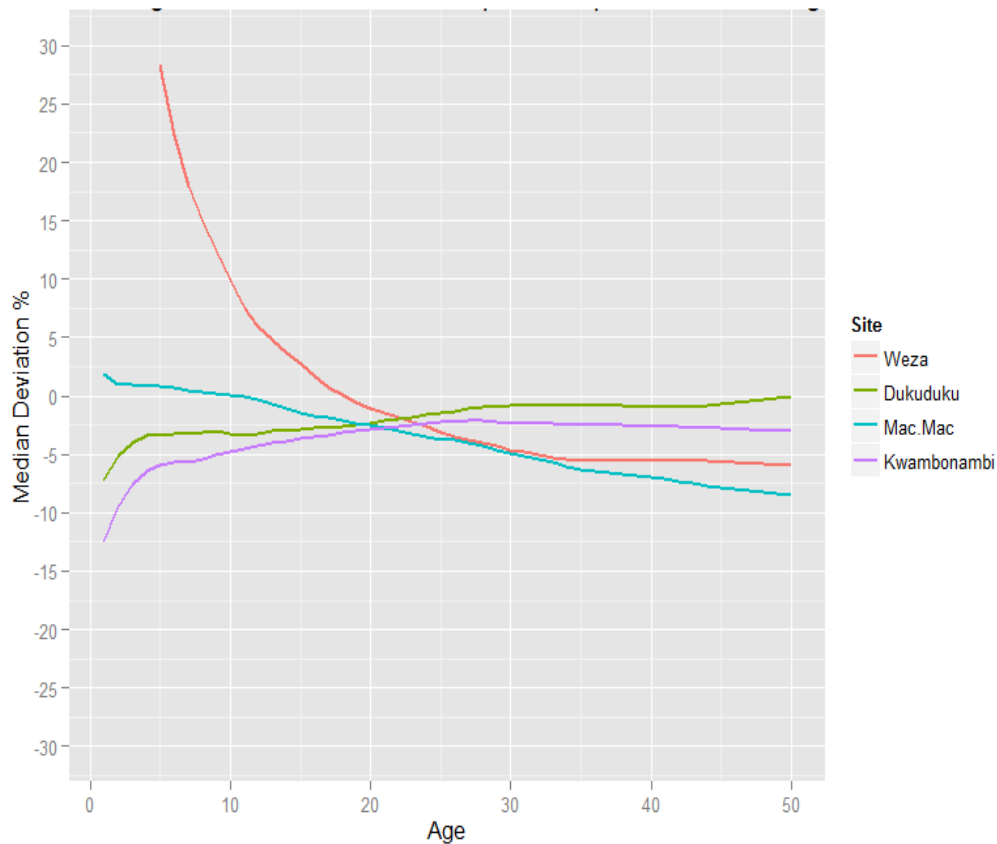


Figure 3-15: Median deviation in percentage of the observed and predicted potential height curves

Overall, considering the vastly differing site conditions, growing conditions and the different shape of the height-age growth curves, the combined model reacts well, and although seems to underpredict slightly at after age 20 (**Fig 3-15**), it seems to stabilise to some degree. It seems that site index can indeed be used as a predictor for potential height modelling, although the values will range at very young and very old ages. The high error before age 10 could also mean that this Chapman Richards based model may be unsuitable for pulpwood rotations, although as these are only potential height values it would first need to be determined how the final predicted model performs once a modifier is applied.

Referencing against independent PSP data

The above validation used sites in which the model was parameterised; however it is necessary to see how this model works for independent data. For this reason a few sites were selected from the PSP dataset which had a suitable number of re-measurements. From this an inspection of how the observed vs. the predicted potentials compare for sites of a given site index can be seen (**Fig 3-21**).

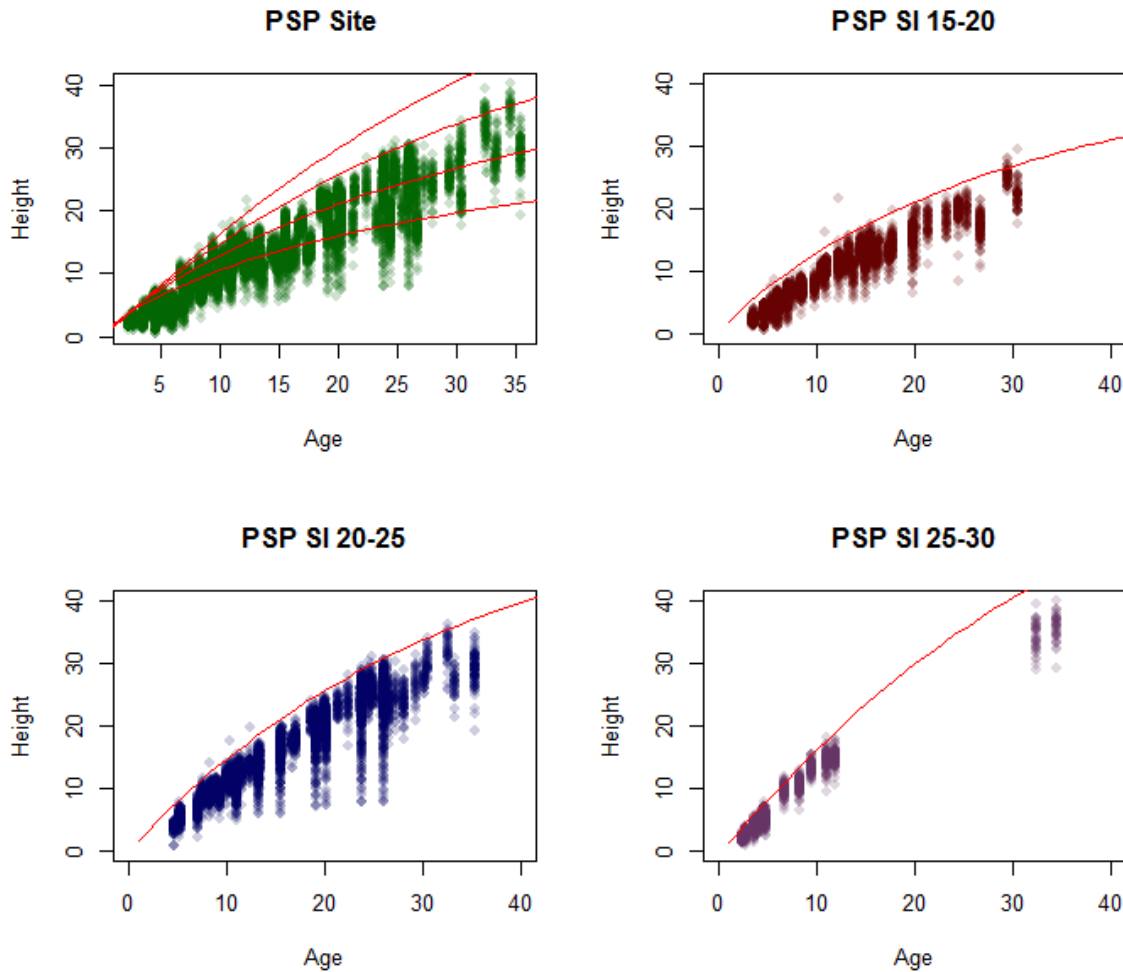


Figure 3-16: Potential height curves plotted over PSP data of different classes. The red lines represent the predicted potential height for the upper bound of the SI classes presented above. The top left image presents all of the data with 15, 20, 25 and 30 SI predicted potentials

In Figure 3-21, the predicted potential from site index seems to capture the upper bound of the PSP dataset, with the top left hand graph representing all of the sites, with lines of the predicted potentials for SI's of 15, 20, 25 and 30. Each of the other graphs represents SI classes within this dataset with the respective potential height curves predicted from SI of 20, 25, and 30 respectively.

To see how this performs on a few selected sites, five sites were selected from the second PSP series, where measured SI's were not available, and thus had to be calculated from the methodology proposed in this chapter. The SI values would then predict a potential curve which would be superimposed on the plot height-age data for visual inspection (**Fig 3-17**).

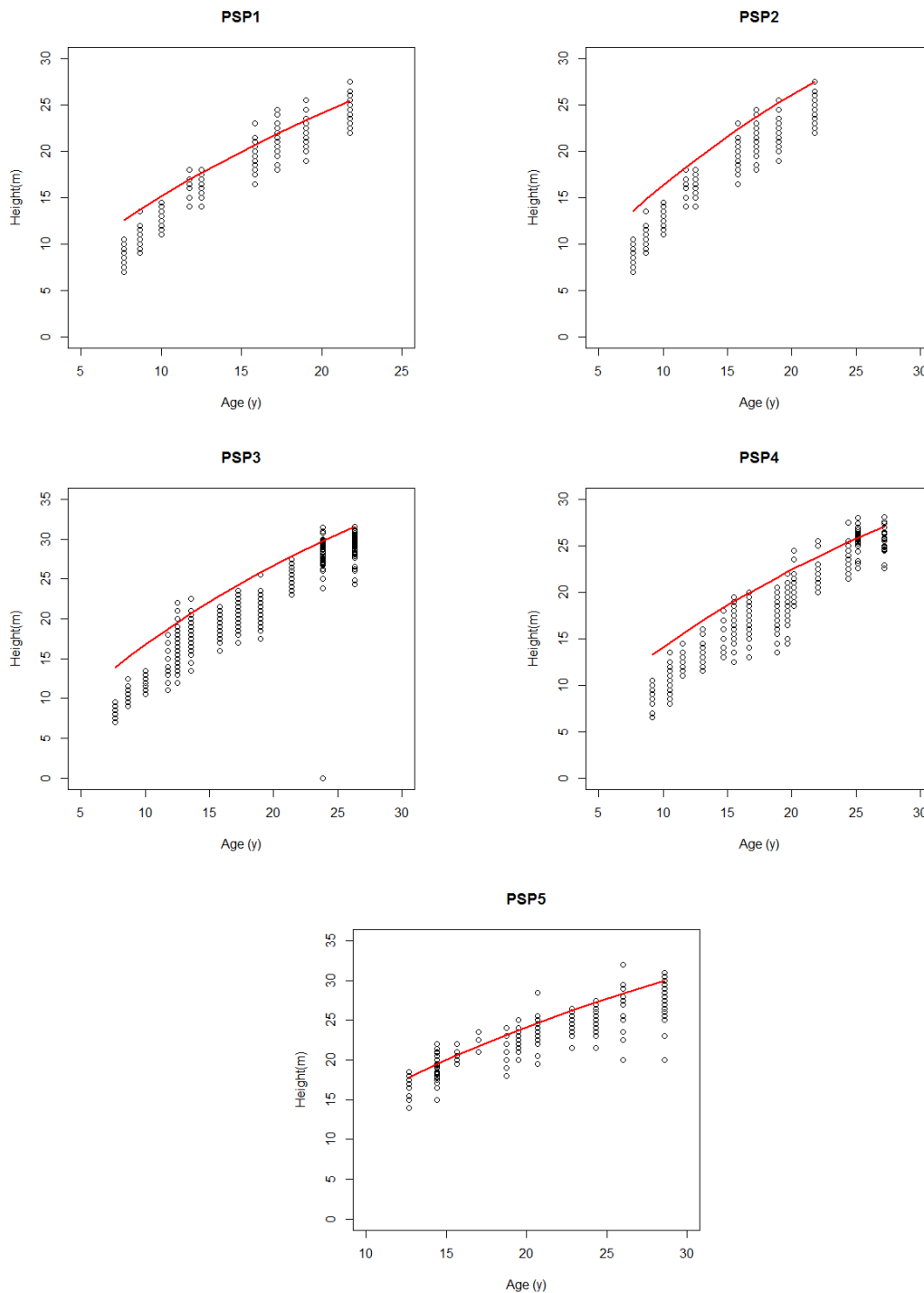


Figure 3-17: Fitted predicted potential height curves fitted on selected independent PSP data

The predicted potential seems to capture the potential heights of each site quite well, with only a few trees which fall above the predicted potential curves.

3.5 Chapter conclusion

The main objective for this chapter, split onto three steps, was to model potential height based on measured site index.

Site index models and potential height models were developed in order to develop a methodology to predict potential height from site index. The relationship between site index and potential height was shown to be a linear relationship, and a reference site index of SI_{20} was used to predict potential height from site index.

The use of nonlinear mixed effects modelling has proved a superior method for modelling dominant height-age curves with the associated clustered data structures and heteroscedasticity and it also represents a potential improvement for other height-age models. However, care must be taken to ensure that including the asymptote as a random effect does not inflate or deflate the asymptote implausibly – as it tends to change the asymptote to fit the observed data, which may affect extrapolation beyond the measured years in the parameterisation training set.

Nonlinear quantile regression proved to be a robust method of fitting observed height potentials. The model training set (CCT trials) covered the range of heights observed PSP data.

Potential height seems to be well correlated to a stand's site index. Visual validation of the results seem to suggest that site index can indeed be used to model the potential height, which covers the first step of the potential modifier method and simulation initialisation by deriving a height potential from an industry standard site index information.

Chapter 4: Modelling diameter increment in response to resource limitations and site classification

4.1 Introduction

The main purpose of this chapter is to create an age-independent diameter increment model, which predicts diameter increment from DBH and the competitive conditions of the tree at any point in time. **Figure 4-1** illustrates this relationship for all of the plots in the Mac Mac spacing trial, where the gradient for the increment-diameter relation flattens out (decreases) over time, with each scatter cloud representing a different measurement activity at a certain age. In the proposed approach, which is following the SILVA methodology, the age effect is substituted by tree size (DBH), while the competition interaction is supposed to cater for the change in increment-diameter relations over age.

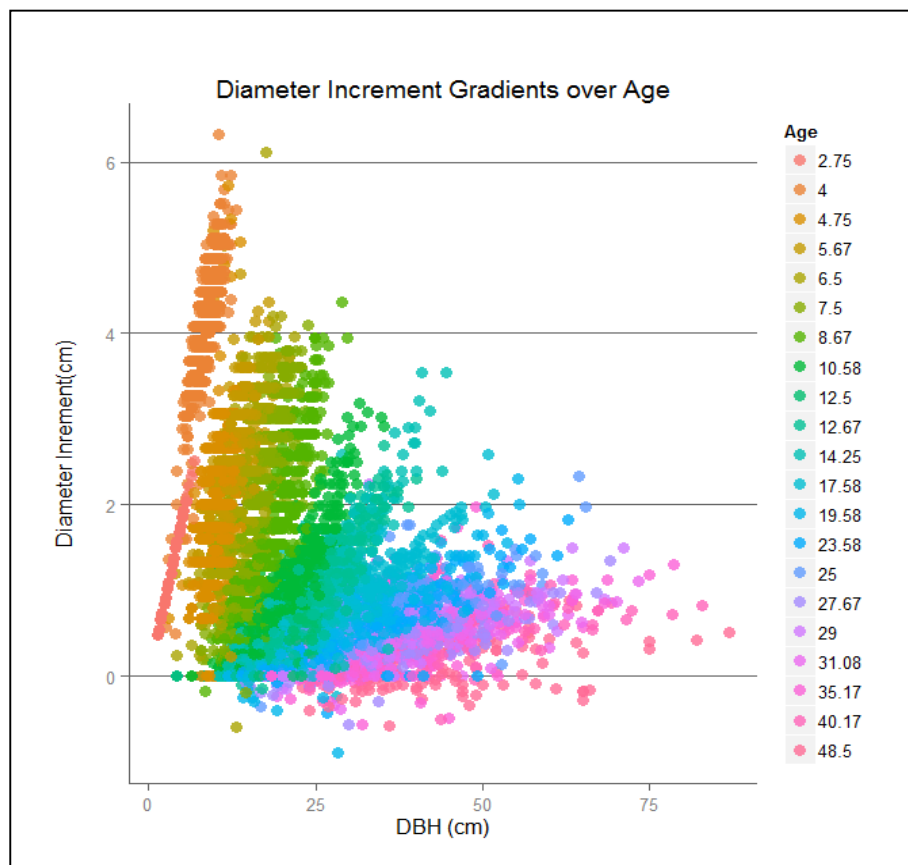


Figure 4-1: DBH Increment - DBH scatterplot on the Mac Mac CCT trial, showing decreasing linear gradient over age

Employing a potential-modifier methodology, two steps were thus required. First the potential increment for any given tree diameter had to be found. Second, the modifier, based on a competition index, had to reduce the increment and display a structure similar

to **Figure 4-1** above, whereby the gradient of diameter increment for a given DBH decreases.

Different competition indices are developed and describe growth differently according to their mathematical structure. For instance the KKL index (Pretzsch *et al.* 2002), with its search crown is intrinsically more suited to describing overtopping and thus competition in light limited environments. Other indices, based solely on distance between individuals or groups of individuals in an area, will be intrinsically more size symmetric, such as the Local Basal Area (LBA) index, described later (Seifert *et al.* in press).

Thus the sensitivity of diameter increment to an index can change according to site location and the resource limitation which is experienced as demonstrated by Seifert *et al.* (in press). For instance, competition index A might describe more of the variation in the predicted model than index B in a wet environment. This can also change according to the quality of the site. The performance of competition indices may also change according to age, density and longitudinal climate changes (wet and dry spells).

4.1.1 Dataset

The Nelder spacing design and the SSS-CCT design at Tweefontein were used for these purposes. It must be noted upfront that these are two very different designs – thus when making comparisons between the behaviour of the indices, some differences could be due to the different spacing design, observation, etc. However, these were the only trials where tree positions could be obtained and most of these problems should be negated by the nature of the competition indices.

4.1.2 Chapter outline

With only two sites from which to test competition with different water availability, this Chapter should be strictly seen as a methodological guideline for predicting diameter increment, instead of an investigation on changing competition mode with different resource limitations, although an indication of shift in importance of competition indices on different sites is seen (4.5.3). The chapter presents possible methodologies for future testing of changing competition modes and their effect on diameter Increment. The methodology was developed according to the following steps:

- Step 1: Classify sites according to water availability
- Step 2: Determine the potential based on site conditions
- Step 3: Fit multiple competition indices
- Step 4: Select competition indices
- Step 5: Use CI's in a deterministic potential modifier equation
- Step 6: Create a stochastic model incorporating natural variability

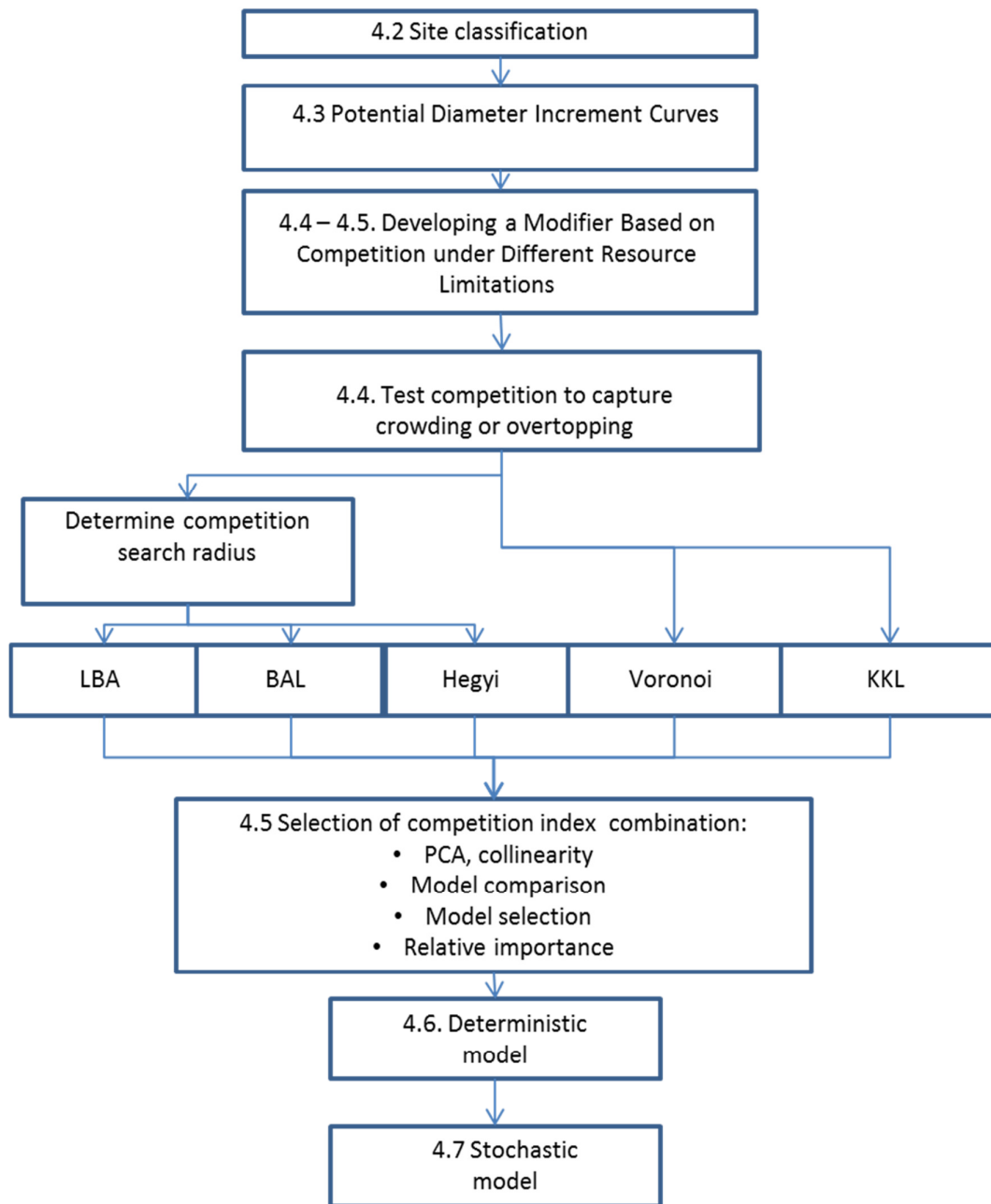


Figure 4-2: Flowchart of the chapter outline showing working steps of the methodological approach used in the study

4.2 Step 1: Site classification according water availability and site index

With the hypothesis in mind that the mode of competition shifts with the availability of edaphic factors, a simple categorisation of the sites according to water availability was sought, subject to the availability of data. The FAO – UNEP index (UNEP 1992) where aridity is calculated by dividing the precipitation by the potential evapotranspiration, as calculated by the Thornthwaite method (Thornthwaite 1948).

In this instance, the aridity index was fixed by using monthly weather data supplied by the Agricultural Research Council (ARC) climate database. This provides an added static classification for the site.

Defining the water index longitudinally (changing over time), either using this index or other candidate indices (McKee, Doesken, and Kleist 1995, Palmer 1965), would improve the interaction with diameter increment. Unfortunately measurements in the trials were not taken at annual intervals; often at 2-5 years, which would mask correlations.

Table 4-1: Classification of the sites according to the FAO-UNEP classification and SI. The Tweefontein and Mac Mac trials used the same weather station.

Location	Precipitation	Aridity index	Classification (FAO-UNEP)	Average SI
Lottering Nelder	950	1.106	Humid	21.2
*Tweefontein SSS-CCT	1222	0.86	Humid	23.0
Kwambonambi	1208	1.2	Humid	20.3
Dukuduku	960	0.76	Humid	17
Weza	927	0.69	Humid	20.3
*Mac Mac	1222	0.86	Humid	24.0

Table 4-1 shows this index for all of the spacing trials used in this thesis. For the aridity index, only two sites were relevant, the Nelder trial and the Tweefontein SSS-CCT trial, as these were the only available trials with tree positions necessary for modelling the effects of competition in this case. However the average site index (averaged over the SI values presented in Chapter 3) was used for the potential increment estimation in Section 4.3 and for the testing of the model in Sections 4.6 and 4.7.

With this in mind, a tentative linear model was drawn between the water index values of the two sites – accepting that more work would need to be done to fully incorporate water availability. This simple methodology would then be developed and tested for future experiments, using this simple water index as a pointer to determine whether there are any

functional relationships between the water availability and the competition indices in this model.

4.3 Step 2: Determine the potential based on site conditions

For the purpose of this study, the methodology proposed by Pretzsch and Biber (2010) was used to determine a workable simulation routine for diameter increment modelling. In this procedure the potential diameter growth/increment was first defined using the *nlrq* package in R for nonlinear quantile regression application (as done in Chapter 3), using the 0.99 tau values, where:

$$di_{pot} = a_0 d^{a_1} e^{-a_2 d} \quad \text{Equation 4-1}$$

Where di_{pot} is the predicted potential increment, d is the diameter (DBH) of the tree and a_0 , a_1 and a_2 are coefficient to be determined in the model. This represents the maximum potential diameter increment for a given DBH – while not a true potential, it is an adequate approximation for the purpose of this thesis.

4.3.1 Comparison of sites

Four sites were considered for potential increment parameterisation, the Mac Mac, Kwambonambi and Dukuduku CCT trials and the Nelder spacing trial. The Weza trial was not included as there was a large period of measurements gaps; the Tweefontein SSS-CCT trial was not included as it was too young. The potential curves fitted for the considered sites are presented in **Figure 4-3** below.

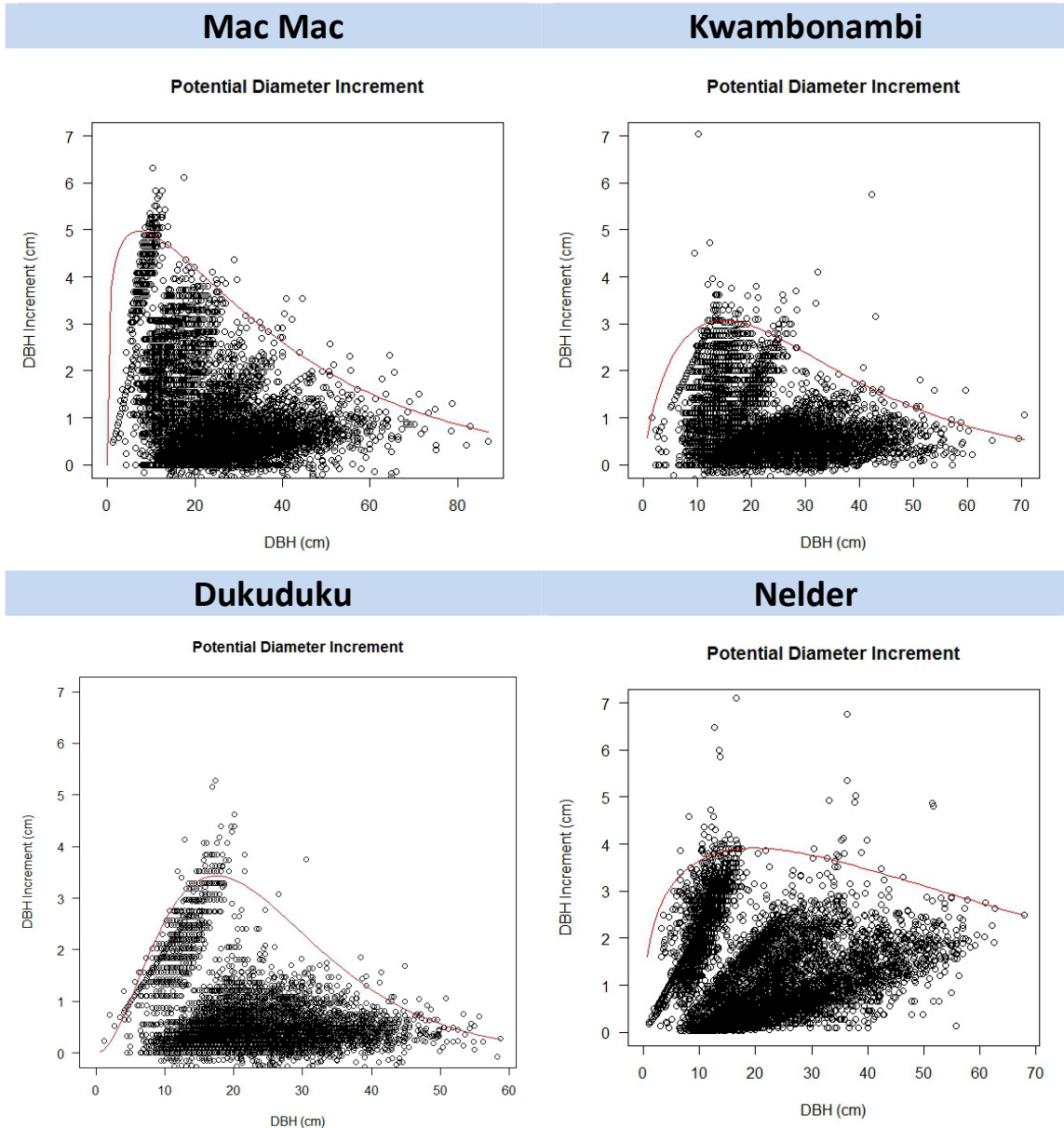


Figure 4-3: Fitted potential increment curves for the four sites considered for increment potential estimation

Including the Nelder trial to the model results in a much flatter curve which gives poor fit to the data upon visual inspection. This was because the Nelder trial was only measured until 26 years, which did not include the flattening of the Increment/DBH relationship, resulting in an interference with the gradient. Thus the results of the potential increment curves of the three CCT trials according to **Equation 4-1** are represented below.

Table 4-2: Potential increment model coefficients according to Equation 4-1.

Parameter	SI 17.0 Dukuduku		SI 20.3 Kwambonambi		SI 23.9 Mac Mac	
	Estimate	SE	Estimate	SE	Estimate	SE
a	0.0532	0.0103	0.8002	0.2241	3.9714	0.5858
b	2.2383	0.0973	0.7967	0.1337	0.2282	0.0799
c	0.128	0.0042	0.0539	0.0052	0.0316	0.0039

From **Figure 4-3** and **Table 4-2** it can be seen that site index had an effect on the determination of potential increment (as seen in the above figures). Thus dominant height – or h_{dom} - can be included as a predictor variable to **Equation 4-1**, where a site quality effect can be included (Pretzsch and Biber 2010) as was done in in Chapter 3, resulting in **Equation 4-2** below

$$id_{pot} = (a_1 + a_2 h_{dom_{20}}) d^{2.3} e^{-2.4d} \quad \text{Equation 4-2}$$

The resulting fit of the above model is tabulated in **Table 4-3** and the curves are illustrated in **Figure 4-4**.

Table 4-3: Coefficients of the potential increment model (Equation 2) using site index and dbh as predictor variables

	Combined Increment/ DBH model (Equation 4-2)			
	Value	Std. Error	t value	Pr(> t)
a_1	0.5372	0.27377	1.96221	0.04975
a_2	0.11434	0.01564	7.30886	0
a_3	0.34421	0.08628	3.9893	0.00007
a_4	0.03724	0.00413	9.01277	0

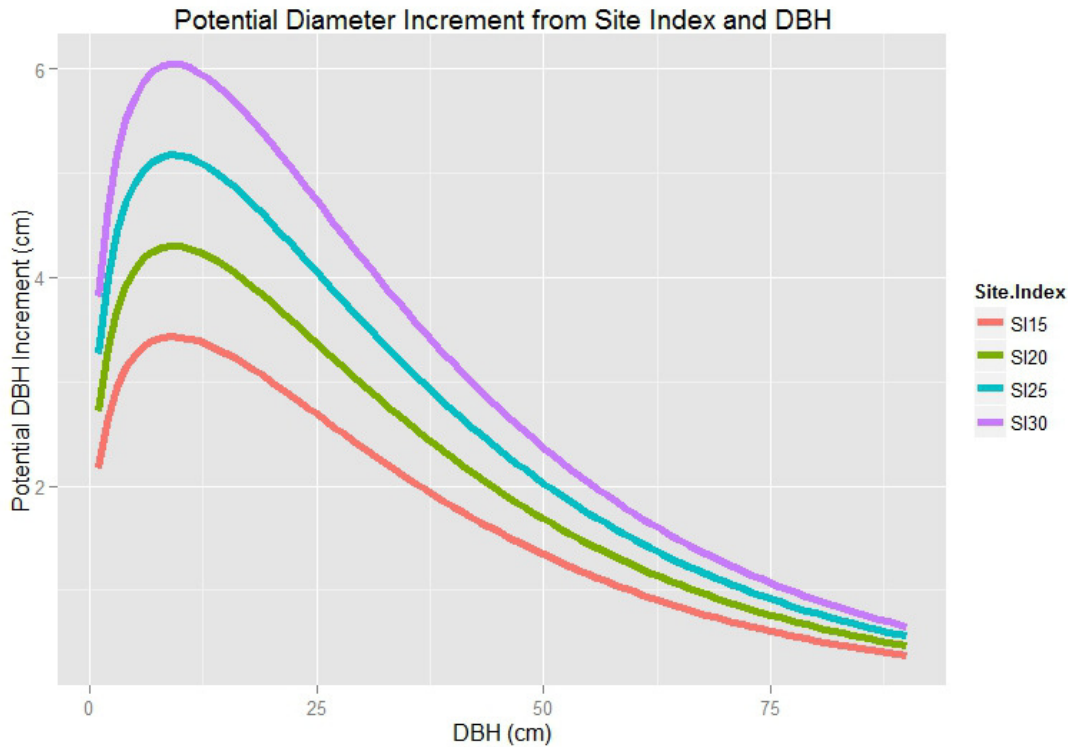


Figure 4-4: Potential increment over DBH curves parameterised from the CCT trial data; the different curves represent different SI values

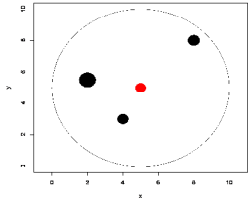
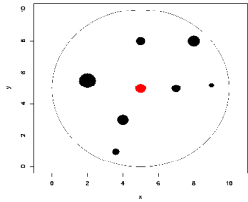
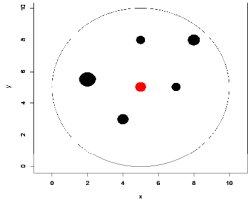
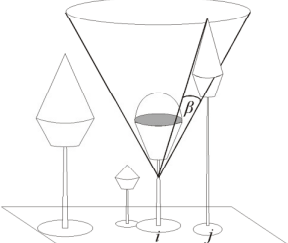
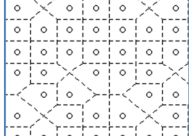
These potentials are the 99% quantiles representing the maximum annual diameter increment that a tree can obtain for a given DBH and site index. The estimated increment was then modelled by including a suitable competition index, discussed in the following sections, which should mimic the structure presented in **Figure 4-1**.

4.4 Step 3: Fit various competition models

4.4.1 Distance dependent competition indices

In this study five distance-dependent competition indices were used. These indices were chosen to capture a gradient, which would describe overtopping or local crowding, with the objective that two indices will be selected which both adequately describe competition and capture different modes of competition (Seifert *et al.* in review). These, summarised in **Table 4-4** and described below, are the KKL (which is used in SILVA), the local basal area (LBA), basal area of larger trees (BAL), Hegyi and Voronoi polygons.

Table 4-4: Illustration of the different competition indices used in this study (Seifert et al. in review), *i* refers to the central tree, *j* refers to the competitor trees. Models shown below are discussed in more detail in text below.

Name	Formula	Graph	Reference
Basal area of larger trees	$BAL_{ij} = \sum_{j=1}^n BA_j$ <p>, where $BA_i > BA_j$</p>		Wykoff et al. (1982)
Local basal area	$IBA = \sum_{j=1}^n BA_j$		Stenerker and Jarvis (1963)
Hegyi	$ITH = \sum_{j=1}^n \left(\frac{D_j}{D_i} \cdot \frac{1}{DIST_{ij}} \right)$		Hegyi (1974)
Crown Competition Index KKL	$KKL = \sum_{j=1}^n b_j \cdot \left(\frac{CCA_i}{CCA_j} \right) \cdot TM_j$ <p>, where $j \neq i$</p>		Pretzsch et al. (2002)
Voronoi	Division of growing space based on connection of bisecting line distances		(Brown 1965)

The KKL index (Pretzsch 2009) uses a search for competitors based on a search cone at an angle of 60° which starts at a point of 60% of the tree height. As most competition indices it works in a combination of competitor selection and the quantification of the competition effect of those identified competitors. Neighbouring trees, which fall inside the search cone, are included as competitors and the angle from the tip of the neighbouring tree and its maximum crown extension to the cone mantle (*angle β in Table 4-4*) is calculated. This is used as a measure of competition, which is then multiplied by the cross-sectional crown area of the competitor in relation to the subject tree (CCA_i/CCA_j). Species specific values for crown dimensions for *P. elliotii* were not available, for this reason the SILVA model was used to calculate the KKL for the stands using *Pinus sylvestris* crown model as a proxy. The KKL is by its design strongly focussing on competition induced by crown competition and overtopping.

The local basal area (LBA) is simply the sum of the basal area of all trees within a competition search radius and the reference tree located in the centre of the circle (Steneker and Jarvis 1963). It has been suggested as an effective measure of local crowding and is more sensitive to measuring edaphic limitations as it does not discriminate between the sizes of the trees included in the influence zone, except for their basal area contribution (Seifert *et al.* in review).

The basal area of larger trees (BAL) is identical to the LBA except that it only includes trees which are larger in DBH than the reference tree, which would then provide a good indication of overtopping and radiation exclusion.

The Hegyi Index (Hegyi 1974) calculates size ratios of the reference tree with its competitors multiplied by an inverse distance weighting. These are then all summed up for each reference tree within a zone of competition.

The Voronoi index is a simple growing space index where the space in a given stand is divided equally among trees, first by drawing distance lines between each tree and its neighbours, these lines are then bisected and lines are drawn connecting the bisected lines, resulting in a polygon for each tree based on its growing area with regards to its neighbour (**Figure 4-5**). This is thus simply a refined measure of a tree's growing area and the relationships would change with age and competition. It is included here as a benchmark – to illustrate the performance of the other indices in relation to a simplistic growing area index. It must be noted that refinements of the Voronoi exist, where the size of the polygons is changed proportional to the size ratio of the tree compared to its neighbours or even to measured crown dimensions (Seifert and Utschig 2002). This was not examined in this study, but would be a potential point of interest for future comparison.

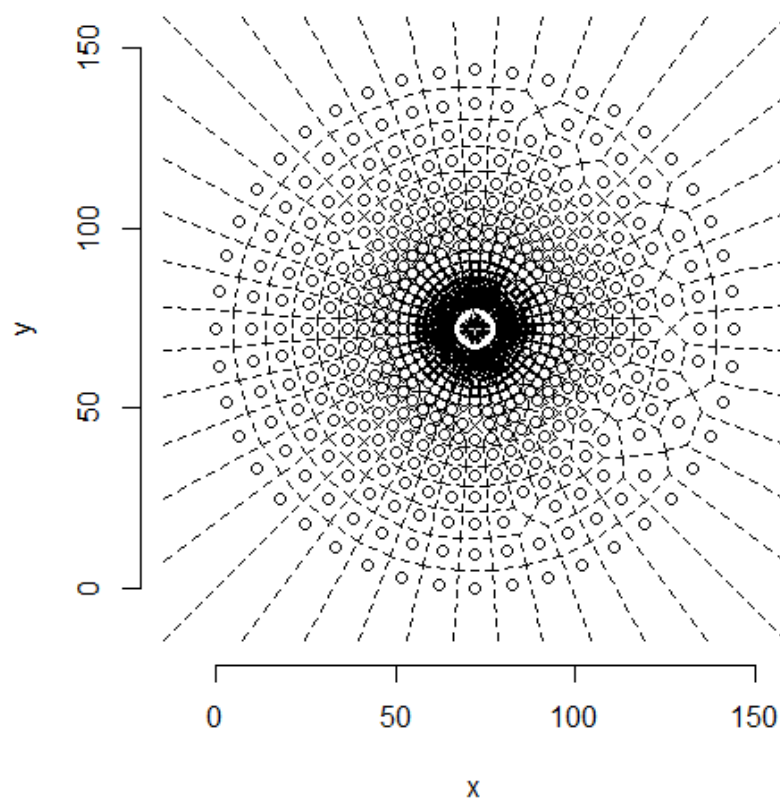


Figure 4-5: Voronoi polygons calculated for the Nelder spacing trial as an example

4.4.2 Edge effects

In the Tweefontein SSS-CCT trial, although border trees were present, they were not measured. With the small size of each stand, 25 trees in a stand, many of the indices will have search zones which fall outside of the boundary. This is noted as a possible limitation of the study.

Pretzsch (2009) mentions three methods to simulate or extrapolate stand structure outside of the edge of the measured stand: plot mirroring, shifting and linear expansion. In this study, the plot shifting technique was used for the SSS-CCT trial, where the plot is shifted along the sides and corners creating eight edge plots, with the measured plot in the centre. This was not necessary for the Nelder trial, the first and last rings were simply repeated along the gradient for positions.

4.4.3 Competition search radius (influence zone)

The definition of the size of the search radius is critical to the performance of the competition indices – if the zone is too big or too small, information on competition will be

lost. The first decision to be made is whether the search radius should be fixed or relative (to stems per hectare for instance). In this study a fixed radius for the different stand densities was chosen, however on which would change relative to the dominant height of a stand at a certain age, the hypothesis being that the size of the influence zone would change as trees become larger in a stand. The Hegyi, LBA and the BAL indices use fixed search radii and it was decided to standardise the radius for all three indices in order to draw reasonable comparisons.

The LBA was used as a reference to determine how large the search radius should be as it is the simplest of the three indices and is not complicated by additional factors such as size of competitors, etc. Furthermore, the LBA index has a much clearly linear correlation with diameter increment (**Figure 4-8**), with a much higher link to linear correlation than the other indices, where more sophisticated nonlinear or transformation applications would have been necessary to determine the maximum correlation with the search radius.

To determine the optimum search radius, different sizes of search radii were tested on every measurement age of the Nelder trial and the correlation of the LBA with diameter increment determined (the minimum of each correlation was chosen since the size of the competition has a negative effect on diameter increment). This is represented in **Figure 4-6** and summarised in **Table 4-5** below.

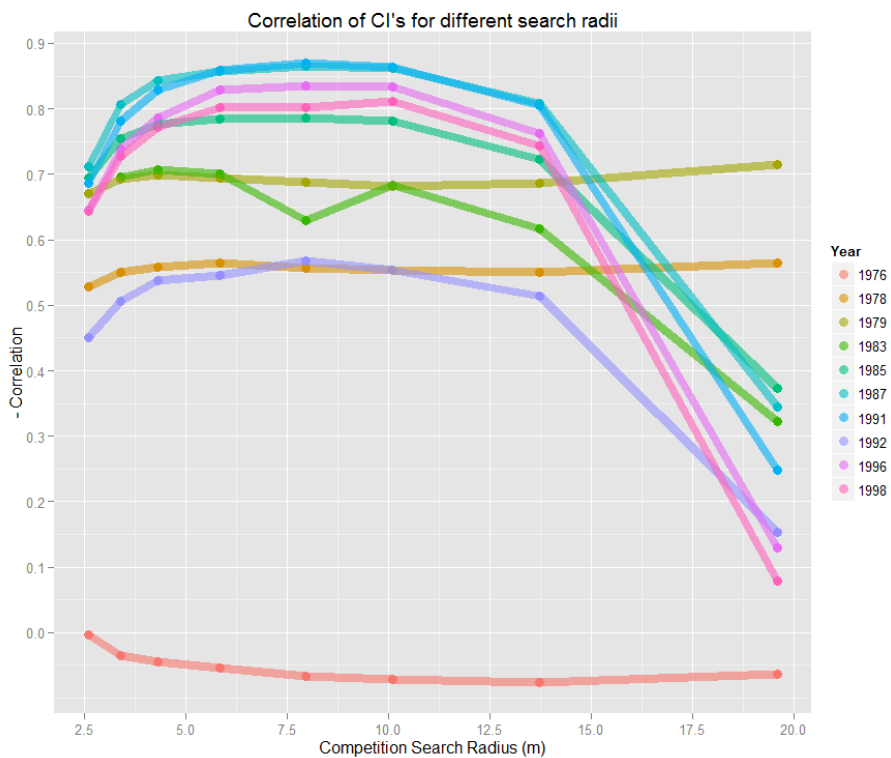


Figure 4-6: Negative Correlation of the LBA competition index with diameter increment at different measurement years (ages) using different competition search radii, the correlation was multiplied by -1 for illustrative purposes.

Table 4-5: Correlation of the LBA competition index at different ages of the Nelder trial with the resulting search radius size included

Age	Dominant height	Minimum correlation	Optimum search radius
4.58	4.314737484	-0.36186	2.622173
6	5.929475765	-0.66753	5.861549
11.08	11.77634706	-0.8528	4.312308
13.33	14.28871325	-0.79171	7.970661
15.58	16.71489227	-0.83003	7.970661
19.17	20.37716099	-0.82632	7.970661
20.25	21.42530206	-0.57752	7.970661
23.83	24.7174421	-0.81951	7.970661
26	26.57690163	-0.84739	10.11444

A simple linear model was then fitted for the values from Ages 4-26 – with their respective dominant heights in order to exclude age from the calculations illustrated in **Figure 4-7**, with the resulting model in **Table 4-6**.

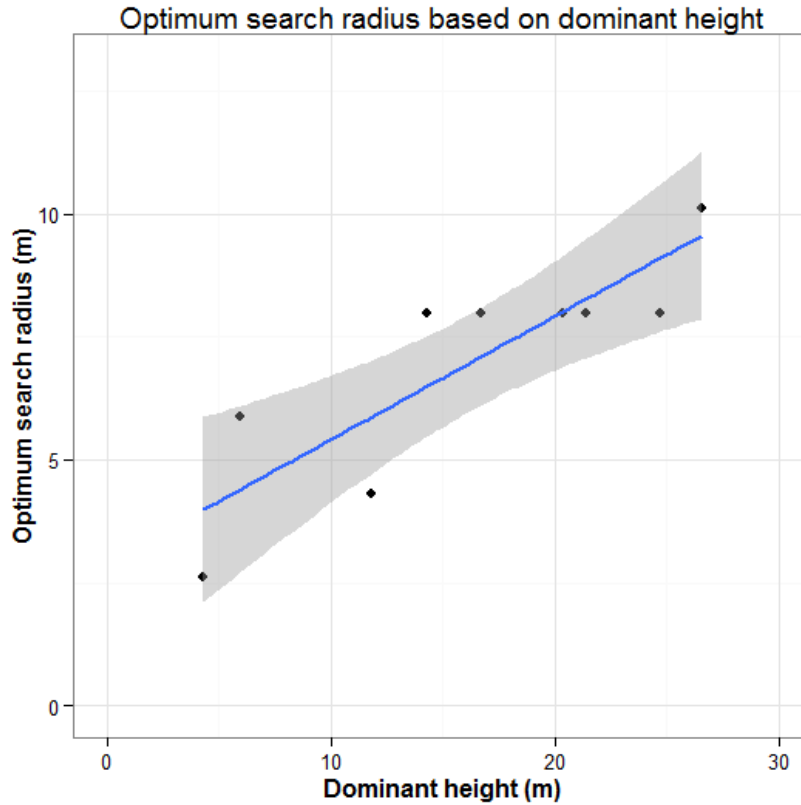


Figure 4-7: Linear regression of the optimum search radius over dominant height

Table 4-6: Linear regression coefficients of the optimum search radius based on dominant height

Linear (OLS) competition zone model					
Parameter	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.90895	1.00781	2.886	0.02344	*
Hdom	0.25036	0.05646	4.434	0.00303	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.256 on 7 degrees of freedom

Multiple R-squared: 0.7375, Adjusted R-squared: 0.7

Thus the competition search zone will increase with the given dominant height, which was used for all three competition indices that required a search radius (LBA, BAL and Hegyi).

4.4.4 Performance of competition indices

Each of the competition indices introduced in Section 4.4.1 was applied to the Lottering and Tweefontein spacing trials resulting in the following diameter increment in relation to the competition index plots below (Figure 4-8).

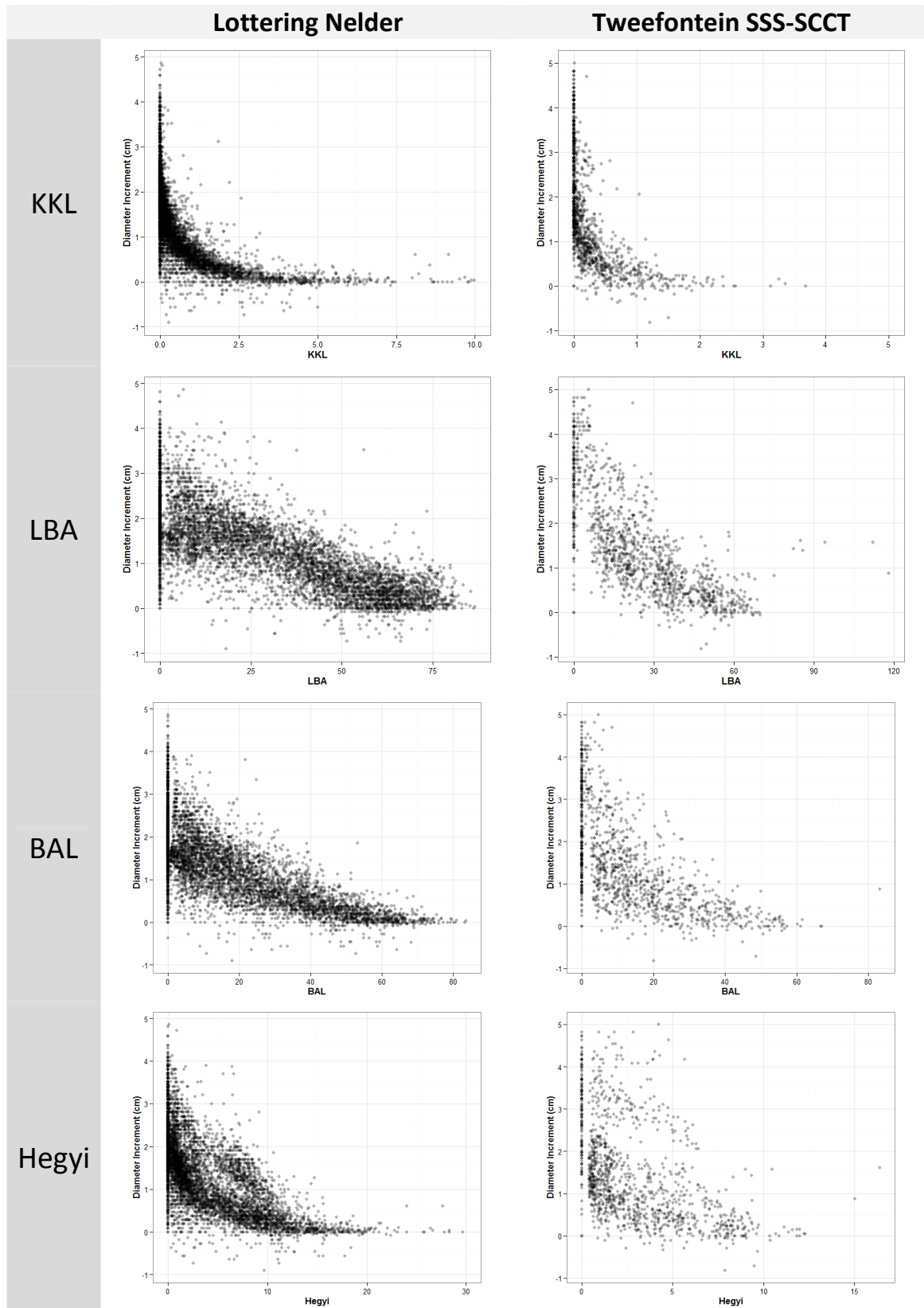


Figure 4-8: Relationship of diameter increment and the competition indices fitted in the two spacing trials used for this study

The Hegyi and KKL indices both had a clear negative exponential trend of increment over the CI value, the BAL index also showed a nonlinear trend and the LBA index shows a linear response. Significant variation exists in the plots, especially when competition indices approach zero (state of no competition). The Voronoi index did not show much correlation (Appendix B); however this could be because the correlation changes at different ages or stages of development of the stand, which this investigation is trying to avoid.

4.5 Step 4: Selecting competition indices

As seen in the previous section, each competition index, by virtue of the different aspects that they measure, covers different explanatory aspects of competition and thus would potentially contribute explanatory value. Thus, combining two or more CI's to the modifier equation by creating a multiple regression equation, more aspects of competition might be covered and a better fit may be obtained with respect to diameter increment.

However, overlap of the explanation between competition indices or collinearity may distort the explanation value of each index used. Collinearity (or multicollinearity, when more than two variables are included in a model) occurs when predictor variables in a multiple regression are highly correlated (Myers 1986). Collinearity does not always affect the predictive power of a model on the data that it is fit. However, collinearity significantly affects the stability of the model and hinders the ability to draw valid conclusions about individual regressor variables; small changes in y-values can significantly alter model coefficients even though measures of fit (e.g. R^2 values) remain relatively unchanged (Myers 1986). This would compromise one of the objectives of this chapter – to determine how the importance of different variables changes under different edaphic and site quality conditions. Additionally, although it is tempting to add regressor variables (the CI's) to obtain a better fit, the concept of Ockham's razor should be followed, where parsimonious models are sought. Furthermore, tedious calculation of every competition index would not be ideal for any future model application.

In view of the objectives of this chapter, the set of chosen competition indices was thus tested individually as well as in combinations to identify index combinations that were not significantly collinear. Ideally, a combination of two CI's was to be obtained that adequately describe diameter increment. The resulting model should provide a good explanatory power with regards to overtopping (light limitations) and local crowding (edaphic limitation) – as in the example of **Figure 1-2**.

This required a variable selection exercise, where correlation (collinearity) and behaviour of the respective indices was sought through principle component analysis (PCA) and variance inflation factors (VIF's). Different combinations of CI's variables were tested sequentially to determine the best combination of CI's. Relative importance measures were then carried out with the two final selected CI's on different sites.

Figures 4-9 and 4-10 show the pairwise matrix scatterplots of the different competition indices against each other in the Nelder and SSS-CCT trials, indicating correlation between certain variables.

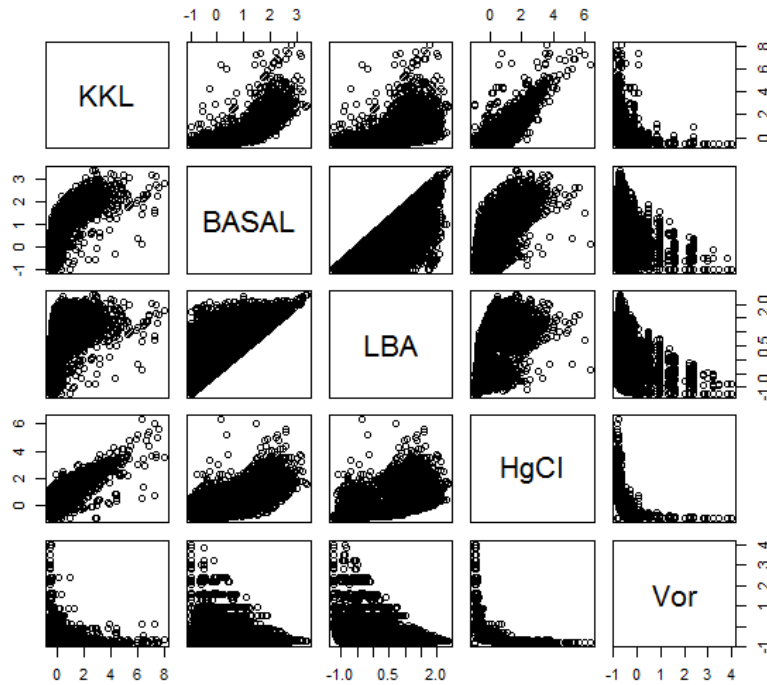


Figure 4-9: Pairwise scatterplot matrix of the various competition indices against each other in the Nelder Trial

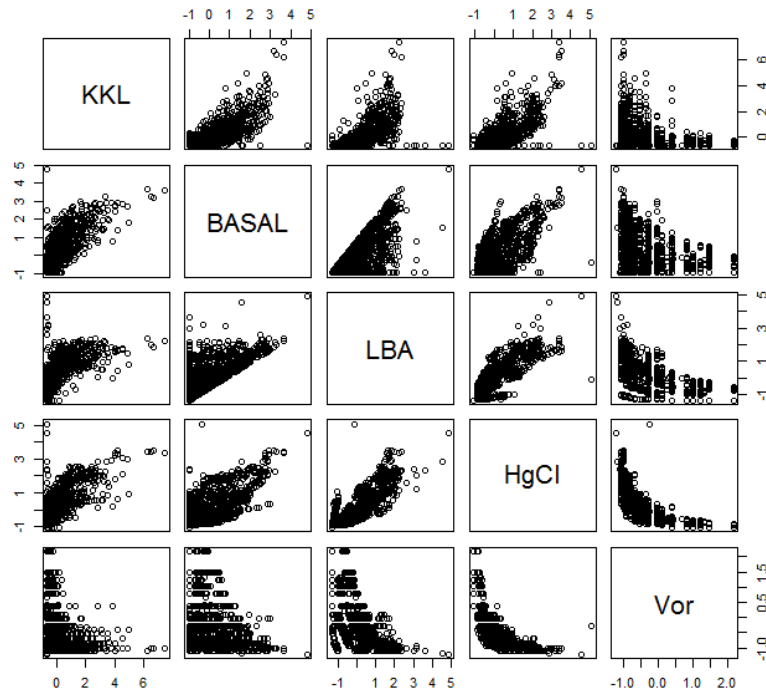


Figure 4-10: Pairwise scatterplot matrix of the various competition indices against each other in the Tweefontein Trial

4.5.1 Variable selection

As seen in **Figure 4-8** most of the competition indices were nonlinear in their effect on *id*. A simple square root transformation was found to be adequate to linearise the response;

$$\sqrt{id} = f(x)$$

where $f(x)$ represents a linear combination of one, several, or all of the competition indices. Square root transformations have the advantage over logarithmic transformations that, for diameter increment where 0 values are prevalent, no infinite values are produced. The transformation produced visually linear fits (ordinary least squares) for all of the models, except for the KKL and Voronoi.

Table 4-7: Linear models coefficients of the square root transformation of diameter increment using the respective competition indices

Model						R ²	AIC
sqrt(DbhIncrement) ~ BAL							
	Estimate	Std. Error	t value	Pr(> t)			
(Intercept)	1.404668	0.004503	312	<2e-16	***	0.6002	2088.409
BAL	-0.01864	0.000177	-105.1	<2e-16	***		
sqrt(DbhIncrement) ~ LBA							
	Estimate	Std. Error	t value	Pr(> t)			
(Intercept)	1.496116	0.005296	282.5	<2e-16	***	0.5834	2392.123
LBA	-0.01444	0.000142	-101.5	<2e-16	***		
sqrt(DbhIncrement) ~ HgCl							
	Estimate	Std. Error	t value	Pr(> t)			
(Intercept)	1.35438	0.005146	263.17	<2e-16	***	0.4606	4291.707
HgCl	-0.07005	0.000884	-79.24	<2e-16	***		
sqrt(DbhIncrement) ~ kkl							
	Estimate	Std. Error	t value	Pr(> t)			
(Intercept)	1.240745	0.004345	285.59	<2e-16	***	0.4524	4403.322
kkl	-0.25707	0.003299	-77.93	<2e-16	***		
lm(formula = sqrt(DbhIncrement) ~ Voronoi, data = sss)							
	Estimate	Std. Error	t value	Pr(> t)			
(Intercept)	0.893277	0.006296	141.88	<2e-16	***	0.1995	7195.024
Voronoi	0.008485	0.000198	42.81	<2e-16	***		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							

Table 4-7 shows that the LBA and BAL indices explain the variance well, (showing the highest R² values), although it must be noted that the KKL could be better described with an alternative transformation.

Combining all of the indices into a full model to describe diameter increment according to **Equation 4-3** shows how a much improved fit can be obtained by adding variables together, indicated by their respective names in the equation, with parameters a-e to be determined (**Table 4-8**).

$$\sqrt{DBH_1} = a * kkl + b * Hegyi + c * LBA + d * Voronoi + e \quad \text{Equation 4-3}$$

However, due to interactions and collinearity between variables, a decision has to be made of which variables should be included in the model that represents a good compromise between model fit and reduced correlation between the variables. Variance inflation is one method used to detect collinearity (Myers 1986):

$$VIF_j = \frac{1}{1 - R_j^2}$$

Equation 4-4

where the **VIF** for a variable is the reciprocal of the inverse of R^2 from the regression. Variance inflation is calculated for each of the explanatory variables. As a guideline a VIF of 5, or a square root of the VIF larger than two, indicate collinearity in the model.

Table 4-8: Full linear model of the square root diameter increment transformation using all of the CI's as regressor variables

Model: sqrt(Diameter Increment) = LBA + BAL + HgCl + KKL + Voronoi							
Parameter	Estimate	Std. Error	t value	Pr(> t)		VIF	sqrt(VIF)
(Intercept)	1.5046982	0.0089776	167.606	< 2e-16	***		
LBA	-0.0072389	0.0002346	-30.863	< 2e-16	***	3.496665	1.869937
BAL	-0.0065211	0.0003478	-18.749	< 2e-16	***	4.750188	2.179493
HgCl	-0.0165476	0.0014857	-11.138	< 2e-16	***	4.710291	2.170321
kkl	-0.0434107	0.0052941	-8.2	2.82E-16	***	4.361191	2.088346
Voronoi	-0.0004587	0.0001793	-2.558	0.0106	*	2.026482	1.423546

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Multiple R-squared: 0.6767, Adjusted R-squared: 0.6765
F-statistic: 3077 on 5 and 7349 DF, p-value: < 2.2e-16

Table 4-8 shows a significant fit for all of the parameters, however VIF is quite high for most of the variables indicating that the model needed to be reduced and dependencies between variables better understood.

4.5.2 Principle component analysis

Principal component analysis (PCA), (Hotelling 1933), can be used to reveal similarities in explanation value and thus collinear structures between variables. It is a multivariate statistical feature extraction method that works by projecting the data onto orthogonal vectors, called principal components, which explain the maximum variation of the data. These principal components are the features of the data.

Detailed information of PCA can be found in multiple texts. Briefly explained though: If \mathbf{X} is a matrix with the columns representing variables and the rows representing sample observations, the first principal component \mathbf{p}_1 is the linear combination where the projection onto \mathbf{p}_1 , given by $\mathbf{t}_1 = \mathbf{p}_1 \mathbf{X}$, yields the maximum variance subject to $\|\mathbf{p}_1\| = 1$. The projection $\mathbf{t}_2 = \mathbf{p}_2 \mathbf{X}$ onto the second principal component \mathbf{p}_2 is the combination that gives the second highest variance subject to $\|\mathbf{p}_2\| = 1$ and \mathbf{p}_2 being orthogonal to \mathbf{p}_1 . This is extended to all further principal components. The orthogonality criterion ensures that each component is completely uncorrelated with all other components, thereby ensuring that the components explain the maximum variation. The principal component loading vectors, \mathbf{p}_i (i

$= 1, \dots, m$), are the eigenvectors of the covariance matrix of the data set, therefore they can be obtained from spectral decomposition of the covariance matrix.

Once the principal components have been determined a biplot can be used to visualise which variables contribute most to each principal component, it also shows similarities or collinearity behaviour of the variables in relation to each other in the transformed space. When a biplot is applied to PCA results, the axes are a pair of principal components. The points represent the principal component scores of the observation, and the vectors represent the coefficients of each variable on the principal components. So in the biplot the longest vector will represent the variable that contributes most to the variation in the data and therefore provides the most important information for the model.

R has a function called *princomp* in its stats package for PCA. A data matrix consisting of the competition indices is passed into it and it performs PCA on it.

Table 4-9: Importance measures of the principle components

Measure	Importance of components:				
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.646021	1.034012	0.879463	0.560748	0.365429
Proportion of Variance	0.541877	0.213836	0.154691	0.062888	0.026708
Cumulative Proportion	0.541877	0.755713	0.910405	0.973292	1

Table 4-9 above shows that Component 1 accounts for 54% of the variation in the data: The loadings of each variable (CI) of the principle components above can be shown numerically in Table 4-10.

Table 4-10: Principle component loadings of the respective principle components

Index	Principle Component Loadings:				
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
KKL	-0.455	0.424	-0.421	-0.285	-0.596
BASAL	-0.479	0.286	0.303	-0.496	0.592
LBA	-0.455		0.687	0.447	-0.348
HgCl	-0.476		-0.503	0.6	0.398
Vor	0.361	0.858		0.336	0.122

Visually this is shown in the biplots for both sites (Figures 4-11 and 4-12), where the length of the arrow for each variable indicates its contribution to the component loading.

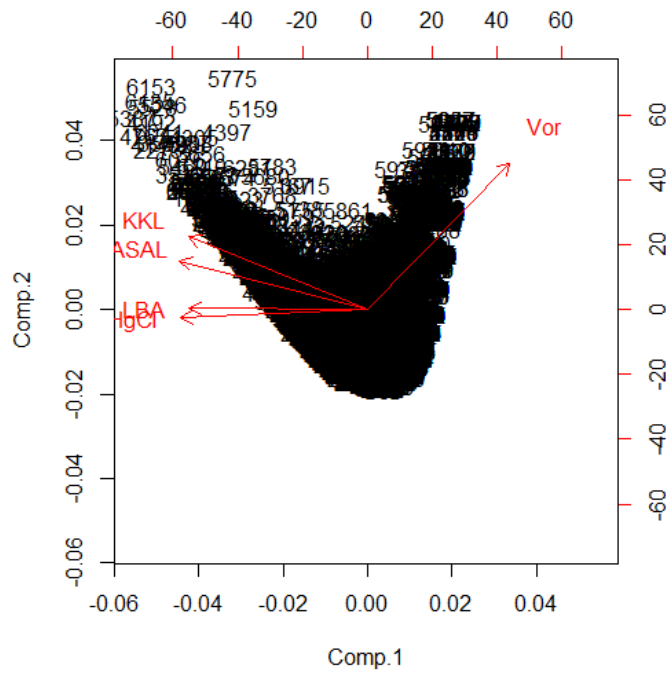


Figure 4-11: PCA biplot of the variable plotted in the transformed space of the first two principle components in the Nelder trial

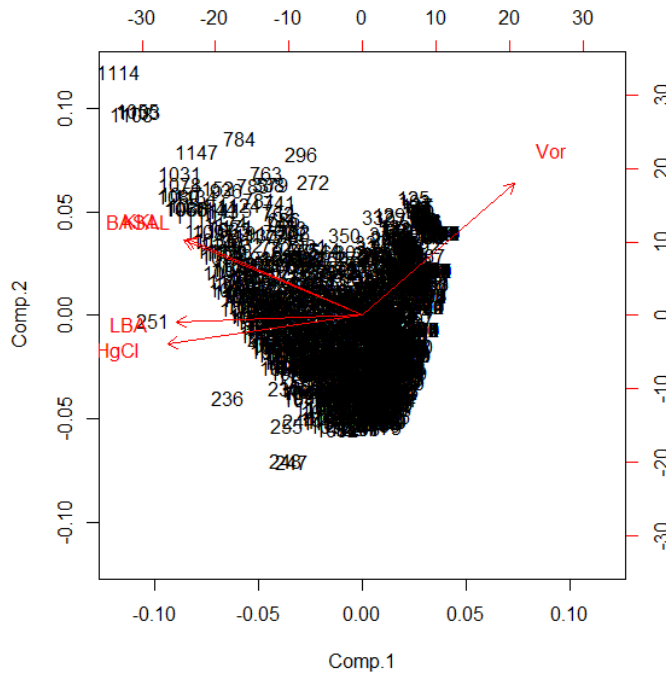


Figure 4-12: PCA biplot of the variable plotted in the transformed space of the first two principle components in the Tweefontein trial

Not much can be determined from the length of the vectors (or their loadings) as they load relatively equally on the two main components. However, **Figures 4-11 and 4-12** show consistent performance on the two sites with regards to behaviour (grouping) of the CI variables. The Voronoi index is negatively correlated to the other potential predictors – a high Voronoi area will mean that there is more growing space and thus a higher increment, as opposed to the other variables which increase in values with increasing competition. However, what is interesting is the correlation between the other four variables, which will be necessary for stratification of the variables to make deductions about their expected similarity. The KKL and the BAL are very similar, as they characterise competition strongly by overtopping. The Hegyi and LBA, which might explain edaphic competition (local crowding) to a higher degree, are closely grouped compared to the other indices.

As explained earlier, collinearity will be a problem for creating a combined model. The PCA loadings above give a qualitative idea of which variables to include. Clearly, the two groups can be seen, with the KKL and BAL closely related, and the Hegyi and LBA closely related as well. It was decided to include one variable from each grouping (e.g. the KKL and Hegyi) for multiple collinearity problems and for parsimony in the model. The Voronoi index was not considered for further modelling purposes due to its low explanation value (**Table 4-7** and **Table 4-6**). **Table 4-11** below shows the results for the various pairings of indices:

Table 4-11: Linear model coefficients for the combinations of the competition indices

Model					VIF	Sqrt(VIF)	R ²	AIC
sqrt(DbhIncrement) ~ kkl + LBA								
	Estimate	Std. Error	t value	Pr(> t)				
(Intercept)	1.467	0.005	300.95	<2e-16	***	1.563	1.250	1005.657
kkl	-0.128	0.003	-39.08	<2e-16	***			
LBA	-0.011	0.000	-65.73	<2e-16	***			
sqrt(DbhIncrement) ~ kkl + HgCI								
	Estimate	Std. Error	t value	Pr(> t)				
(Intercept)	1.322994	0.005085	260.17	<2e-16	***	2.915	1.707	3671.775
kkl	-0.136476	0.005358	-25.47	<2e-16	***			
HgCI	-0.040179	0.001447	-27.77	<2e-16	***			
sqrt(DbhIncrement) ~ LBA + BAL								
	Estimate	Std. Error	t value	Pr(> t)				
(Intercept)	1.4841496	0.004857	305.55	<2e-16	***	3.033	1.742	1089.763
LBA	-0.007428	0.000227	-32.73	<2e-16	***			
BAL	-0.010904	0.000289	-37.77	<2e-16	***			
sqrt(DbhIncrement) ~ HgCI + BAL								
	Estimate	Std. Error	t value	Pr(> t)				
(Intercept)	1.4324895	0.004447	322.12	<2e-16	***	2.024	1.423	1457.380
HgCI	-0.026661	0.001037	-25.71	<2e-16	***			
BAL	-0.014221	0.000242	-58.82	<2e-16	***			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1								

The combination of the KKL and LBA indices had the best R² values, lowest AIC values and had lowest VIF – suggesting that they were the least collinear of the above combinations and represented the best fit. A combination of LBA and BAL also resulted in a good fit; however the VIF was the highest as they are based on the same competition index.

The combination of KKL and LBA intuitively represent different modes of competition, which is quantified in the better fit obtained and the low collinearity between the two CI's. These two were thus selected as a candidate combination to test the changing mode of competition and combination of CI's for the modifier function.

4.5.3 Relative importance

Now that a candidate combination was obtained, where collinear effects would be minimal to the explanation value of the two CI's - a comparison between the importance of variables in each site is given. The full model and the suggested reduced model (KKL and LBA) are introduced to see how they change over the two sites – assuming that the water index is representative of a gradient.

The *relaimpo* package developed by Grömping (2006) in R was used for this purpose. The package provides different methods for decomposing the contribution of R^2 with different predictor variables provided that the regressors are not correlated. Grömping (2006) recommends that two particular methods “**lmg**” method and “**pmvd**” method, with **lmg** more appropriate for causal analysis and **pmvd** more appropriate for predictive analysis (Grömping 2006). The other methods were included here for a complete picture to analyse whether the relative importance changes consistently with other variance decomposition methods.

The results for the Nelder and Tweefontein trials are summarised in **Table 4-12** below, using different variance decomposition methods for illustration and to see consistency with other methods, although the **lmg** and **pmvd** methods are of importance.

Table 4-12: Relative importance proportions (explanation contribution) of the KKL and LBA indices respectively using different importance measures

Method	Nelder		Tweefontein	
	KKL	LBA	KKL	LBA
lmg	0.431	0.569	0.407	0.593
pmvd	0.329	0.671	0.255	0.745
betasq	0.329	0.671	0.255	0.745
car	0.413	0.587	0.379	0.621
last	0.329	0.671	0.255	0.745
first	0.457	0.543	0.443	0.557
Total R2	61.71%		69.16%	

The LBA was attributed consistently a higher importance to the variance explanation of the model on every method type. In the **lmg** method, this was small, suggesting a small difference between sites, however the **pmvd** method showed a much larger difference. The results are illustrated in **Figure 4-13** and **Figure 4-14** for the Tweefontein and Nelder sites respectively, with 95% bootstrapped confidence intervals for the respective importance values for the different CI's.

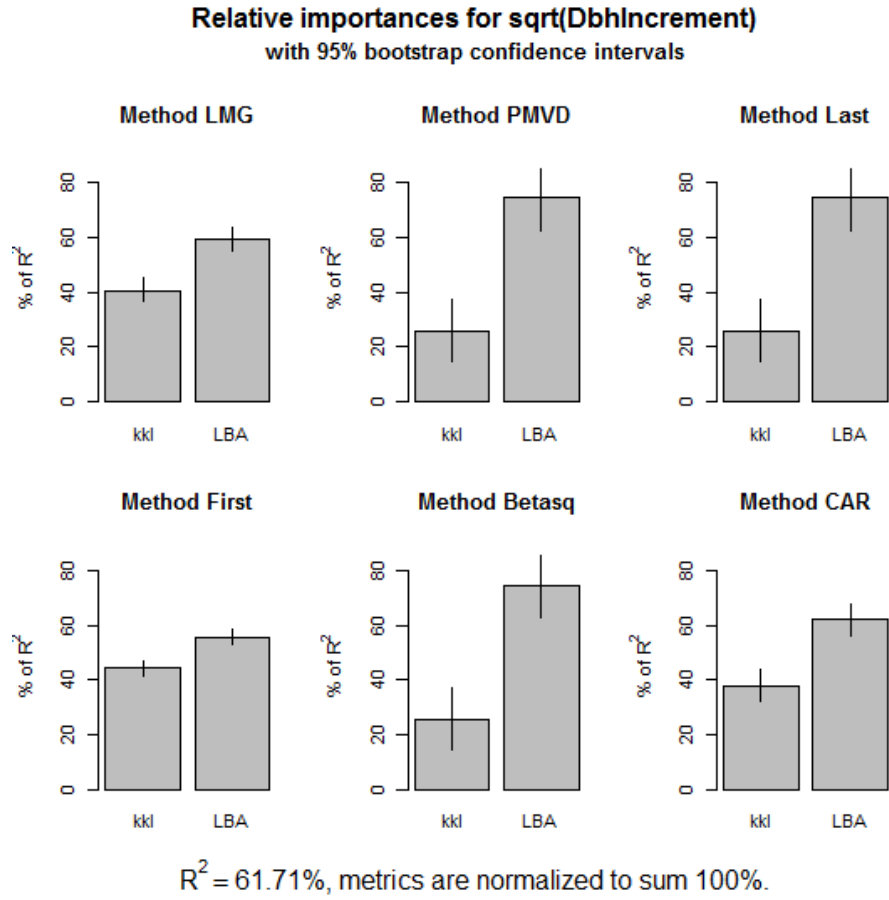


Figure 4-13: Relative importance graphs using different methods for the Tweefontein spacing trial

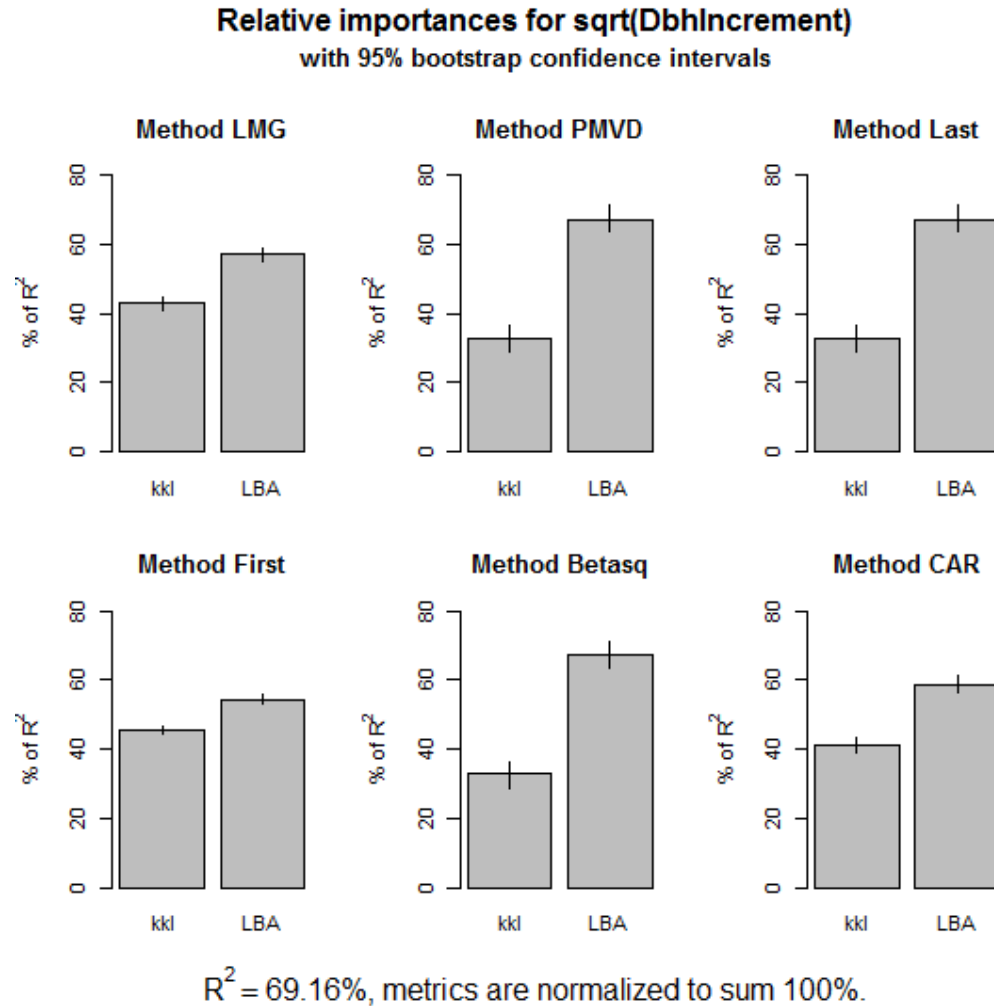


Figure 4-14: Relative importance graphs using different methods for the Nelder spacing trial showing improved KKL importance as compared to the Tweefontein site

Regardless of the method it is clear the LBA has the largest contribution to the R^2 value in both sites, with the KKL increasing in importance in the Nelder site, which was classified as a wetter site according to the aridity index (Table 4-1) for both the **lmg** and **pmvd** methods consistently. It can be tentatively hypothesised that this could be due to the better moisture conditions on the site. This is an encouraging result for changing importance of competition indices based on resource limitations and changing mode of competition. However more sites are necessary for proper validation of such a hypothesis. For the purpose of this thesis water index was used as a considered addition to the modifier function.

4.6 Step 5: Use CI's in a deterministic potential modifier equation

Step 2 defined the potentials, which can be modelled for any given diameter on a certain site. Using the information gained in the previous section, the competition indices can now be incorporated into a modifier. Two objectives are represented here: to see whether the

incorporation of an additional competition index, which would balance the light oriented KKL with an index more focussed on local crowding and edaphic limitations, would improve the model. The second objective is to see if the incorporation of a water index, which can scale the CI's according to the available water, improves the model fit.

4.6.1 Objective 1: Incorporating the LBA competition index

Using the KKL and the LBA in a linear model within the modifier of the potential modifier approach results in **Equation 4-5**:

$$id = id_{pot} * e^{(a+b*KKL+c*LBA)} \quad \text{Equation 4-5}$$

Where i_d is the mean annual diameter increment, id_{pot} is the potential diameter increment for that specific site with its given site index calculated in **Equation 4-2** and illustrated in **Figure 4-4**, and the exponential function with the linear equation representing the modifier. For the Nelder and Tweefontein sites together this was represented by:

Table 4-13: Coefficients of the diameter increment potential modifier formula (Equation 4-5)

Formula: Equation 4-5					
Parameter	Estimate	Std. Error	t value	Pr(> t)	
a	-0.4224962	0.0067322	-62.76	< 2e-16	***
b	-1.1991936	0.0408444	-29.36	< 2e-16	***
c	-0.0036113	0.0005382	-6.71	2.09E-11	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.6703 on 7493 degrees of freedom					

The exponent of the linear equation scaled the linear equation between 0 (absolute theoretical competition: no growth) and 1 (no competition: maximum growth). The intercept presented a potential problem as it will always reduce the potential (i.e. the modifier will never reach 1 (the potential growth). However, excluding this intercept resulted in an over scaled, unrealistic model where too many trees are growing at the potential or not growing at all, which would not be ideal for the simulation. For this reason, it was decided not to exclude the intercept from the model in order to obtain a good average, with a stochastic component to be added later (Step 6).

As stated, the first objective is to see whether including the local basal area in addition to the KKL improved the model. This was done by only including either the LBA or KKL in **Equation 4-5**, comparing each of them in addition to the combined model.

Table 4-14: Anova comparison if Equation 4-5 with KKL, LBA and both indices included in Model 1, 2 and 3 respectively

Analysis of Variance Table									
Model	Res.Df	Res.Sum Sq	Df	Sum Sq	AIC	RMSE	F value	Pr(>F)	
1	7494	3385.5			15320.34	0.67204			
2	7494	4139.9	0	0	16828.38	0.743157			
3	7493	3366.5	1	773.37	15280.29	0.670158	1721.3	<	2.20E-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The Anova comparison shows a significant improvement to the AIC, RMSE and F-statistic. Including both indices seems to indicate that including the LBA in conjunction with KKL in Equation 4-5 improves the predictive fit for diameter increment modelling.

4.6.2 Objective 2: Incorporating a water index

In order to include the water availability with the aridity index presented in Table 4-1, a new linear model inside of the modifier can be incorporated:

$$Modifier = W \cdot KKL + W \cdot LBA + a \tag{Equation 4-6}$$

Where *W* represents the FAO – UNEP (1992) water index, *KKL* and *LBA* are the CI's and *a* is the intercept. This can be incorporated as the modifier into Equation 4 -7:

$$id = id_{pot} * e^{(a+b \cdot W \cdot KKL + c \cdot W \cdot LBA)} \tag{Equation 4-7}$$

Table 4-15: Coefficients of the diameter increment potential modifier formula (Equation 4-7) incorporating the water index

Formula: Equation 7					
Parameter	Estimate	Std. Error	t value	Pr(> t)	
a	-0.43025	0.006868	-62.647	< 2e-16	***
b	-1.15627	0.040222	-28.747	< 2e-16	***
c	-0.00298	0.00053	-5.622	1.96E-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6782 on 7493 degrees of freedom

Section 4.5.3 (Relative importance) showed that for the two sites being used for parameterisation, the importance of the indices changed for the square root transformed model. Thus the next objective was to see whether the water index does indeed improve the model by comparing Equation 4-7 and Equation 4-5

Table 4-16: Anova comparison of the diameter increment model with (Model1) and without (Model 2) the water index

Model	Res.Df	Res.Sum Sq	Analysis of Variance Table				F value	Pr(>F)
			Df	Sum Sq	AIC	RMSE		
1	7493	3446.9			15457.02	0.678104		
2	7493	3366.5	0	0	15280.29	0.670158		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4-16 shows no significant difference between the two models based on the F-statistic. This could mean that with only two sites containing numbers of observations and only two water index values, not enough information was available to test whether this should be included in the full model, despite the plausible relative changes of the variable importance observed (4.5.3). Furthermore, as the water index was a static value for each site, it may be improved by a water availability index in order to introduce dynamic effects. Unfortunately, such an index was not available in this study. In order to test whether the model performance was due to the relative size of the datasets – random samples from the Nelder experiment were taken to match the number of observations of the Tweefontein site and a comparison tested (**Table 4-17**).

Table 4-17: Anova comparison of the diameter increment model with (Model1) and without (Model 2) the water index by randomly subsetting data from the Nelder trial to match the number of observations in the Tweefontein trial

Model	Res.Df	Res.Sum Sq	Analysis of Variance Table				F value	Pr(>F)
			Df	Sum Sq	AIC	RMSE		
1	2439	1093.3			4975.732	0.669118		
2	2439	1048.8	0	0	4874.234	0.655356		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

This result was consistent with multiple randomised samples, suggesting that the water gradient provided here between the two sites may not have been large enough to cover changes in CI behaviour with water availability. For further modelling it was decided to use the model (**Equation 4-7**) with the water index on the different sites nonetheless since it showed a logical behaviour and was statistically not different in the explanation value than **Equation 4-5** without the water index.

4.6.3 Model behaviour

Predictions from Equation 7 were used on the two datasets to analyse the structure of the predicted values compared to the observed increment values. **Figure 4-15** shows the procedure where the potentials for each given site are calculated (a), after which a modifier (b) reduces the increment. The observed and predicted increments for each site are shown sequentially (c-f).

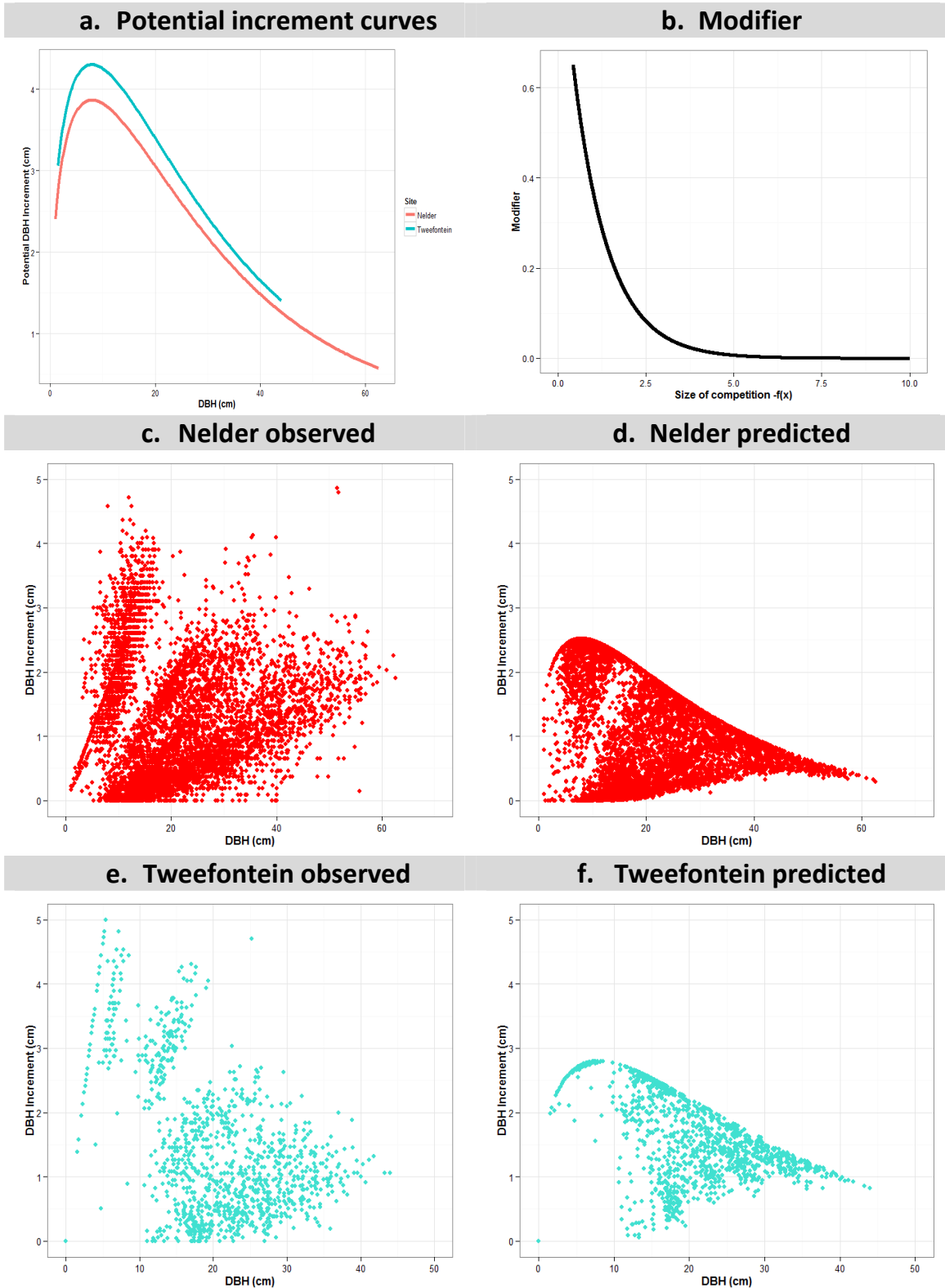


Figure 4-15: Illustration of the potential modifier approach on the two spacing trials, with the top right hand diagram representing the change of the modifier over the size of competition

The predicted increment follows a pattern that mimics the trend of competition at different DBH values (**Figure 4-1**); however the predicted values show much less variability than the observed increments shown above.

What is also evident is that the predicted increments are constrained both by the potential and by the maximum value of the modifier (which was 0.65 in this case), which means that the increment cannot increase above 0.65 times the increment potential for a given DBH. The depression of the modifier was due to the inclusion of an intercept in the modifier equation (α in **Equation 4-7**), which was nonetheless included because it includes variation from factors unexplained in the model and prevents extreme values in the modifier. It was decided to test the performance of this model, and then to incorporate random variability as a stochastic component for further improvement in the next section.

4.6.4 Model validation

The chosen model (**Equation 4-7**) was tested on the Nelder spacing trial, with the first measurement age (4.58 years in the Nelder trial) used as a starting point. The simulation time step period was not initiated (in SILVA it is 5 years). In this case the simulation is performed between measurement periods to test how the model reacts in the different time steps. This means that the effect of competition is constant between two measurement ages – thus the simulated increment will remain the same for that period.

The simulated series of the deterministic model (without variance component) produces a series where the diameter increment and thus DBH is overpredicted from age 6 to 23, after which point it begins to converge again (Appendix C). The full range of simulated periods is shown in Appendix D. **Figure 4-16** below shows the end of the simulation at 26 years of the Nelder trial (the final measured age of the trial).

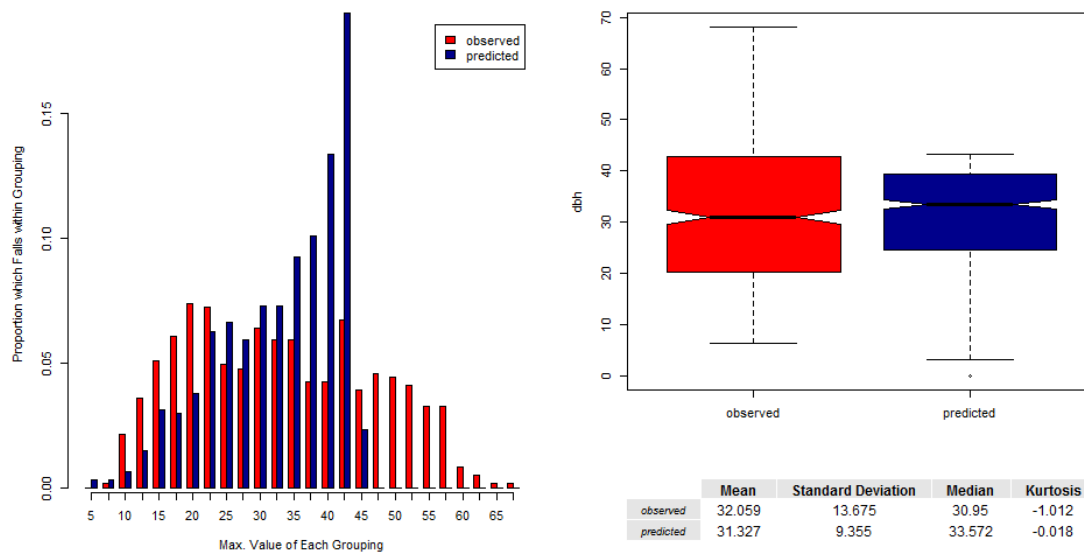


Figure 4-16: Comparison between the observed (red bars) and the simulated diameter (DBH) represented as a distribution of 2.5cm diameter classes. The mean DBH for the site is represented in the notched boxplot. This showed a skewed prediction grouped around a narrower DBH band

As can be seen from the simulation end (**Figure 4-16**) and Appendix D the diameter simulated DBH distribution shifts towards higher DBH classes, clustered around a narrow range of DBH. While this is still a significant improvement from a simple mean model, this could be problematic in the simulation as it will reduce the variability of the diameter classes in the model and reduce the competitive complexity. Thus with each simulation step the model continues to shift in this trend, thereby compounding the error. For this reason it was necessary to see whether the increment prediction with variability improved the model – and finally in which rings (and corresponding SPHA) of the Nelder trials it overpredicted.

4.7 Step 6: Create a stochastic model incorporating natural variability

4.7.1 Incorporating natural variation

Due to the complex nature of biological systems, and our inability (or difficulty) to capture this complexity in a model with distinct variables, it is often necessary to account for this variability around the mean value with the help of stochastic models (Pretzsch 2009). In this case it is desired to approximate (or mimic) the high degree of variability of diameter increment, which could be caused by factors not included in the model, e.g. micro-site variation in water and nutrient supply, weather, genetics, etc. As is seen by the scatter plot of **Figure 4-17**, which represent a residual plot of the predicted value deviation, while much of the increment patterns can be discerned from stand age/competition and the diameter of the tree, there remains a need to incorporate random variance into the model in order to mimic natural variability. Furthermore, the application of the model led to a mismatched

diameter distribution, skewed to the right and highly concentrated around a narrow band of diameter classes (**Figure 4-16**). For this reason a stochastic model was tested that incorporates variability by including a random deviate in the increment function.

It was decided to incorporate natural variance by analysing the residual scatter plots of the predicted values for the two parameterisation sites, with the residual standard deviation of the observed minus the predicted values plotted over the predicted values (**Figure 4-17**). Linear quantile regression models of the residual plots were created to represent the upper and lower bounds and a new random deviate would then be generated between these two bounds based on a normal distribution.

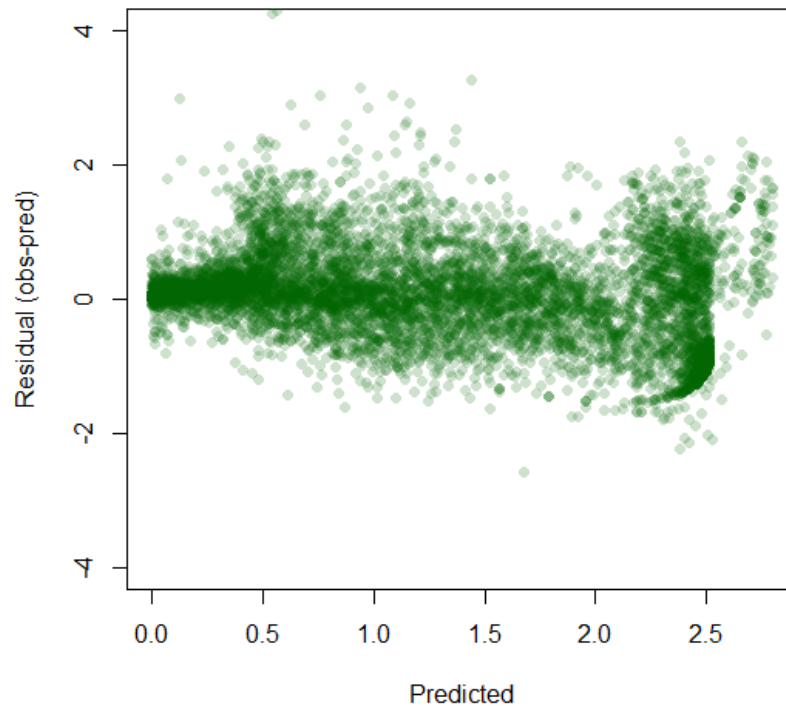


Figure 4-17: Residual plot of the predicted values of the spacing trials

The scatter plot shows that the variance seems to increase slightly with increasing predicted increment values. This is more clear in the negative tail of the distribution, which is expected since the increment is bound by the lower zero value (diameter does not usually become negative). To model this relationship the 0.95 and 0.05 tau value (5% and 95%) linear quantile regression lines were used to represent the upper and lower bounds respectively of the residual plot (**Figure 4-18**). The benefit of the quantile line is that it allows continuous variance bounds for any given predicted value.

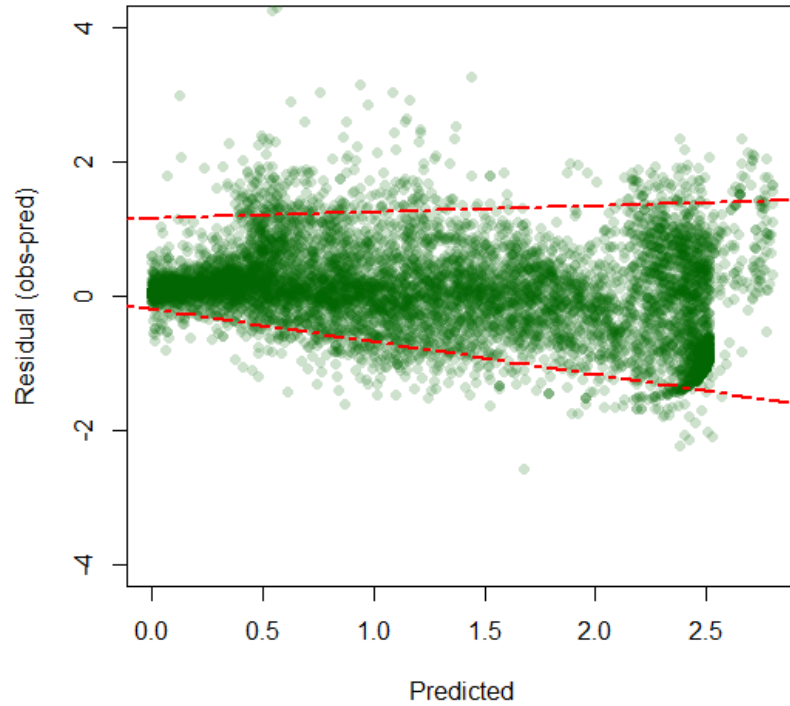


Figure 4-18: 5% and 95% linear quantile regression lines which represents the upper and lower bounds for the truncated normal distribution

This significant positive ($\tau=0.95$) and negative ($\tau=0.05$) slope values for the quantile lines illustrate an increasing variability for higher predicted values, with strong deviation in the negative direction, while the variance in the positive residual direction only increases slightly. The values of the quantile lines are shown in **Table 4-18** below. For instance this means that bounds of the variance for a diameter increment of 1.5 would be between -0.92 and 1.30.

Table 4-18: Linear quantile regression coefficients of 0.05 and 0.95 tau values representing the upper and lower bounds of simulated residual prediction

rq(formula = Increment residuals ~ Predicted Increment, tau = c(0.05, 0.95))					
tau value	Parameter	Value	Std. Error	t value	Pr(> t)
tau: 0.05	(Intercept)	-0.18648	0.02259	-8.25416	0
	Increment	-0.48784	0.01464	-33.3323	0
tau: 0.95	(Intercept)	1.15066	0.03729	30.85437	0
	Increment	0.09447	0.0254	3.71888	0.0002

While the quantiles do not capture all of the data points, they proved to be the most robust with some observed variation sacrificed.

As is seen from **Figure 4-18** the predicted bounds from the quantile regression is quite large, indicating a high degree of variability, however, the points are still strongly clustered around the mean or predicted value and the distance of the bounds increases with predicted increment. For this reason and because the distance between the bounds changes over the size of the predicted increment), a weighting factor was added to the standard deviation value of the normal distribution, where the standard deviation is defined by:

$$StDev = 0.5 * (upper - lower) \quad \text{Equation 4-8}$$

where **StDev** is the standard deviation of the normal distribution, **upper** and **lower** are the truncation bounds for any predicted value.

The results of the predicted diameter increment based on the complete stochastic model produced a scatter plot illustrated below overlain on the observed residual plot (**Figure 4-19**), with the full visual procedure illustrated in **Figure 4-20**.

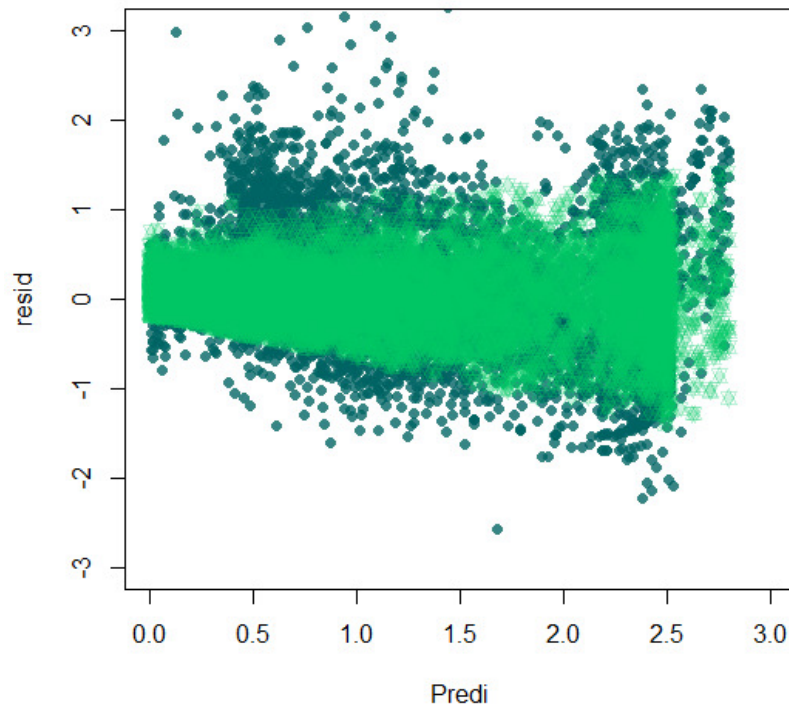


Figure 4-19: Predicted residual generated from a random deviate between two bounds (Figure 18) based on a normal distribution

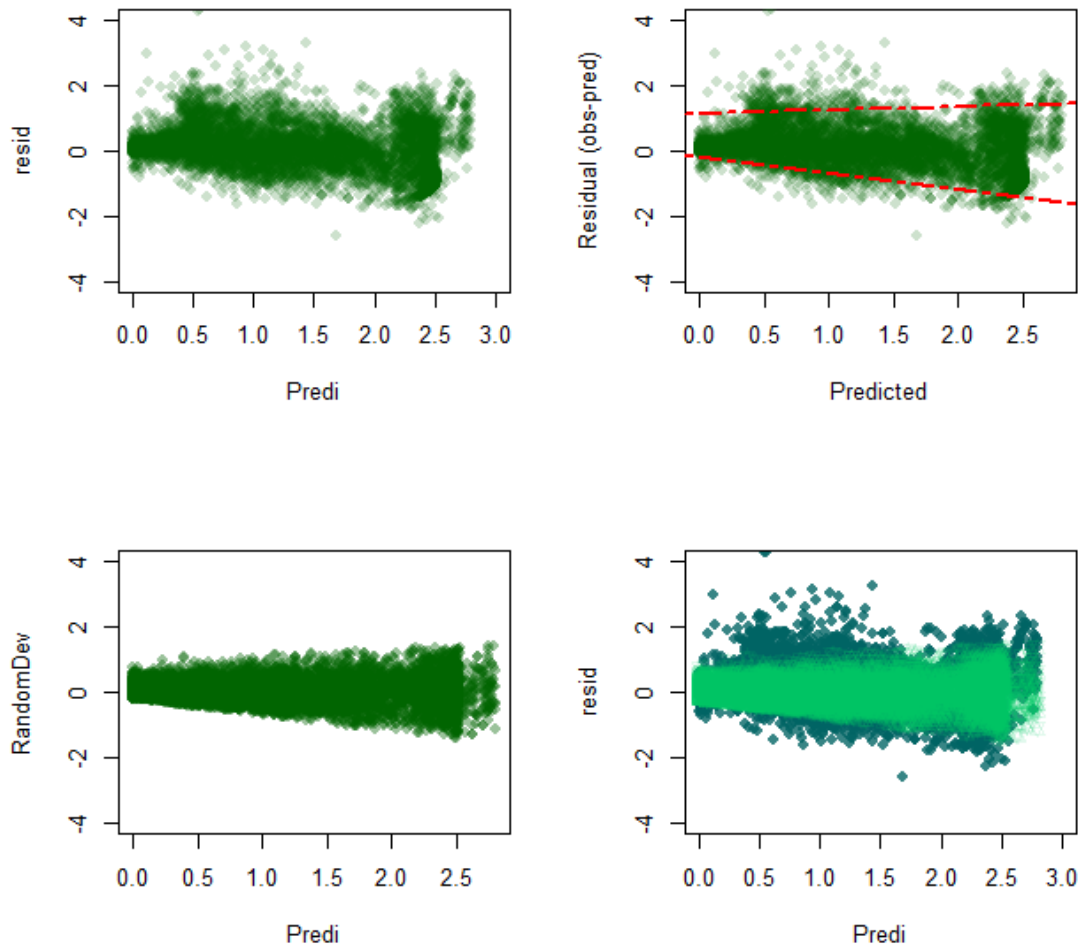


Figure 4-20: Sequential process of generating random deviation, from plotting residuals (top left), predicting bounds (top right), generating random deviation (bottom left) superimposed on the residual plot (bottom right)

The new predicted scatterplot showed a similarity to the observed natural deviation. It remained to be seen how these deviate result in an improvement of the increment model. The new predicted increment values with the random deviates added had to be compared to the increment values. This was done visually initially to see if it matches the structure of the increment-DBH scatter plots, with its distinctive pattern. **Figure 4-21** and **Figure 4-22** below shows the improvement of the structure from the observed increment for the Nelder and the Tweefontein sites respectively, with the predicted plots without deviation and the predicted plots with deviation.

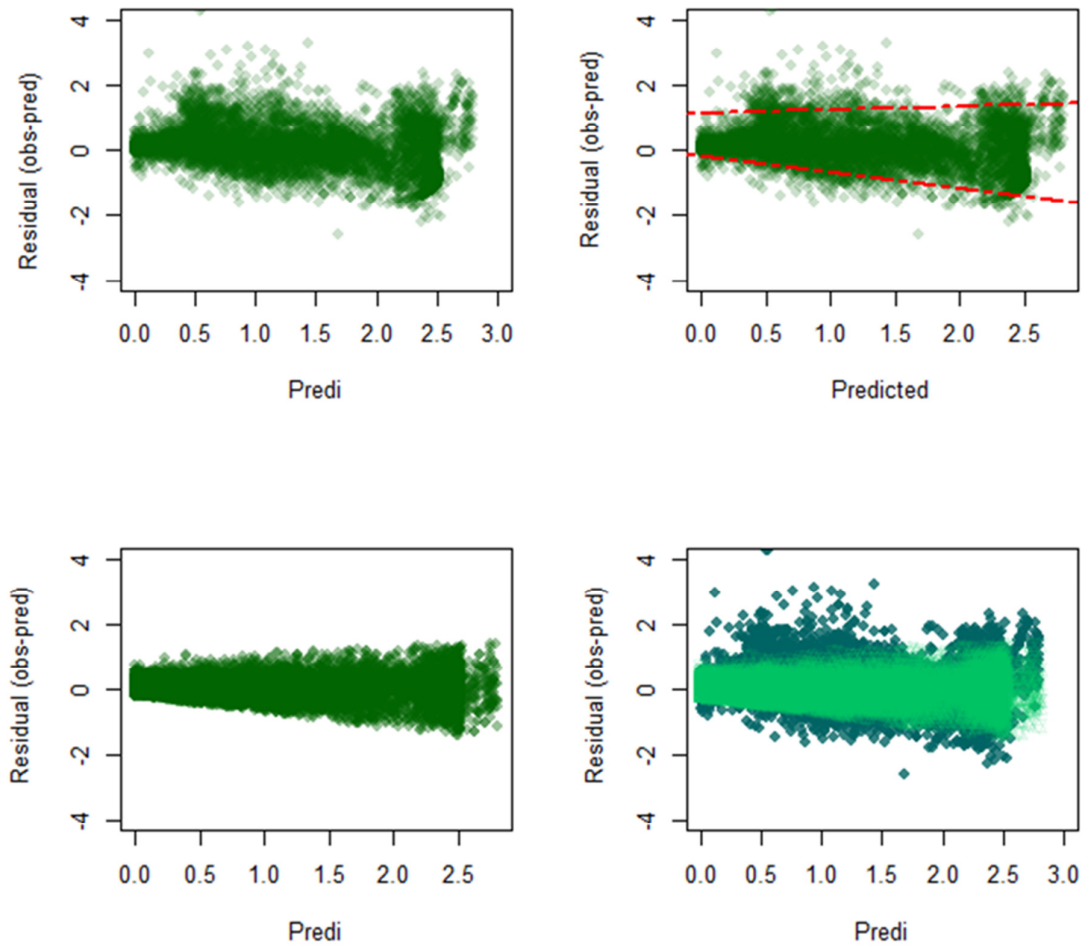


Figure 4-21: Observed and predicted scatterplots for the Tweefontein SS-CCT spacing trial, with black points representing the observed and green representing the predicted values. The left hand image represents the scatter of the average model – with the linear quantile in the top right, the bottom left and right images represent the scatter of the model with added modelled random variance

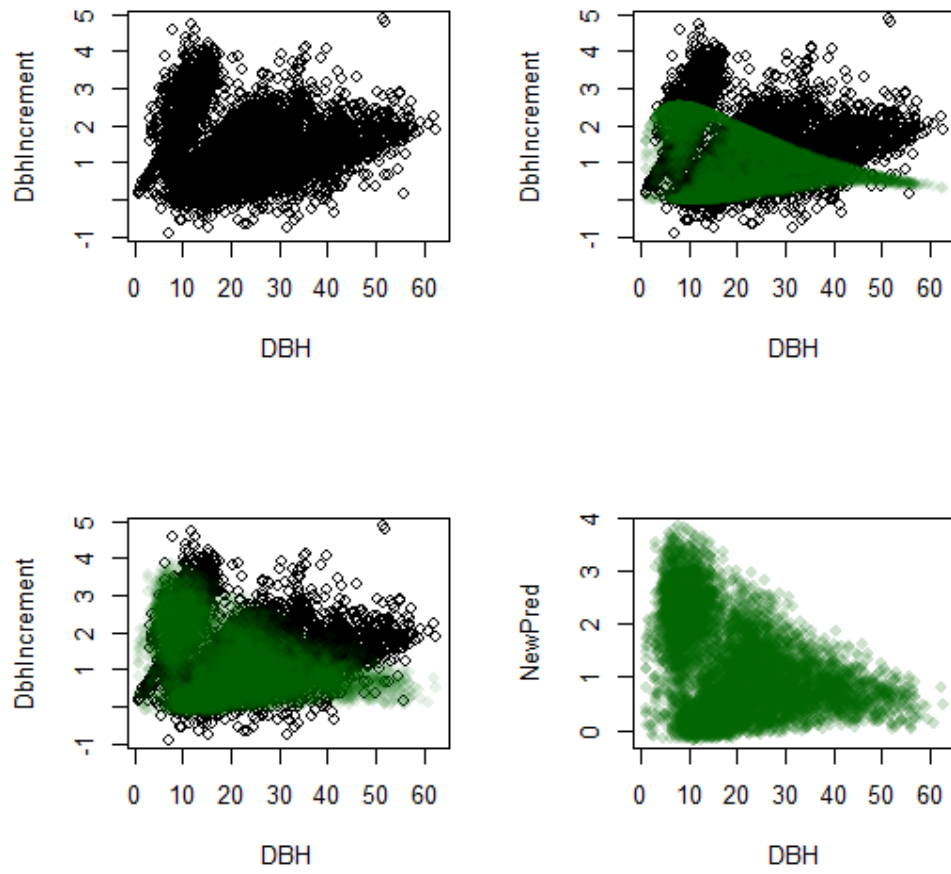


Figure 4-22: Observed and predicted scatterplots for the Lottering Nelder spacing trial, with black points representing the observed and green representing the predicted values. The top right image shows the deterministic model, the bottom left and right images represent the scatter of the model with added modelled random variance

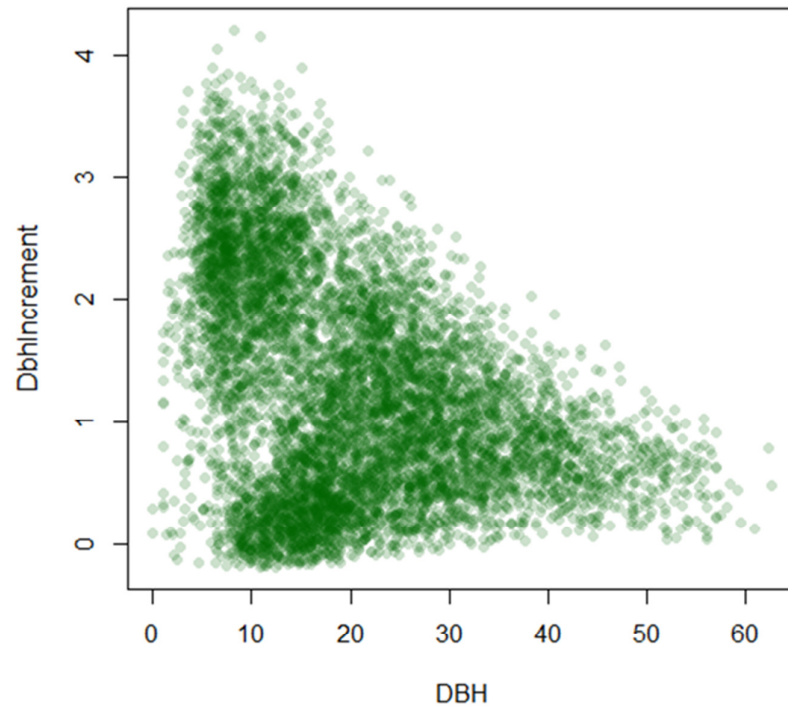


Figure 4-23: Residual scatter of the added random variance model for the parameterised sites

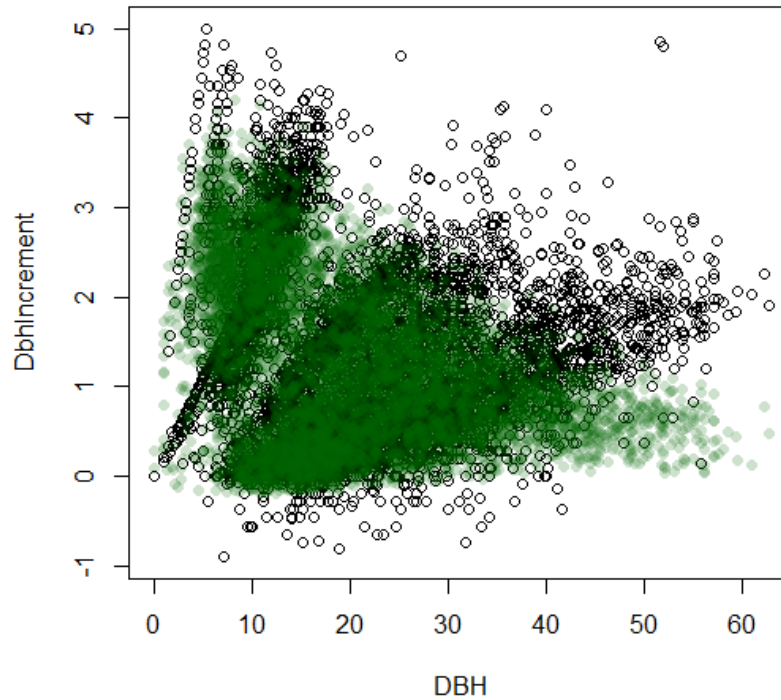


Figure 4-24: Residual scatter overlay on the observed increment scatter of the added random variance model for the parameterised sites. The black points represent the observed and the green points represent the predicted values from the stochastic model.

The new plots (**Figure 4-23 and 4-24**) seemed to mimic the observed increment quite well, matching the point clouds obtained from the competitive situation at different measurement ages, except for a depressed tail of points at higher DBH values (with lower predicted increment curves). It is now necessary to see how the new random deviates behave in a simulation over time. This is done in the validation section below.

Model Validation

Adding variability should prevent model from collecting or converging towards a mean around one DBH class, as was seen from the simulated model (Appendix D). The same simulation was done as in Section 4.6.4, this time with the random variance added to the prediction, shown in Appendix E with the end point at 26 years shown in **Figure 4-25**. Diameter increment is clearly overestimated at most ages, for both the model including variability and without.

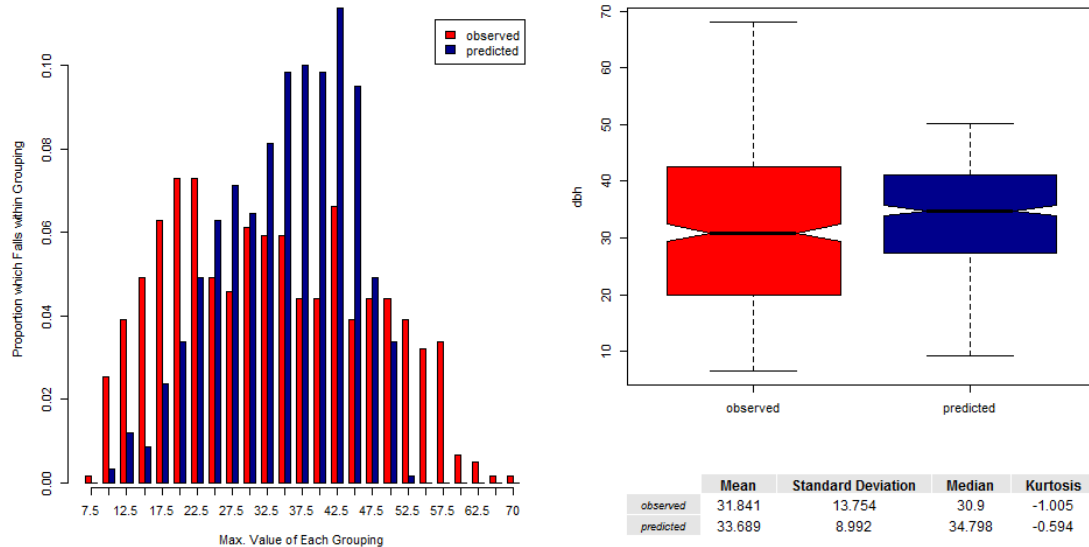


Figure 4-25: Observed (red bars) vs. simulated (blue bars) for the Nelder trial with random deviation added at each point. The model shows a clear overall overprediction of DBH, however with an improved distribution compared to the model without deviation (Figure 4-16).

From a diameter distribution point of view, the predicted model is clearly superior – although shifted to the right due to the overpredicted dbh originating from the deterministic model part. The model overall showed an overprediction for the Nelder trial, however with the extreme planted densities that exist (from 6700- 126 spha) it is necessary to see at which densities the model is not predicting well. For the end of the simulation run the following **Table 4-19** shows the average observed vs. predicted DBH values for the different Nelder rings (and corresponding stems per hectare) at 26 years:

Table 4-19: Average of observed and predicted values for the different Nelder plots, showing clearly that the model underperforms at extreme densities

Plot (ring)	SPHA (Planted)	Average of predicted DBH(cm)	Average of observed DBH (cm)	Difference
4	126	42.46	53.15	-10.69
5	167	43.18	51.40	-8.22
6	222	42.64	48.68	-6.04
7	295	39.56	44.93	-5.37
8	391	40.14	42.00	-1.85
9	520	37.56	36.48	1.08
10	691	36.37	34.65	1.71
11	917	34.16	31.94	2.22
12	1219	32.22	28.88	3.34
13	1619	29.99	23.86	6.13
14	2151	27.40	22.50	4.90
15	2857	25.79	19.67	6.12
16	3796	23.41	18.31	5.10
17	5043	22.94	16.40	6.54
18	6700	22.35	14.78	7.57

Table 4-19 shows that the model overpredicts at high densities and underpredicts at very low densities; however, at moderate densities at which most stands would be planted and managed under in plantation stands (highlighted in blue), the difference is relatively small. This difference is illustrated as a percentage in **Figure 4-26** below.

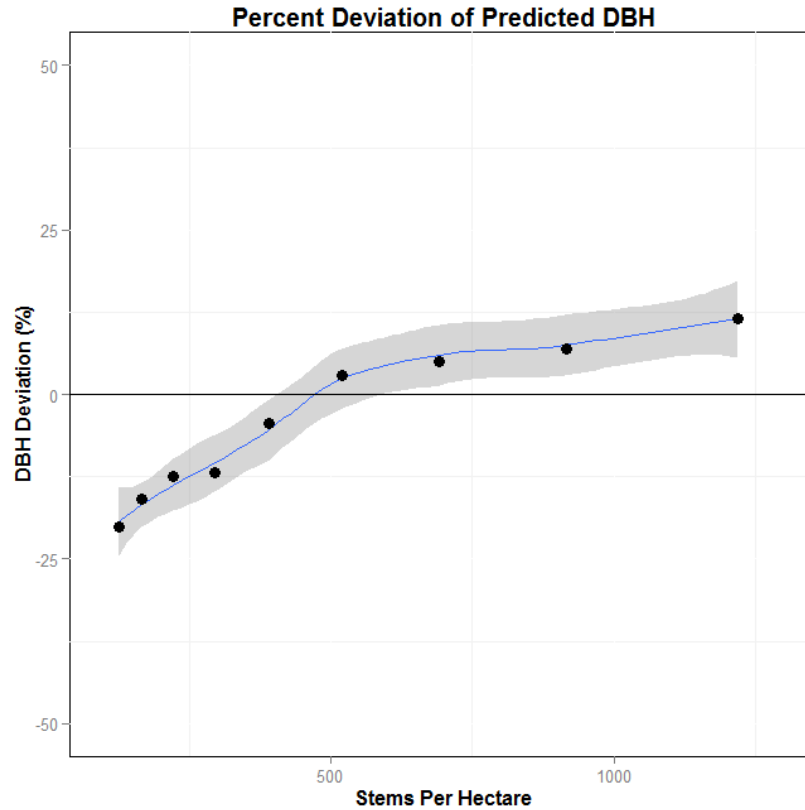


Figure 4-26: Deviation of predicted average diameter increment for a more realistic of planted stand densities for a plantation industry setup in the Nelder trial

4.8 Chapter conclusion

The main objective of this chapter was to model diameter increment in as a function of resource limitations and site classification for application of a methodological based approach for a specific dataset to outline problems and future work. This was split into the 6 sub-objectives which formed the outline of the working steps of the chapter. This main objective was achieved, however some key issues arose which will be discussed according to the six working steps/objectives of the Chapter.

The water index introduced in the first step proved not to be adequate to capture a large enough gradient between the two sites for modelling purposes. However, it did result in a shift between importance of the competition indices discussed later. Potential increment curves were applied based on a given DBH on different stands, which showed clear differences between sites. Incorporating site index as a site quality measure into the potential increment model resulted in a gradient of predicted potentials from site quality.

In order to include a modifier multiple distance-dependent competition indices were tested on two sites with available tree positions with the objective to identify a combination of

indices that showed small collinearity while explaining diameter increment well. The key idea behind this was to capture a gradient between overtopping (light limitation) and local crowding (edaphic limitations) modes of competition as highlighted by Seifert *et al.* (in press). The chosen competition indices reacted differently, but in order to identify and quantify the differences the behaviour and correlation between the indices were examined.

Differences in PCA biplots and improved fit with low collinearity resulted in the decision to choose the LBA and KKL indices in a combined model. Incorporating a competition index which captures more size symmetric competition (LBA) seemed, under the fitted model constructed above, to improve prediction compared to model relying purely on the standard SILVA competition index KKL. The relative importance of the KKL and LBA to the transformed model changed under different site conditions on the two sites, with the KKL increasing in importance in the less drought affected site. Incorporating both indices in one model has potential to capture changing competition symmetry under changing water availability - although this needs to be investigated under a much wider gradient of sites with varying water availability and perhaps of different years of the same site in a longitudinal study approach.

The results have shown that including the aridity index in the modifier equation presents potential for inclusion of a water index into the diameter increment prediction. Although the chosen approach was very simple, restricted by site and data availability, and reliable conclusions are difficult to make with regards to changing importance of CI models, it could be used for further parameterisation of SILVA. The ideal situation would be to combine a water index and weather data with a soil water balance model to estimate water availability, which was outside the scope of this thesis. In this study the longitudinal data aspect was not considered due to time and resource limitations, however, it must be noted that this may have a significant effect, which should be analysed in future studies.

The presented model matched competitive stages, achieving the objective of obtaining an age-independent model. The potential modifier model with the competition indices were still not able to capture the wide range of variability observed in diameter increment predictions; this lead to a skewed, narrow prediction of DBH distribution of predicted compared to observed distributions.

The use of a linear quantile regression for the truncation of a normal distribution between residuals, and a random prediction based on a normal distribution proved to be a robust methodology to incorporate natural variation. However, the captured variability was not large enough to mimic the variance in observed stands. It is the opinion of the author that the problem originates from the deterministic model, where the competition indices were still not sensitive enough to model diameter based solely on a potential and competition in the modifier. It is thought that including allometric relationships, such as diameter/height or crown dimensions into the model would greatly improve the situation. The structure of the KKL index may also not currently be suited to capture the competitive situation in the South

African context; changes could be made to the angle of the search cone which detects and counts potential competitors. This could be tested in a sensitivity analysis. The problem could also be related to the size of the trees – larger trees may be expected to deviate in the upper regions of the random spectrum for instance. This was not addressed in the model and represents a key issue for future work.

Further refinements to the stochastic model could be made by testing the normality assumption and by using a Weibull distribution for instance. The variation predicted also did not take into account density effects in the observed residual plot, which may affect the size of the variation of diameter increment under different degrees of competition (a higher variation at low competition for instance); this could be a consideration for an investigation for future studies, which may make the stochastic model more sensitive to different stand density ranges.

Visual analysis of the new predicted increments showed a good likeness to observed Increment-DBH scatterplots. Incorporating a variability to move into stochastic modelling drastically improved the DBH distribution problem, matching observed DBH values reasonably well. The model seemed to capture the average situation quite well.

This Chapter serves as an example of a methodological approach to simulate diameter increment combinations of distance dependent competition indices and stochastic modelling. Future studies can greatly improve the process by considering further sites with different water indices and site quality and by studying the shifting importance of competition indices under different water availability in order to increase the sensitivity of the model to changing climatic conditions.

Chapter 5: Conclusion and recommendations

This chapter provides a brief overview of what was achieved and recommendations for future work, as most of the issues have been discussed in the individual chapters already. This is divided into two sections: the work which was applied in this thesis (the potential height and diameter increment models in Chapters 3 and 4) and steps for further parameterisation for SILVA, which were not applied in this thesis.

5.1 Potential height modelling

The overall objective of this chapter was to model potential height and to see whether site index can be used as a predictor variable for potential height. In this sense the overall objective was achieved as it seemed to be a viable approach.

Nonlinear quantile regression proved to be a robust method for conditional median modelling, in this case of the potential height. The algorithm used seems to be less affected by outliers than other least squares estimation techniques. There were some limitations worth noting: for one, as with many models, it is bound by the observed values of the data represented, and may thus be unable to quantify the true potential. This limits the potential height model as it is then by definition unable to further model exceptional trees, which have a large effect on competition and neighbours. This can be remedied by the inclusion of stochastic modelling techniques, as applied in Chapter 4 whereby the allowance of a random deviation procedure could allow some trees to vary above the allocated potential after the application of a modifier.

Modelling site index using different fitting techniques provided potential new and interesting techniques for site index modelling. Nonlinear mixed effects modelling proved superior in predictive quality compared to the NLS and GNLS techniques for the data range acquired. Including the asymptote as the random effect provided the most robust and accurate inclusion of random effects into the NLME methodology.

Using site index as a predictor, the overall objective of the chapter, proved to be possible. Some validation and refinements to the process could be made, however it is thought that site index is able to model potential height quite well. A major concern would be the effect of density on site index, while this thesis used the South African site index definition as is, it could lead to substantial errors at extreme (very high or low) stand densities and is an issue that should be addressed in future work.

Future work should focus on a site quality predictor based on edaphic conditions (Esler 2012, Louw and Scholes 2002), which should improve the sensitivity of the model to site conditions and move the model into a hybrid modelling approach.

5.2 Diameter increment modelling

The overall objective of the chapter was, under data limitations, to present a methodology for future parameterisation and to highlight problems associated with this. In this sense, the objective was achieved; however improvements to this methodology can still be made.

Stratifying the indices based on the collinear relationships in order to obtain models with a combination of good predictive power to diameter increment and low collinearity to each other using PCA and VIF's was a novel and powerful approach.

As noted earlier, the discussion on the behaviour of competition indices on different site types must be taken very tentatively, as two sites does not represent a good gradient. However, some trends can be noted and the validation produced some interesting results, showing clearly that the importance of the different indices.

Including the Local Basal Area index, which captures a more growth symmetric competition structure representing edaphic factors, improved the model and often performed better than the KKL for *P. elliottii* under the water limited South African conditions.

It is necessary though to see how this changes for different growth periods, taking into account rainfall and water availability over different seasons and years to see whether the competitive nature of the trees can switch between symmetric and asymmetric competition. It is strongly suggested that a study be undertaken to understand the cross-sectional (site stratification) and temporal shifts in competition mode. This could for example be done in combination with a dendrochronological study.

5.3 Additions for model completion

As this thesis did not cover the entire modelling and simulation process done in Silva, further work must be done in the following areas, although no strong deductions can be made until they are tested in the South African growing context.

The omission of the species specific crown model and light transmission factor may have been a major factor in the behaviour of the KKL model. A model should be developed for South Africa to model the crown structure of *P. elliottii* and other species to be used in future parameterisations, which could be used in the competition indices and allometric equations to improve growth prediction. Height increment modelling based on potential height and the stand initialisation should be studied. Although height increment is not as strongly affected by competition as diameter increment, competition in addition to allometric models and site quality could be investigated. Mortality was not parameterised in this study. Some surrogates could be used, however a more size and competition based approach should be used for parameterisation into SILVA.

5.4 Overall thoughts

While the two components studied are very different in their model approaches and are not combined into a full model yet, the overall objectives for each were obtained. The above process in the thesis gives a good indication of the complexity of constructing single-tree growth models. However, the thesis resulted in a strong indication that this type of modelling process would be feasible, with the main focus to shift on completing the simulation structure of the entire growth modelling structure and to focus on shifts in competition index and competition modes under changing water availability. Problems encountered, and the solutions suggested, with the models tested could be of use for further research.

Chapter 6: References

- Akaike, H. (1972). Statistical predictor identification. *Ann. Inst. Math.*, 22, 203–207.
- Ackerman, S.A., Ackerman, P.A., Seifert, T. (2013). Effects of irregular stand structure on tree growth, crown extension and branchiness of plantation grown *Pinus patula*. *Southern Forests*, 75(4), 247–256.
- Botkin, D. B., Janak, J. F., & Wallis, J. R. (1972). Some Ecological Consequences of a Computer Model of Forest Growth. *Journal of Ecology*, 60(3), 849–872.
- Bragg, D. C. (2001). Potential relative increment (PRI): a new method to empirically derive optimal tree diameter growth. *Ecological Modelling*, 137, 77–92.
- Bredenkamp, B.V. (1984). The C.C.T. concept in spacing research – a review. Proceedings: Symposium on site and productivity of fast growing plantations. Vol 1:313-332.
- Bredenkamp, B. V. (1990). The Triple-S CCT design. In *Proceedings of a Symposium arranged by the Forest Mensuration and Modelling Working Group in collaboration with the Southern African Institute of Forestry and the Eucalyptus grandis Research Network on “Management of Eucalyptus grandis in South Africa”* (pp. 198–205).
- Bredenkamp, B.V. (1993). Top height: a definition for use in South Africa. *Southern African Forestry Journal*, 167: 55.
- Brown, G. (1965). *Point density in stems per acre* (p. 12).
- Cade, B., & Noon, B. (2003). A gentle introduction to quantile regression for ecologists. *Front Ecol Environ*, 1(8), 412– 420.
- Chertov, O. G. (1990). SPECOM - A Single Tree Model of Pine Stand/Raw Hums Soil Ecosystem. *Ecological Modelling*, 50, 107–132.
- Corral-Rivas, J., Sanchez Orois, S., Kotze, H. and Gadow von, K. (2009). Testing the suitability of the Nepal-Somers stand table projection method for *Eucalyptus grandis* plantations in South Africa. *Southern Forests* 2009, 71 (3): 207-214.
- Dalgaard, P. (2008). *Introductory Statistics with R*. New York: Springer.
- Du Toit, B. (2012). Matching site, species and silvicultural regime to optimise the productivity of commercial softwood species in Southern Africa. In *South African Forestry Handbook* (5th ed., pp. 43–50). Southern African Institute of Forestry.

Dye, P. ., Jacobs, S., & Drew, D. (2004). Verification of 3-PG growth and water-use predictions in twelve Eucalyptus plantation stands in Zululand, South Africa. *Forest Ecology and Management*, 193 (1-2), 197–218. doi:10.1016/j.foreco.2004.01.030

Dzierson H, Mason EG. (2006). Towards a nationwide growth and yield model for radiata plantations in New Zealand. *Can J For Res.* 36: 2533-2543.

Ek, A., & Dudek, A. (1980). Development of individual tree based stand growth simulators: progress and applications.

Esler, B. (2012). *On the development and application of indirect site indices based on edaphoclimatic variables for commercial forestry in South Africa*. Stellenbosch university.

Fairbanks, D.H.K., Scholes, R.J. (1999). South African Country Study on Climate Change: Vulnerability and Adaptation Assessment for Plantation Forestry. Prepared for: National Research Facility.

Fang, Z., & Bailey, R. L. (2001). Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *Forest Science*, 47(3), 287–300.

Fekedulegn, D., Siurtain, M. P. Mac, Colbert, J. J., Siurtain, M., & Parameter, J. J. (1999). Parameter Estimation of Nonlinear Growth Models in Forestry. *Silva Fennica*, 33(4), 327–336.

Gea-Izquierdo, G., Cañellas, I., & Montero, G. (2008). Site index in agroforestry systems: age-dependent and age-independent dynamic diameter growth models for *Quercus ilex* in Iberian open oak woodlands. *Canadian Journal of Forest Research*, 38(1), 101–113.

Grömping, U. (2006). Relative Importance for Linear Regression in R : The Package relaimpo. *Journal of Statistical Software*, 17(1).

Hegy, F. (1974). A simulation model for managing Jack-pine stands. In J. Fries (Ed.), *Growth models for tree and stand simulation* (pp. 74–90). Royal College of Forest, Stockholm, Sweden.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441.

Kassier, H. W. (1993). *Dynamics of diameter and height distributions in even-aged pine plantations*. University of Stellenbosch.

Koenker, R. (2006). Quantile regression in r: a vignette. Retrieved July 16, 2012, from <http://www.econ.uiuc.edu/~roger/research/rq/vig.pdf>

Koenker, R., & Bassett, G. (1978). No Title. *Econometrica*, 46(1), 33–50.

- Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, 15(4), 143–156. doi:10.1257/jep.15.4.143
- Komarov, A., Chertov, O., Zudin, S., Nadporozhskaya, M., Mikhailov, A., Bykhovets, S., ... Zoubkova, E. (2003). EFIMOD 2—a model of growth and cycling of elements in boreal forest ecosystems. *Ecological Modelling*, 170(2-3), 373–392.
- Kotze, H., Kassier, H. W., Fletcher, Y., & Morley, T. (2012). Growth Modelling and Yield Tables. In B. Bredenkamp & S. Upfold (Eds.), *South African Forestry Handbook* (5th ed., pp. 175–219). Southern African Institute of Forestry.
- Landsberg, J. J., & Waring, R. H. (1997). A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management*, 95(3), 209–228.
- Lee, W.K. and Gadow, K.v. (1997). Iterative Bestimmung der Konkurrenzbaume in *Pinus densiflora* Beständen. *AFJZ* 168(3/4): 41-44.
- Lekwadi, S. O., Nemesova, a., Lynch, T., Phillips, H., Hunter, a., & Mac Siúrtáin, M. (2012). Site classification and growth models for Sitka spruce plantations in Ireland. *Forest Ecology and Management*, 283, 56–65.
- Louw, J. H., & Scholes, M. (2002). Forest site classification and evaluation : a South African perspective. *Forest Ecology and Management*, 171, 153–168.
- McKee, T., Doesken, N., & Kleist, J. (1995). Drought monitoring with multiple time scales. In *9th Conference on Applied Climatology* (pp. 233–236). Dalas.
- Munro, D. (1974). Forest growth models - a prognosis. In J. Fries (Ed.), *Growth models for tree and stand simulation* (Research N., pp. 7–21). Stockholm, Sweden.
- Myers, R. H. (1986). *Classical and Modern Regression with Applications*. Boston: PWS Publishers.
- Nelder, J. A. (1962). New kinds of systematic designs for spacing experiments. *Bometrics*, 18, 283–309.
- O'Connor, A.J. (1935). Forest research with special reference to planting distances and thinning. British Empire Forestry Conference, South Africa. 30 pp.
- Odhiambo, B., Meincken, M., & Seifert, T. (2014). The protective role of bark against fire damage: a comparative study on selected introduced and indigenous tree species in the Western Cape, South Africa. *Trees*, DOI 10.1007/s00468-013-0971-0
- Palmer, W. (1965). *Meteorological Drought. Research Paper 45*. Washington DC.

- Payandeh, B. (1974). Nonlinear site index equations for several major Canadian forest timber species. *For. Chron* 50, 194-196.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Porte, A., & Bartelink, H. H. (2002). Modelling mixed forest growth : a review of models for forest management. *Ecological Modelling*, 150, 141–188.
- Poynton, R. J. (1979). *Tree Planting in Southern Africa, Volume 1, The Pines*. Pretoria: Department of Forestry.
- Pretzsch, H, Biber, P., & Dursky, J. (2002). The single tree-based stand simulator SILVA : construction , application and evaluation. *Forest Ecology and Management*, 162, 3–21.
- Pretzsch, Hans. (2009). *Forest Dynamics, Growth and Yield* (p. 664). Berlin and Heidelberg: Springer.
- Pretzsch, Hans, & Biber, P. (2010). Size-symmetric versus size-asymmetric competition and growth partitioning among trees in forest stands along an ecological gradient in central Europe. *Canadian Journal of Forest Research*, 40(2), 370–384.
- Reed, D. D., Jones, E. A., Tome, M., & Araujo, M. C. (2003). Models of potential height and diameter for Eucalyptus globulus in Portugal. *Forest Ecology and Management*, 172, 191–198.
- Ritchie, M. W., & Hann, D. W. (1990). Equations for Predicting the 5-Year in Southwest Oregon Height Growth of Six Conifer Species, (March).
- Rötzer T, Seifert T, Pretzsch H. (2009). Above and below ground carbon dynamics in a mixed beech and spruce stand influenced by environmental changes. *European Journal of Forest Research* 128(2): 171 – 182.
- Rötzer T, Seifert T, Gayler S, Priesack E, Pretzsch H. (2012). Effects of stress and defence allocation defence on tree growth: Simulation results at the tree and stand level. In: Matyssek R, Schnyder H, Ernst D, Munch J-C, Oßwald W, Pretzsch H (eds) *Growth and Defence in Plants: Resource Allocation at Multiple Scales*. Ecological Studies 220, Springer. 401-432.
- Schwinning, S., & Weiner, J. (1998). Mechanisms determining the degree of size asymmetry in competition among plants. *Oecologia*, 113(4), 447–455.
- Seifert, T., & Seifert, S. (2014). Modelling and simulation of tree biomass. In T. Seifert (Ed.), *Bioenergy from Wood: Sustainable Production in the Tropics* (Managing F., pp. 42–65). Springer.

Seifert, T., Seifert, S., Seydack, A., Durrheim, G., & von Gadow, K. (n.d.). Competition effect in an afrotemperate forest. *Forest Ecosystems (in review)*

Seifert, T. (2007). Simulating the extent of decay caused by *Heterobasidion annosum* s. l. in stems of Norway spruce. *Forest Ecology and Management*, 248(1-2), 95–106.

Seifert, T., & Pretzsch, H. (2004). Modeling growth and quality of Norway spruce (*Picea abies* Karst.) with the growth simulator SILVA. In *Proceedings of the 4th Workshop "Connection between Forest Resources and Wood Quality: Modelling Approaches and Simulation Software"* (pp. 562–574). IUFRO Working Party S5.01.04, Harrison Hot Springs, British Columbia, Canada.

Seifert T, du Toit B, Jooste GH, Seibold S, Durrheim G, Seydack A (in press), Towards an individual tree forest growth model for natural forests in South Africa: Testing competition indices to model individual tree diameter growth in Afro-temperate forests. 5th Natural Forests and Woodlands Symposium, Richards Bay, South Africa, 10.–14. April 2011.

Seifert S, Utschig, H. (2002). Standraum und Wachstum - Voronoi-Diagramme zur Bestimmung des Wachstums in Beständen. Bericht des Forschungszentrums Waldökosysteme, Universität Göttingen, Reihe B(68), 111-112.

Steneker, G. A., & Jarvis, J. M. (1963). A preliminary study to assess competition in a white spruce – trembling aspen stand. *Forestry Chronicles*, 39, 334–336.

Stoll, P., Weiner, J., Muller-Landau, H., Müller, E., & Hara, T. (2002). Size symmetry of competition alters biomass-density relationships. *Proceedings. Biological sciences / The Royal Society*, 269(1506), 2191–5.

Team, R. C. (2013). R: A Language and Environment for Statistical Computing.

Thornthwaite, C. (1948). An approach towards the rational classification of climate. *Geographical Review*, 38, 55–94.

UNEP. (1992). *World Atlas of Desertification*. London: Edward Arnold.

Uhl E, Metzger H-G, Seifert T (2006) Dimension und Wachstum von solitären Buchen und Eichen. Tagungsband der Jahrestagung der Sektion Ertragskunde im Deutschen Verband Forstlicher Forschungsanstalten, 47 – 53.

Vanclay, J. K. (1995). Growth models for tropical forests: A synthesis of models and methods. *Forest Science*, 41(1), 4–42.

van Staden, Vida, Erasmus, B.F.N., Roux, J., Wingfield, M.J., van Jaarsveld, A.S. (2004). Modelling the spatial distribution of two important South African plantation forestry pathogens. *Forest Ecology and Management* 187: 61–73.

Von Gadow, K., & Bredenkamp, B. (1992). *Forest management* (p. 151). Pretoria: Academia.

Warburton, M., Schulze R. (2006). Climate change and the South African commercial forestry sector: an initial study. Report to Forestry SA. *ACRUcons Report 54*, December 2006.

Wichmann, L. (2001). Annual Variations in Competition Symmetry in Even-aged Sitka Spruce. *Annals of Botany*, 88(1), 145–151. doi:10.1006/anbo.2001.1445

Wykoff, W.R., Crookston, N.L. and Stage, A.R. (1982). User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Rep. No. INT-133.

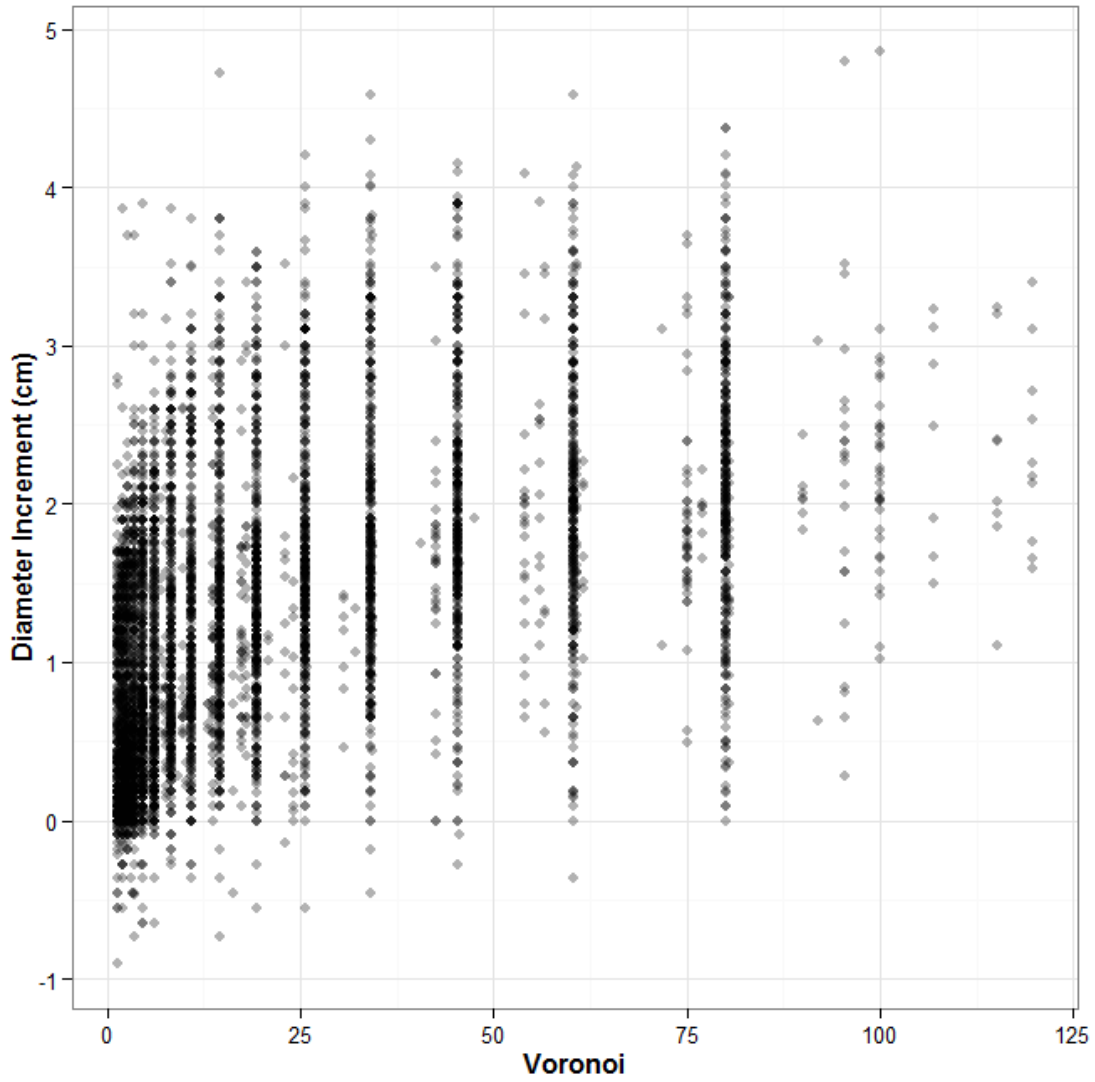
Chapter 7: Appendices

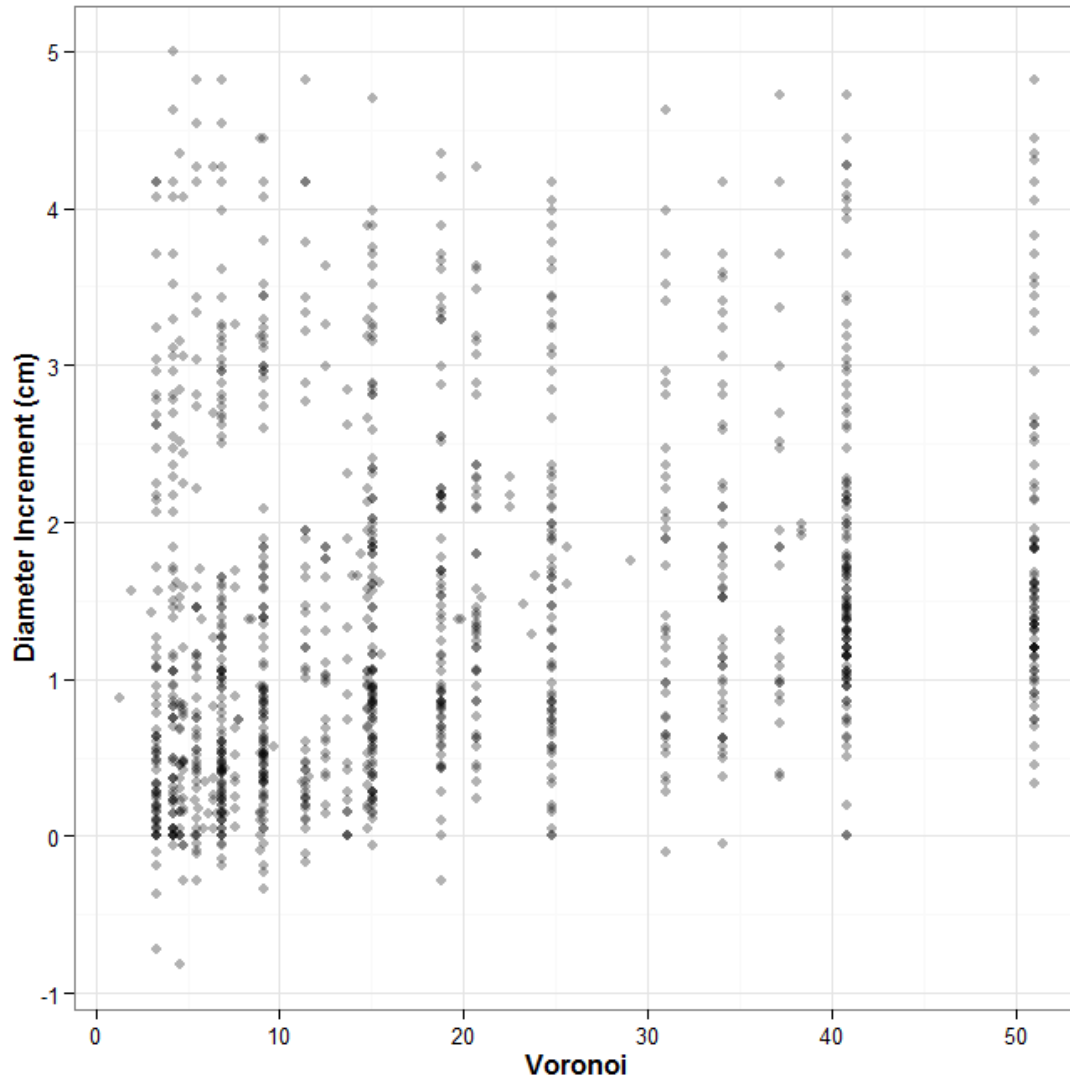
Appendix A: NLME random effects anova table

Density/Plot	Model	Parameter	Convergence (Y,N)	df	AIC	BIC	logLik	Test	L.Ratio	p-value
124	1	a,b,c	N							
	2	a,b	N							
	3	a,c	Y	9	55.72202	65.94146	-18.86101			
	4	a	Y	7	51.58846	59.53692	-18.79423			
	5	b,c	Y	9	55.72365	65.94309	-18.86182			
	6	b	N							
	7	c	Y	7	51.72368	59.67214	-18.86184	5 vs 7	3.11E-05	1
247	1	a,b,c	Y	12	109.375	132.0768	-42.6875			
	2	a,b	N							
	3	a,c	Y	9	103.87055	120.8969	-42.93528	1 vs 3	0.4955555	0.9199
	4	a	N							
	5	b,c	Y	9	103.37498	120.4014	-42.68749			
	6	b	Y	7	99.85563	113.0984	-42.92781	5 vs 6	0.4806494	0.7864
	7	c	Y	7	99.85563	113.0984	-42.92782			
371	1	a,b,c	N							
	2	a,b	N							
	3	a,c	N							
	4	a	Y	7	148.517	164.8322	-67.25852			
	5	b,c	Y	9	151.3502	172.3268	-66.67508			
	6	b	Y	7	149.1949	165.51	-67.59743	5 vs 6	1.8447	0.3976
	7	c	Y	7	148.3134	164.6285	-67.15669			
494	1	a,b,c	Y	12	174.6198	204.4834	-75.30988			
	2	a,b	Y	9	168.6184	191.0161	-75.30919	1 vs 2	0.001373	1
	3	a,c	Y	9	174.2355	196.6333	-78.11777			
	4	a	Y	7	165.1243	182.5447	-75.56214			
	5	b,c	N							
	6	b	Y	7	163.8356	181.2561	-74.9178	6 vs 3	6.399935	0.0408
	7	c	Y	7	170.2361	187.6566	-78.11805			
741	1	a,b,c	N							
	2	a,b	Y	9	269.5768	295.922	-125.7884			
	3	a,c	Y	9	269.5766	295.9219	-125.7883			
	4	a	Y	7	265.5755	286.0662	-125.7877			
	5	b,c	Y	9	272.0675	298.4128	-127.0338			
	6	b	Y	7	268.0678	288.5586	-127.0339	6 vs 5	0.000245222	0.9999
	7	c	Y	7	269.1501	289.6409	-127.5751			

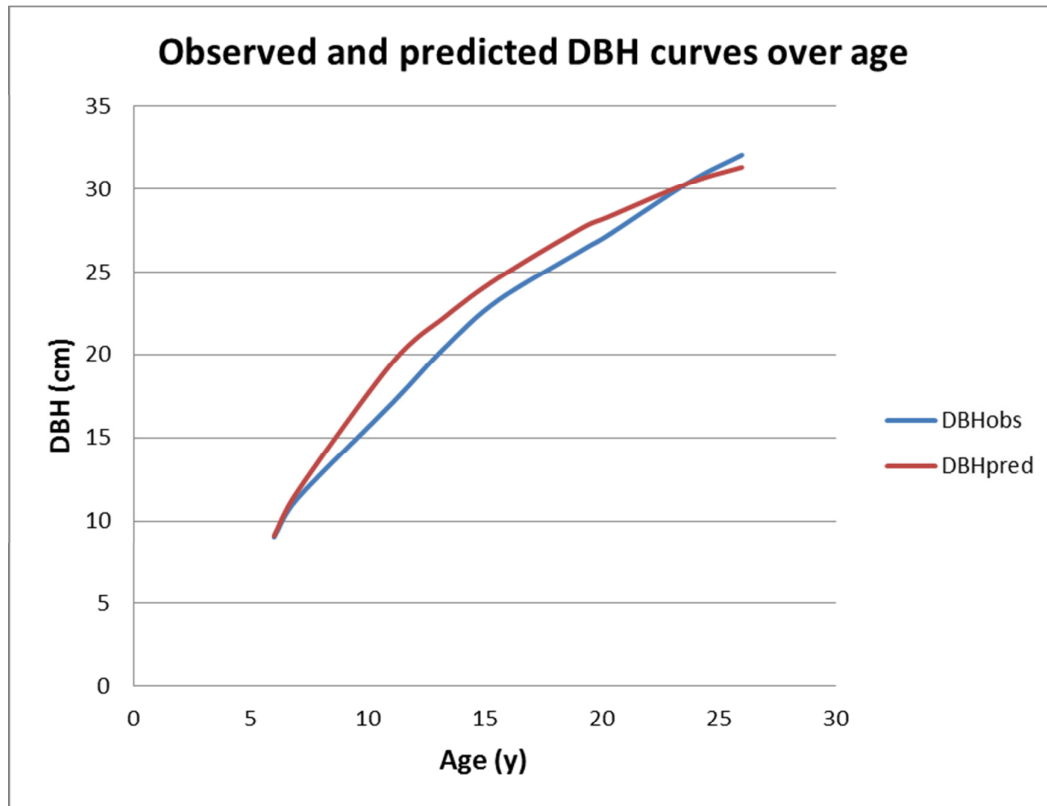
Density/Plot	Model	Parameter	Convergence (Y,N)	df	AIC	BIC	logLik	Test	L.Ratio	p-value
988	1	a,b,c	Y	12	331.6635	369.0794	-153.8317			
	2	a,b	Y	9	325.6821	353.7441	-153.8411	1 vs 2	0.018633177	0.9993
	3	a,c	Y	9	325.675	353.7369	-153.8375			
	4	a	Y	7	321.6714	343.4974	-153.8357	3 vs 4	0.003530949	0.9982
	5	b,c	N							
	6	b	N							
	7	c	N							
1483	1	a,b,c	Y	12	498.8548	540.1638	-237.4274			
	2	a,b	Y	9	492.8549	523.8366	-237.4274	1 vs 2	2.27E-05	1
	3	a,c	N							
	4	a	Y	7	488.8548	512.9518	-237.4274			
	5	b,c	N							
	6	b	Y	7	488.8548	512.9518	-237.4274	6 vs 2	2.47E-05	1
	7	c	Y	7	488.285	512.382	-237.1425			
2965	1	a,b,c	nlme.fit1	12	763.7114	812.588	-369.8557			
	2	a,b	nlme.fit2	9	757.7237	794.3811	-369.8618	1 vs 2	0.012269636	0.9996
	3	a,c	nlme.fit3	9	757.7107	794.3681	-369.8553			
	4	a	nlme.fit7	7	753.7078	782.2191	-369.8539			
	5	b,c	nlme.fit4	9	758.2332	794.8906	-370.1166			
	6	b	nlme.fit5	7	754.2328	782.7441	-370.1164	5 vs 6	0.000403051	0.9998
	7	c	nlme.fit6	7	744.5137	773.025	-365.2569			

Appendix B: Voronoi polygon increment relationship for the Nelder and Tweefontein sites respectively

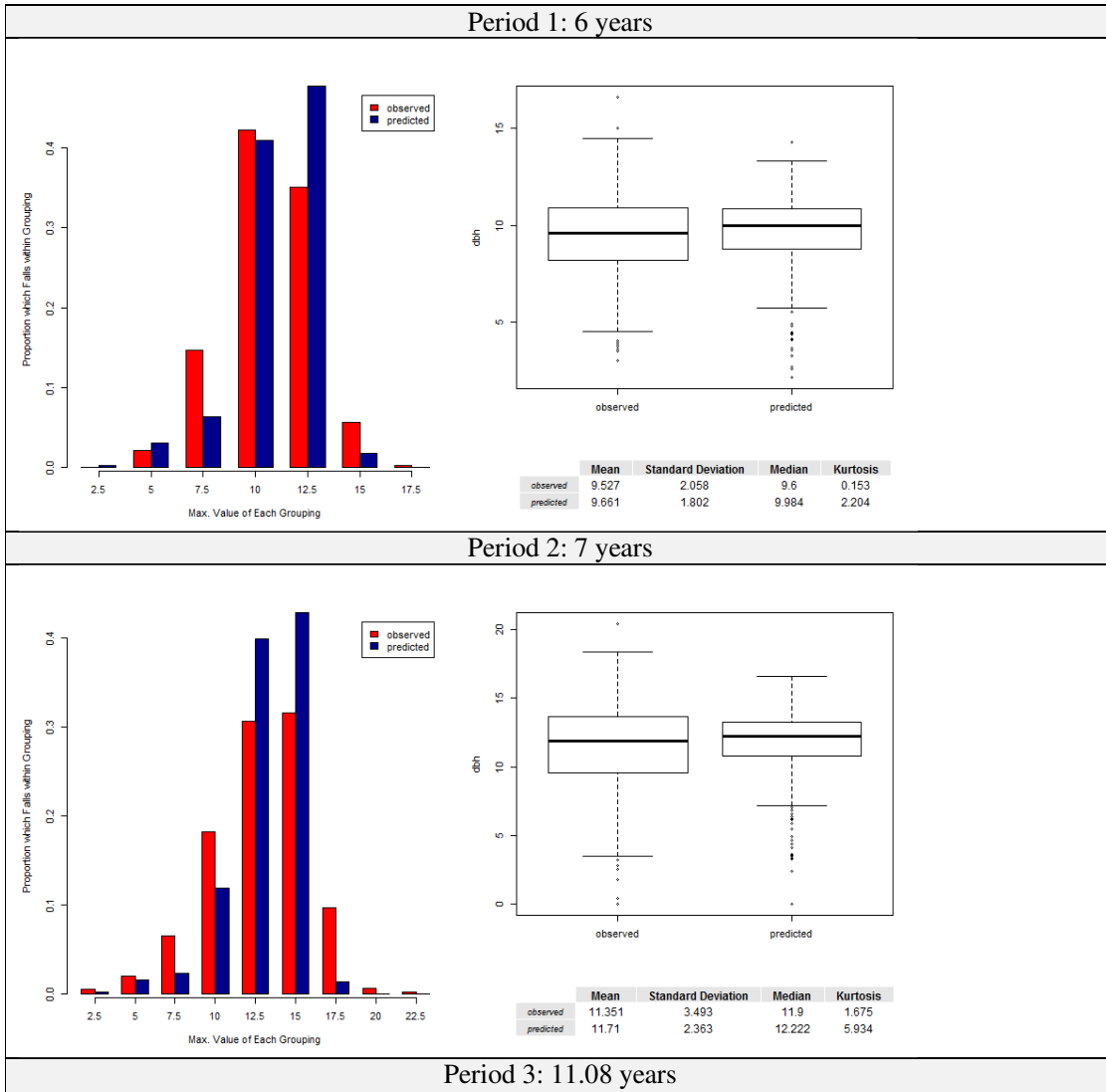


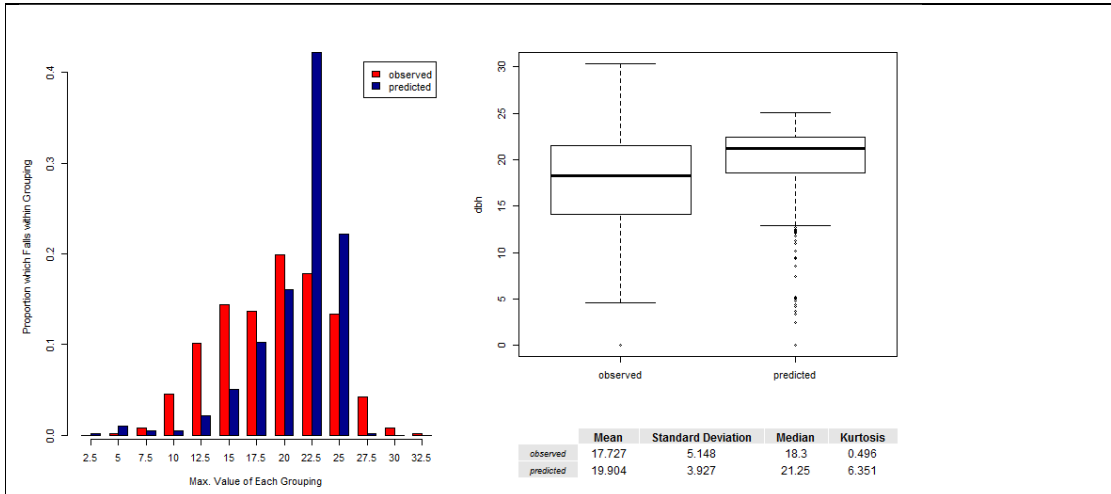


Appendix C: Observed and predicted DBH for the deterministic model

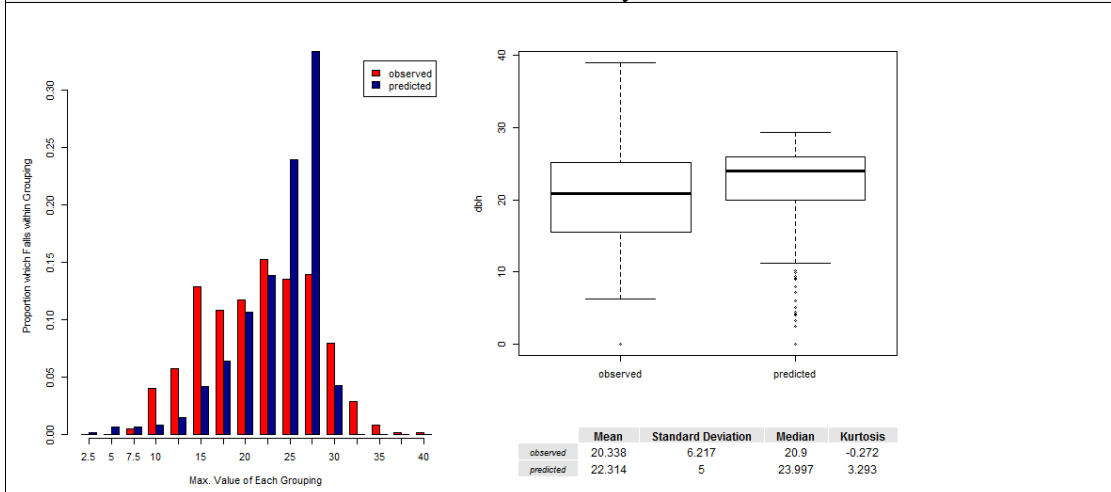


Appendix D: Simulation steps for the deterministic model

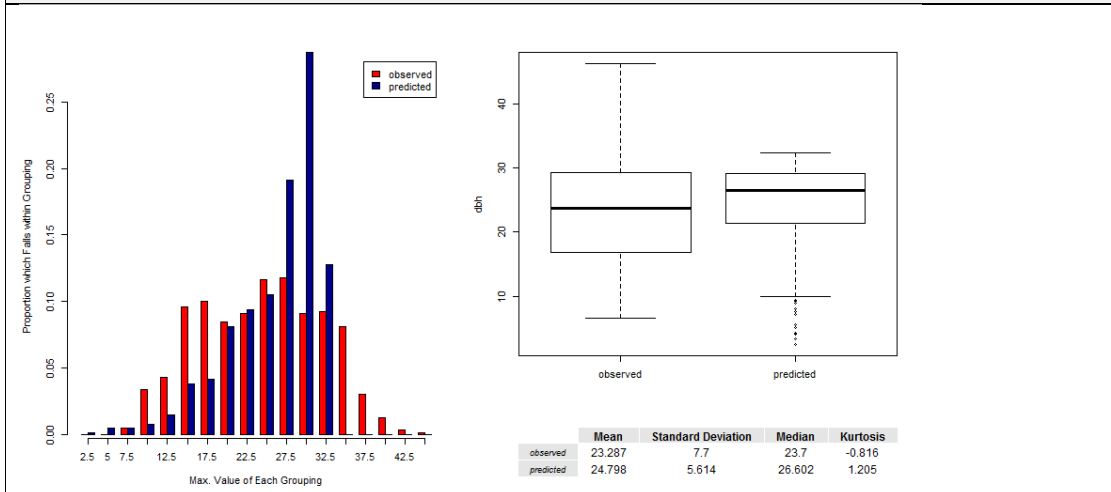




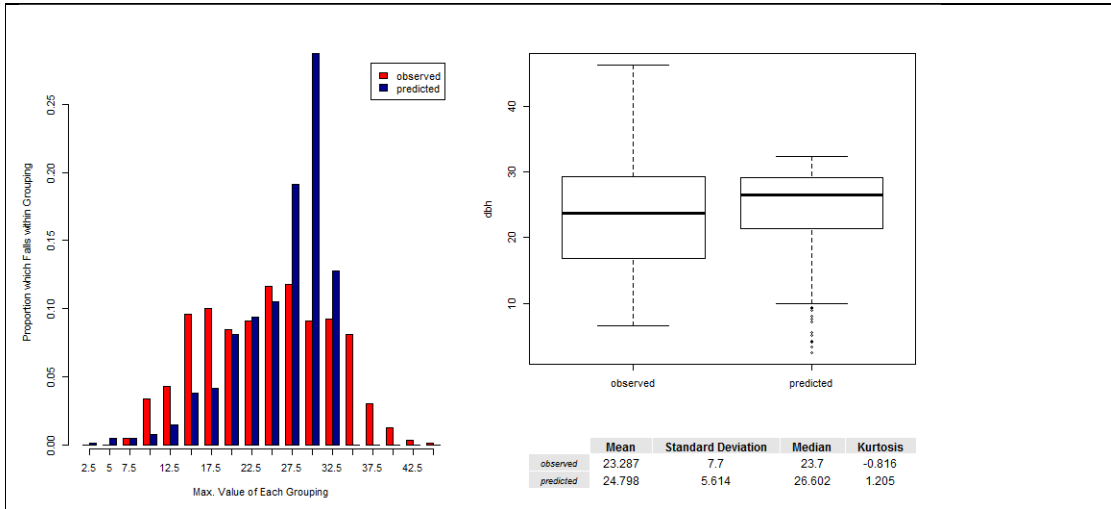
Period 4: 13.33 years



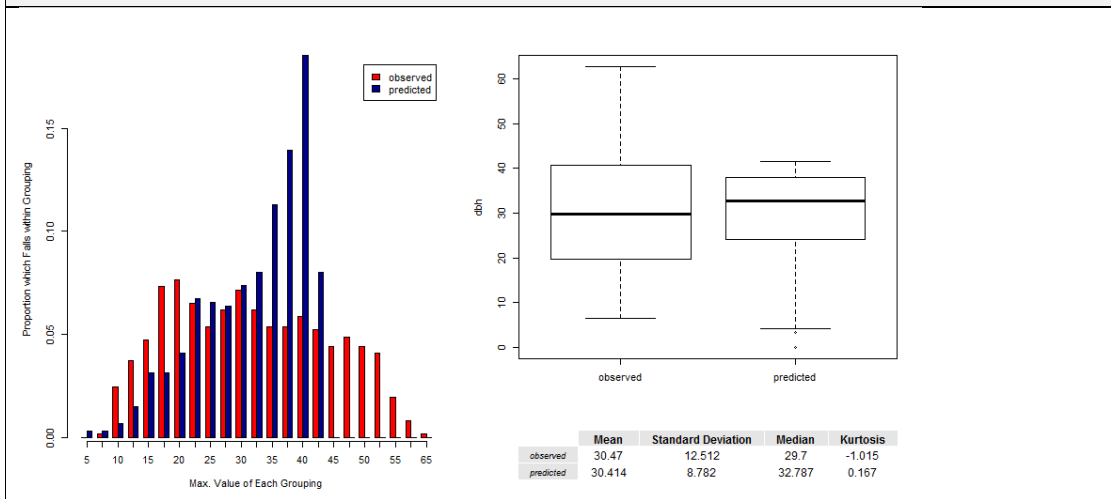
Period 5: 15.58



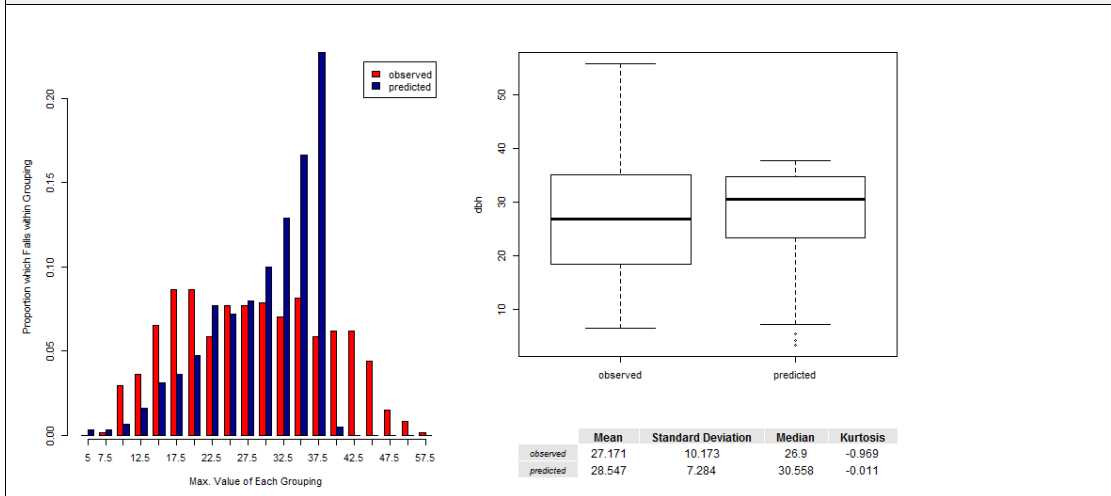
Period 6: 19.17



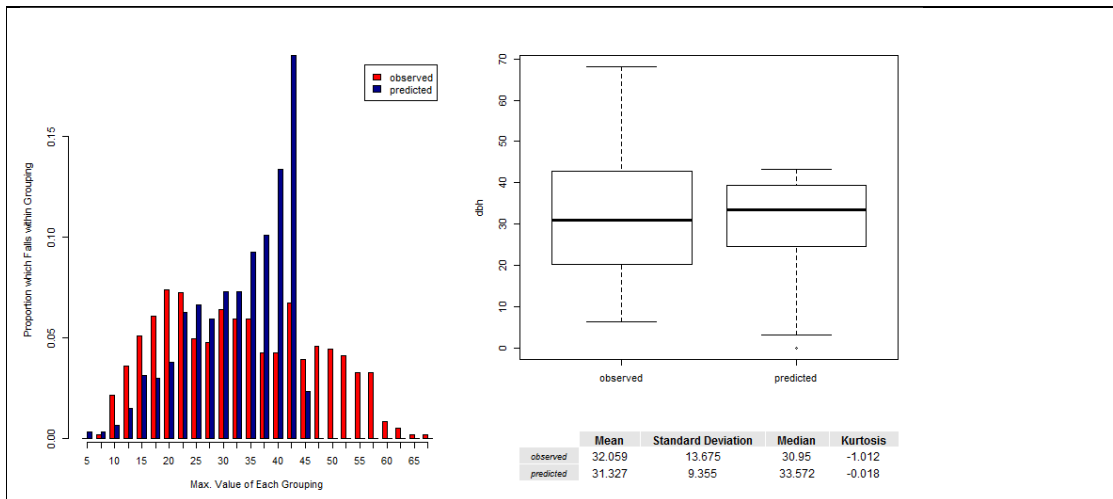
Period 7:20.25



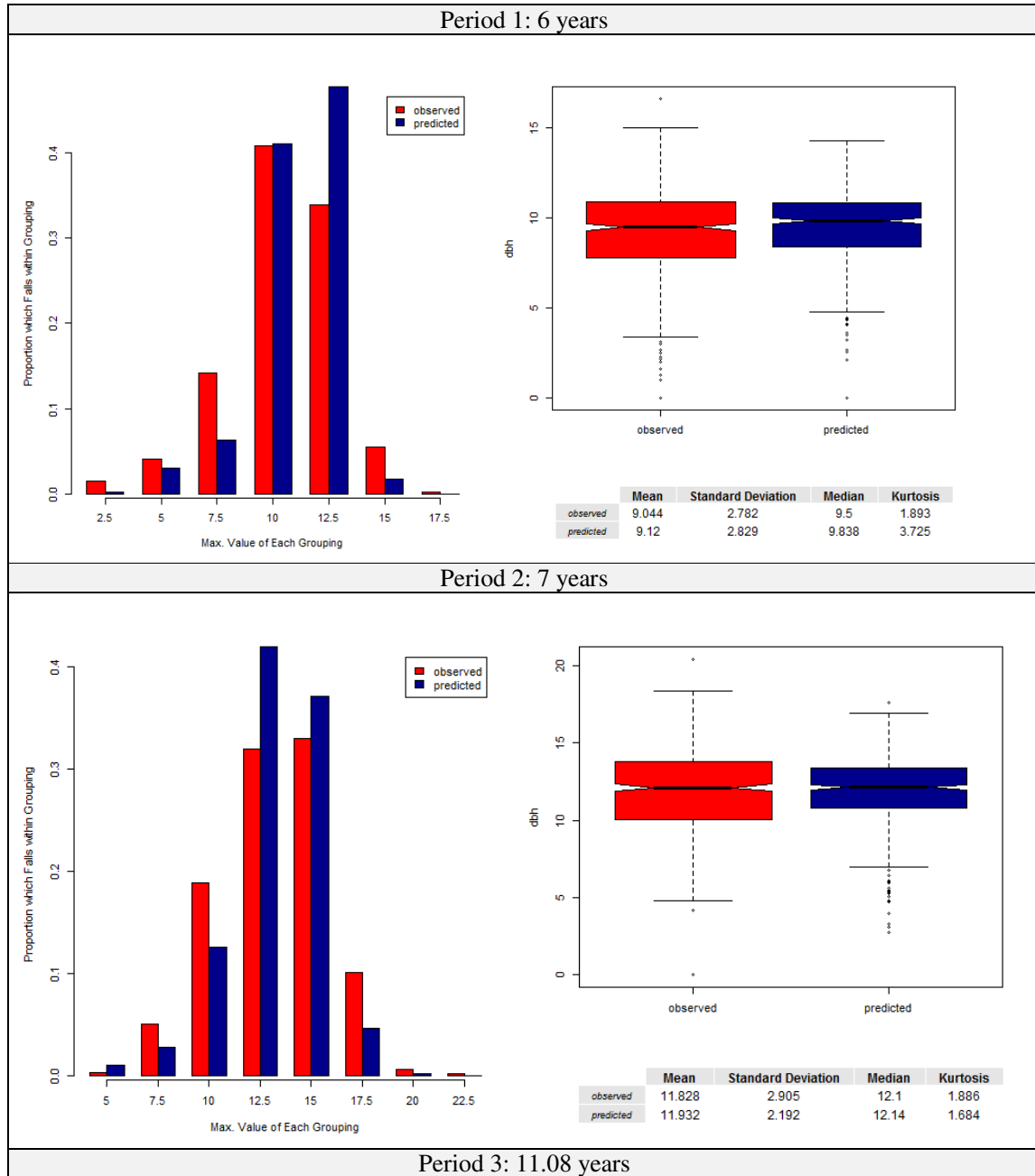
Period 8:23.83

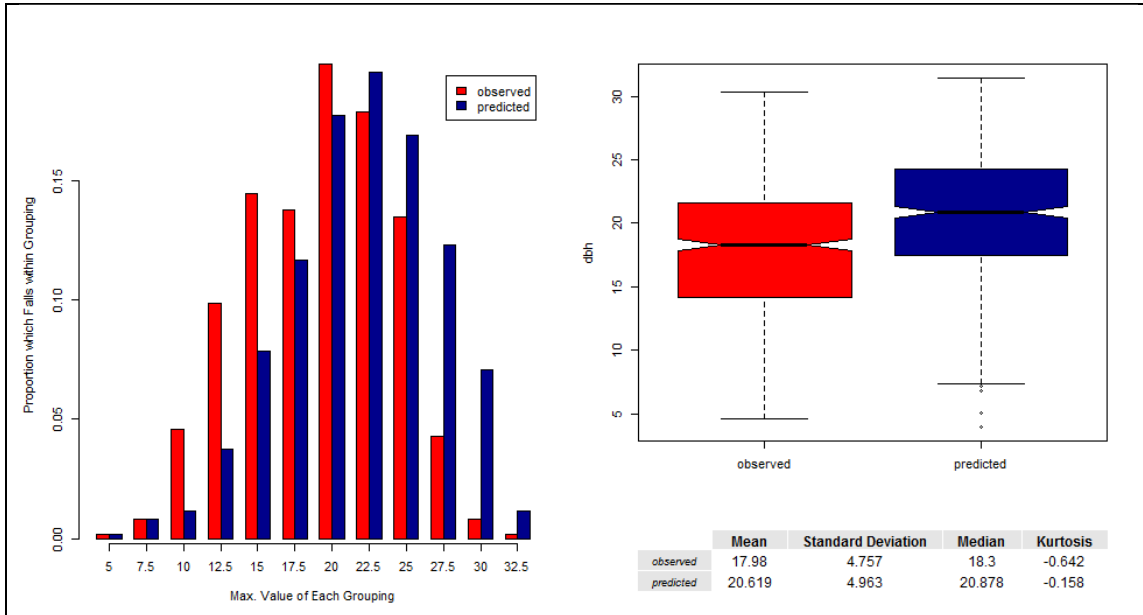


Period 9:26

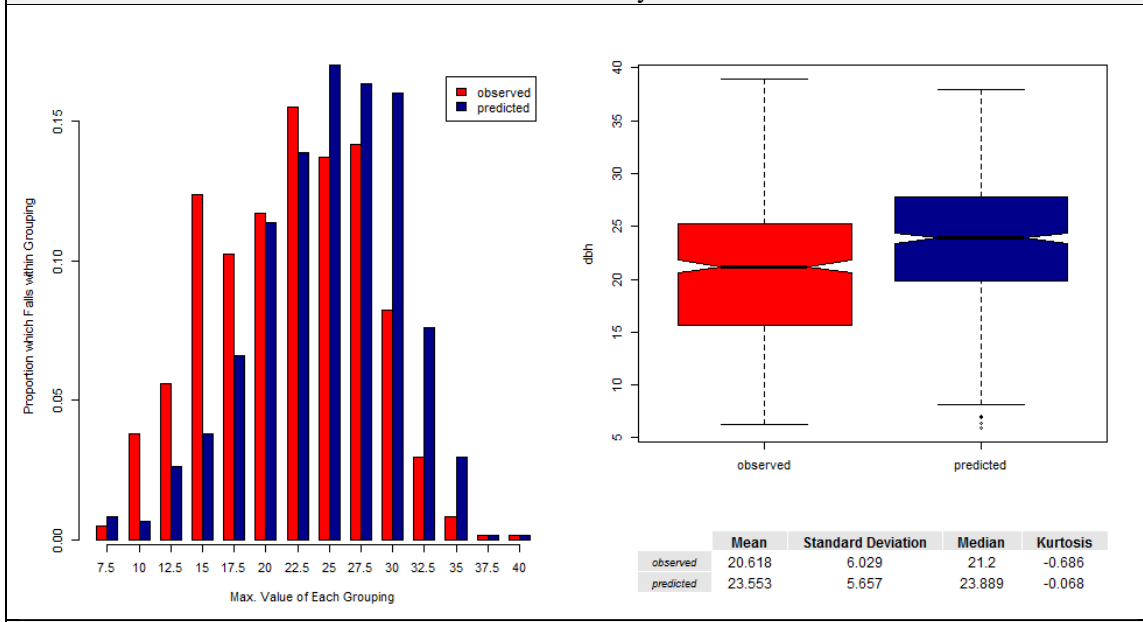


Appendix E: Simulation steps for the stochastic model

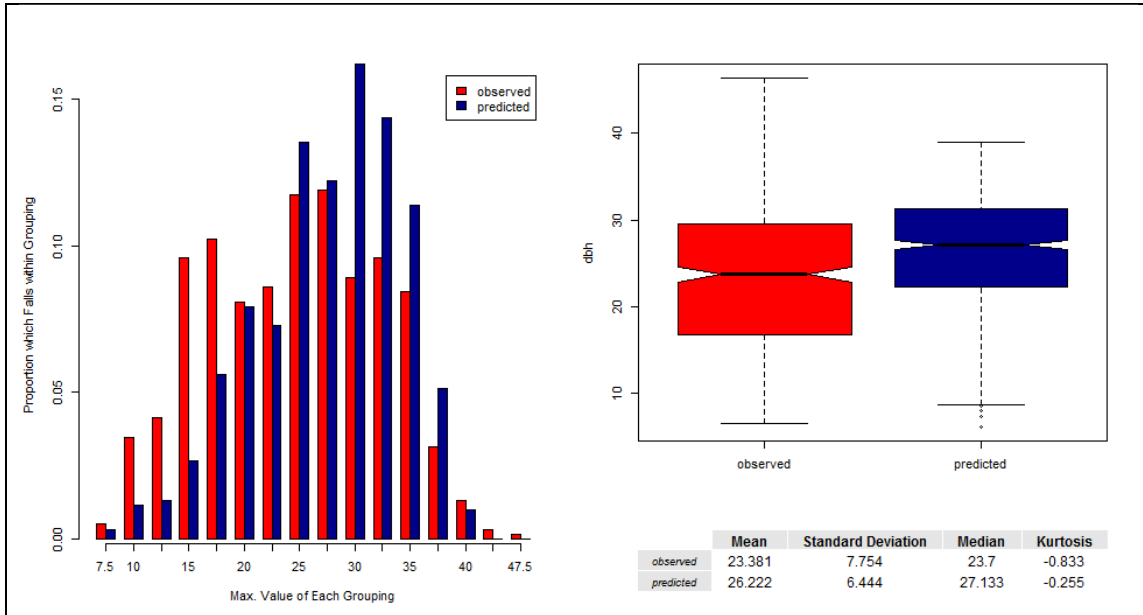




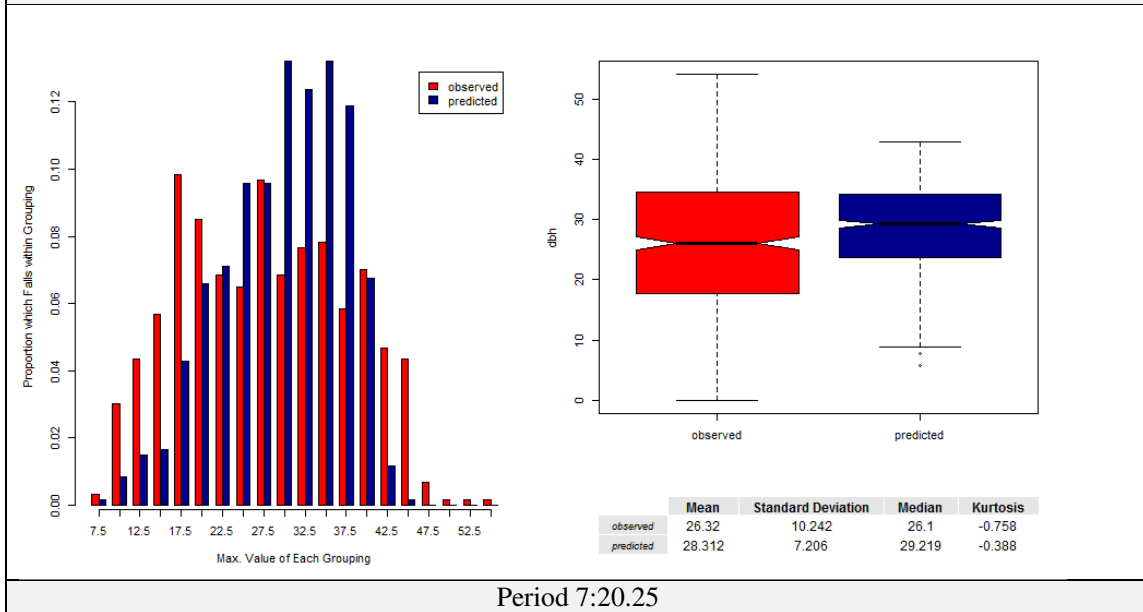
Period 4: 13.33 years



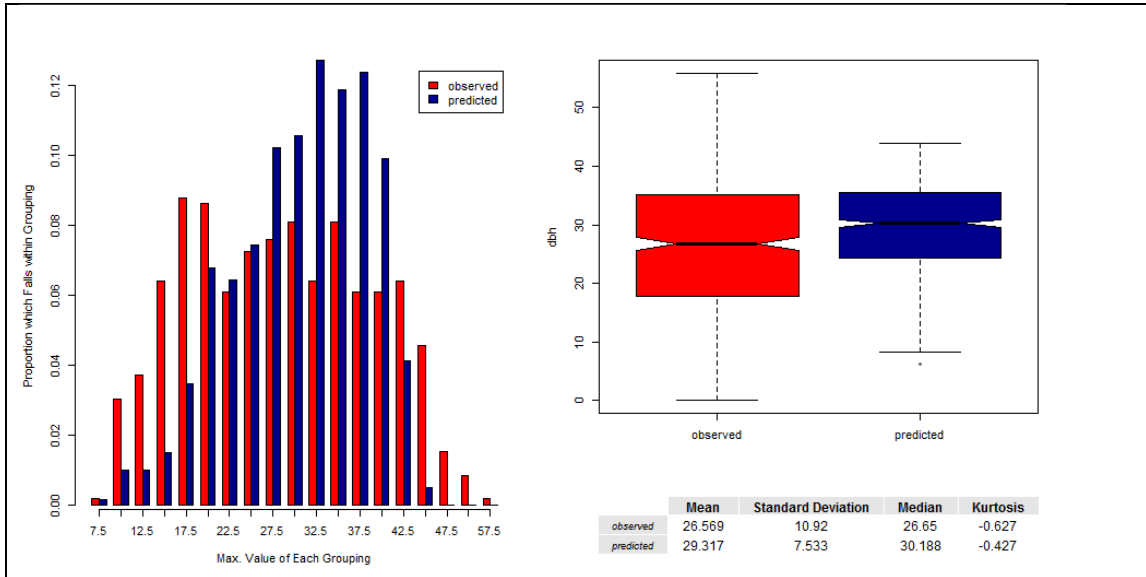
Period 5: 15.58



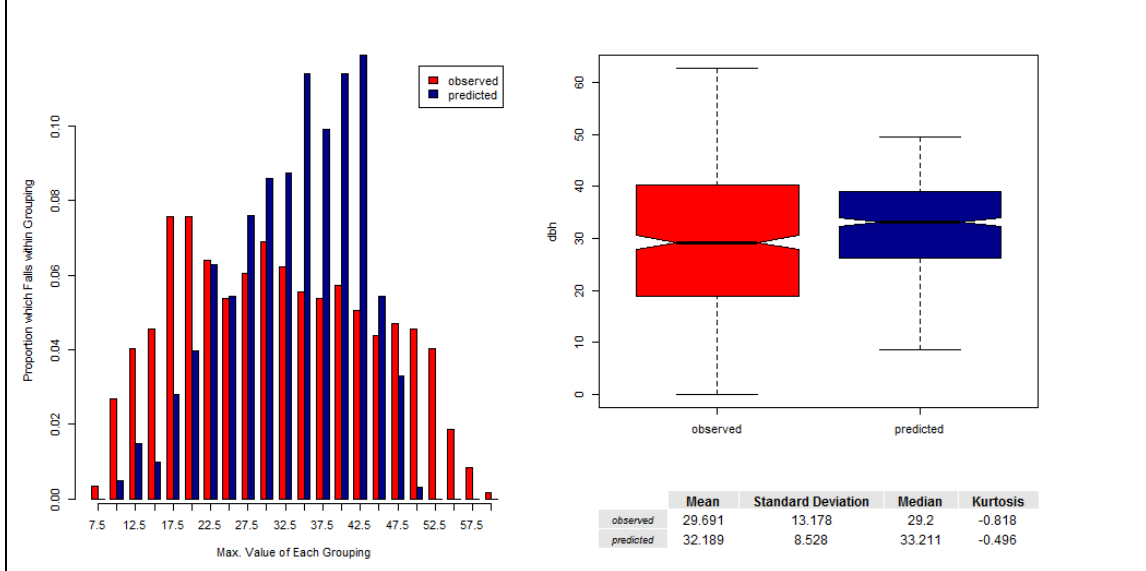
Period6: 19.17



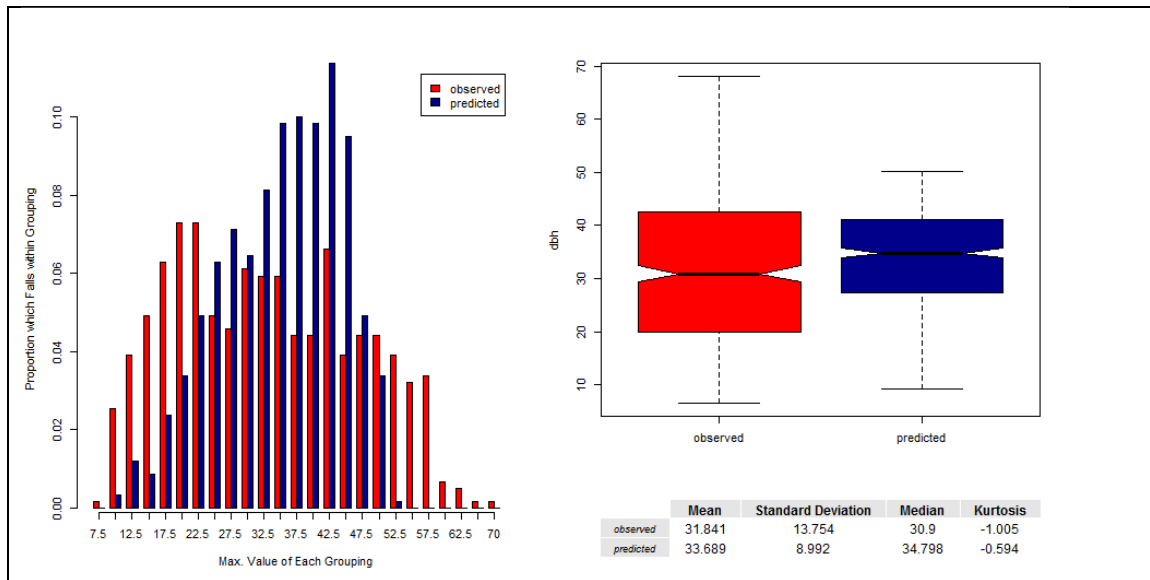
Period 7:20.25



Period 8:23.83



Period 9:26



Appendix F: Select examples of R- Code

```
#####
##### Nonlinear quantile regression example #####
#####

Allccttrials <- subset(Allccttrials, Height > 0)

str(Allccttrials)
head(Allccttrials)

###1.2 Subset the data into the different trials

Weza <- subset(Allccttrials, Location == "Weza")
Mac <- subset(Allccttrials, Location == "Mac Mac")
Kwam <- subset(Allccttrials, Location == "Kwambonambi")
Duku <- subset(Allccttrials, Location == "Dukuduku")

###2. Deriv For functions

#2.1.1 Chapman Richards Three Parameter
height.chapman <-
deriv(~ a * (1 - exp(b*x))^c,
c("a", "b", "c"),
function(x, a, b, c){},
hessian = TRUE)

Weza2965 <- subset(Weza, Spha == "2965")

a <- max(Weza2965$Height)
b1 <- max(Weza2965$Height) - min(Weza2965$Height)
b2 <- max(Weza2965$Age) - min(Weza2965$Age)
b3 <- (b1/b2)
b <- b3/(max(Weza2965$Height))
c <- 0.66
a
b
c

Weza.0.9 <- nlrq(Height ~ height.chapman(Age, a, b, c),
start = list(a = a, b = -b, c = c),
data = Weza2965, tau=0.9, trace=TRUE)

Weza.0.95 <- nlrq(Height ~ height.chapman(Age, a, b, c),
start = list(a = a, b = -b, c = c),
data = Weza2965, tau=0.95, trace=TRUE)

Weza.0.975 <- nlrq(Height ~ height.chapman(Age, a, b, c),
start = list(a = a, b = -b, c = c),
data = Weza2965, tau=0.975, trace=TRUE)
```

```
#Test0.9
handy.Height.hat <- function(Age)
predict(Weza.0.9, newdata = data.frame(Age = Age))
#Call
par(las = 1)
Plotnlrq <- plot(Height ~ Age, data = Weza2965,
xlim = c(0, max(Weza2965$Age, na.rm=TRUE)),
ylim = c(0, max(45)),
ylab = "Height(m)", xlab = "Age (y)", main = "0.9 Quantile")
curve(handy.Height.hat, add = TRUE)

summary(Weza.0.9)

#Test0.95
handy.Height.hat2 <- function(Age)
predict(Weza.0.95, newdata = data.frame(Age = Age))
#Call
par(las = 1)
Plotnlrq <- plot(Height ~ Age, data = Weza2965,
xlim = c(0, max(Weza2965$Age, na.rm=TRUE)),
ylim = c(0, max(45)),
ylab = "Height(m)", xlab = "Age (y)", main = "0.95 Quantile")
curve(handy.Height.hat2, add = TRUE)

summary(Weza.0.95)

#Test0.975
handy.Height.hat3 <- function(Age)
predict(Weza.0.975, newdata = data.frame(Age = Age))
#Call
par(las = 1)
Plotnlrq <- plot(Height ~ Age, data = Weza2965,
xlim = c(0, max(Weza2965$Age, na.rm=TRUE)),
ylim = c(0, max(45)),
ylab = "Height(m)", xlab = "Age (y)", main = "0.975 Quantile")

+ curve(handy.Height.hat, add = TRUE, lty = 2, lwd = 2, col =
"Green")
+ curve(handy.Height.hat2, add = TRUE, lty = 2, lwd = 2, col =
"Blue")
+ curve(handy.Height.hat3, add = TRUE, lty = 2, lwd = 2, col =
"Red")

legend(1, 45, c("0.90", "0.95", "0.975"), lty=c(2,2,2),
lwd=c(2,2,2), col=c("Green", "Blue", "Red"))

summary(Weza.0.975)

#####
##### Site index modelling example #####
#####
```

```
Wezadom1 <- subset(dominantheightclass, Location == "Weza")

Wezadom247 <- subset(Wezadom1, Spha == "247")
Wezadom247 <- subset(Wezadom247, Height > 0)

a <- max(Wezadom247$Height)
b1 <- max(Wezadom247$Height) - min(Wezadom247$Height)
b2 <- max(Wezadom247$Age) - min(Wezadom247$Age)
b3 <- (b1/b2)
b <- b3/(max(Wezadom247$Height))
c <- 0.66
a
c
b
nlc <- nls.control(maxiter = 200, tol = 1e-05, minFactor = 1/1024,
  printEval = FALSE, warnOnly = FALSE)
handy.nls2 <-
nls(Height ~ Height.growth(Age, a, b, c),
  start = list(a = a, b = -b, c = c),
  data = Wezadom247, control = nlc)
#Test
handy.Height.hat <- function(Age)
predict(handy.nls2, newdata = data.frame(Age = Age))
#Call
par(las = 1)
Plotted2 <- plot(Height ~ Age, data = Wezadom247,
  xlim = c(0, max(Wezadom247$Age, na.rm=TRUE)),
  ylim = c(0, max(45)),
  ylab = "Height (m)", xlab = "Age (y)", main = "Average Height")
curve(handy.Height.hat, col = "red", add = TRUE)

summ <- summary(handy.nls2)
require(nlstools)
resid <- nlsResiduals(handy.nls2)
plotresid <- plot(resid, type = 0)

library(nlme)
Weza247group <- groupedData(Height~Age|Number, Wezadom247)
plot(Weza247group)
head(Weza247group)

cor(coef(handy.nls2))

####Nlme

a <- max(Weza247group$Height)
b1 <- max((Weza247group$Height) - min(Weza247group$Height))
b2 <- max(Weza247group$Age) - min(Weza247group$Age)
b3 <- (b1/b2)
b <- b3/(max(Weza247group$Height))
c <- 0.66
```

```

a
b
c

nlc1 <- nlmeControl(maxiter = 100000, tol = 1e-05, minFactor =
1/1024,
                    printEval = FALSE, warnOnly = FALSE)

nlc2 <- nls.control(maxiter = 500, tol = 1e-05, minFactor = 1/1024,
                    printEval = FALSE, warnOnly = FALSE)

Weza247nlslist <-
nlsList(Height ~ Height.growth(Age, a, b, c), data = Weza247group,
start = list(a = 36.74088, b = -0.03314, c = 1.05124), control =
nlc2)
summary(Weza247nlslist)

pairs(Weza247nlslist, id = 0.1, na.action = na.exclude)
plot(intervals(Weza247nlslist), layout = c(3,1))

start <- c(a = a, b = -b, c = c) # starting value
nlme.fit <- nlme(Height ~ Height.growth(Age, a, b, c),
                fixed = a + b + c ~ 1, random = a + b + c ~ 1,
                data = Weza247group, start=c(start),
control=nlmeControl(opt='nlm'),
                weights = varPower(form = ~ Age), corr =
corAR1(0.296739967))
summary(nlme.fit)
pairs(nlme.fit)

anova(nlme.fit, handy.nls2)
###3 - Substantial correlation between b and c random effects -
makes it difficult to converge
intervals(nlme.fit, which = "var-cov")

nlme.fit.a <- update(nlme.fit, random = a ~ 1)
nlme.fit.b <- update(nlme.fit, random = b ~ 1)
nlme.fit.c <- update(nlme.fit, random = c ~ 1)
nlme.fit.bc <- update(nlme.fit, random = b + c ~ 1)

anova(nlme.fit.a, nlme.fit.b, nlme.fit.c, nlme.fit.bc)
anova(nlme.fit.a, nlme.fit.bc)
summary(nlme.fit.a)
summ

##Fit diagonal variance covariance matrix assuming random effects
are independent

nlme.fit.diag <- update(nlme.fit, random = pdDiag(a + b + c ~ 1),
control=nlmeControl(opt='nlm'))

#doesnt converge, thus try other random effects
nlme.fit.diagbc <- update(nlme.fit, random = pdDiag(b + c ~ 1))
nlme.fit.diagac <- update(nlme.fit, random = pdDiag(a + c ~ 1),
control=nlmeControl(opt='nlm'))

```

```
nlme.fit.diagab <- update(nlme.fit, random = pdDiag(a + b ~ 1),
control=nlmeControl(opt='nlm'))
nlme.fit.diaga <- update(nlme.fit, random = pdDiag(a ~ 1),
control=nlmeControl(opt='nlm'))
nlme.fit.diagb <- update(nlme.fit, random = pdDiag(b ~ 1),
control=nlmeControl(opt='nlm'))
nlme.fit.diagc <- update(nlme.fit, random = pdDiag(c ~ 1),
control=nlmeControl(opt='nlm'))

anova(nlme.fit.diagac, nlme.fit.diagab, nlme.fit.diaga,
nlme.fit.diagb, nlme.fit.diagc)

####With a, b and c all substantially corellated, it may not be
necessary to include them all in the model
# Just using a seems to be the best model
summary(nlme.fit.diaga)
summary(handy.nls2)

anova(nlme.fit.diaga, handy.nls2)
plot(handy.nls2)
plot(nlme.fit.diaga)
plot(nlme.fit.diagb)
plot(nlme.fit.diagbc)

#Now use the weights
nlme.fit.diaga <- update(nlme.fit.diaga, weights = varPower(form = ~
Age) )
summary(nlme.fit.diaga)

plot(nlme.fit.diaga)

anova(nlme.fit.diaga, nlme.fit)

###gnls
library(nlme)
Weza247gnls <- gnls(Height ~ Height.growth(Age, a, b, c), data =
Weza247group,
start = list(a = 46.86677, b = -0.04293, c = 1.45860))
summary(Weza247gnls)
summ

Nelder126gnls.Power3 <- update(Weza247gnls, weights = varPower(form
= ~ Age), corr = corAR1(0.296739967))
summary(Nelder126gnls.Power3)

plot(Nelder126gnls.Power3)

gnlspowerwightedresiduals <- Nelder126gnls.Power3
nlmepowerweightedresiduals <- nlme.fit.diaga
nlme <- nlme.fit
nls <- handy.nls2

anova(Nelder126gnls.Power3, nlme.fit.a, handy.nls2)
```

```
####rmse
RMSE(Nelder126gnls.Power3)
RMSE(nlme.fit.a)
RMSE(handy.nls2)

anova(gnlspowerweightedresiduals, nlmepowerweightedresiduals, nlme,
nls)

plot(nls)
plot(nlme)
plot(gnlspowerweightedresiduals)
plot(nlmepowerweightedresiduals)

##Test the autocorrelation regression improvements for the best nlme
(weighted and unweighted)

#Unweighted:
plot(ACF(nlme.fit, maxLag = 10), alpha = 0.05)
ACF(nlme.fit)
nlme.fit.acf <- update(nlme.fit, corr = corAR1(0.11717144))
plot(ACF(nlme.fit.acf, maxLag = 10), alpha = 0.05)
summary(nlme.fit.acf)
summary(nlme.fit)
anova(nlme.fit, nlme.fit.acf)
plot(nlme.fit.acf)

# Weighted
plot(ACF(nlme.fit.diaga, maxLag = 10), alpha = 0.05)
ACF(nlme.fit.diaga)
nlme.fit.diaga.acf <- update(nlme.fit.diaga, corr =
corAR1(0.29005147))
plot(ACF(nlme.fit.diaga.acf, maxLag = 10), alpha = 0.05)
ACF(nlme.fit.diaga.acf)
summary(nlme.fit.diaga)
summary(nlme.fit.diaga.acf)
anova(nlme.fit.diaga.acf, nlme.fit.diaga)
plot(nlme.fit.diaga.acf)
plot(nlme.fit.diaga)

anova(nlme.fit, nlme.fit.acf, nlme.fit.diaga, nlme.fit.diaga.acf)

##See if the Nlmeweighted and unweighted acf's compare
summary(nlme.fit.acf)
summary(nlme.fit.diaga.acf)

anova(nlme.fit.acf, nlme.fit.diaga.acf)

##See if changing the parameters used for random effects has and
effect

nlme.fit.diagbc <- update(nlme.fit, random = pdDiag(b + c ~ 1))

plot(nlme.fit.diagbc)
```

```

nlme.fit.diag.bc <- update(nlme.fit.diagbc , weights = varPower(form
= ~ Age) )
ACF(nlme.fit.diag.bc)
nlme.fit.diaga.bc.acf <- update(nlme.fit.diag.bc, corr =
corAR1(0.29005147))
plot(nlme.fit.diaga.bc.acf)
#unweighted
nlme.fit.diaga.unweight <- update(nlme.fit.diagbc, corr =
corAR1(0.29005147))
plot(nlme.fit.diaga.unweight)

summary(nlme.fit.diaga.bc.acf)
summary(nlme.fit.diaga.unweight)

anova(nlme.fit.diaga.bc.acf, nlme.fit.diaga.unweight)

anova(gnlspowerweightedresiduals, nlmpowerweightedresiduals, nlme,
nls, nlme.fit.diaga.acf)

#####
##### Competition indices exapmle #####
#####

####BALIndex

#df <- read.csv("C:/Users/Gerard/Documents/Masters/r voronoi/Nelder
Year by year/1998.csv")
df <- read.csv("F:/Tweefontein BAL/Tweefonteinnew.csv")

#df<- subset(df, TPH0 == "245")
#df<- subset(df, TPH0 == "403")
#df<- subset(df, TPH0 == "665")
df<- subset(df, TPH0 == "1097")
#df<- subset(df, TPH0 == "1808")
#df<- subset(df, TPH0 == "2981")

head(df)
# computes the BAL competition index as follows:
# all neighbour trees inside the competition zone radius (CZR)
# contribute to the total sum of the Basal area, if their
# DBH is thicker than the central tree.
#
# alle trees are assumed to be on a axias allinged rectangular plot
#
# stefan seifert 2013-1

require(RANN)

df<- subset(df, Age == "1.83")

```

```

df<- subset(df, Age == "2.5")
df<- subset(df, Age == "3.58")
df<- subset(df, Age == "6.25")
df<- subset(df, Age == "7.83")
df<- subset(df, Age == "10.5")
df<- subset(df, Age == "12.67")
df<- subset(df, Age == "14.58")
df<- subset(df, Age == "16.58")
df<- subset(df, Age == "18.33")
df<- subset(df, Age == "20.42")

d <- df
head(d)
createtorus = function(d, dx=NA, dy=NA) {
  if (is.na(dx))
    dx=diff(range(d$x))
  if (is.na(dy))
    dy=diff(range(d$y))
  dnew = d
  dnew$outside=F
  # centerrow
  dc=rbind(dnew,transform(d, x=x-dx, outside = T))
  dc=rbind(dc,transform(d, x=x+dx, outside = T))
  #+top
  dnew = rbind(dc,transform(dc, y=y-dy, outside=T))
  #+bottom row
  dnew = rbind(dnew, transform(dc, y=y+dy, outside=T))
  return(dnew)
}

head(createtorus(d))

neighbours.BAL = function(x,y,r=plotsize,k=100) {
  # returns all neighbours in a distance of r
  # neighbour number is limited to k
  # warning will be given if there might be more neighbours than k
  require(RANN)
  k = min(k,nrow(x))
  neigh = nn2(x[,c("x","y")], x[,c("x","y")], searchtype="radius",
radius = r-1e-6, k=k)
  if(max(neigh$nn.idx[,k])>0 & k<nrow(x)) {
    warning("neighbours.BAL: k might be too less to find all
potential neighbours")
  }
  return(neigh)
}

p = createtorus(d)
head(p)
str(p)

BAL = function(p,CZR=plotsize) {
  # compute the BAL index, BASAL AREA, and DIAM for all trees in p

```



```

# p is a data.frame with the columns x, y for the tree position
#   and D for the DBH of the tree
#   a column newID with unique IDs
# CZR is teh radius to which neighbours are counted
#   if CZR is NA then CZR will be estimated as 2 times the
#   squareroot of the specific tree area
# bounding
BAL = rep(NA,nrow(p))
BASAL = BAL
CR13 = BAL
mimax = range(p$x)
mimay = range(p$y)
if (is.na(CZR)) {
  a = diff(mimax)*diff(mimay)
  spa = a/nrow(p) # 1/intensity : specific area per tree
  CZR = sqrt(spa) * 2 # k-facor 2 : radius is 2-times the
specific tree area
}
border = with(p, x<=mimax[2]-CZR & x>=mimax[1]+CZR &
y<=mimay[2]-CZR & y>=mimay[1]+CZR)
sqha = CZR^2*pi/10000 # circle ha
ne = neighbours.BAL(p,CZR, k=100) # all neighbours for all trees!
# sum up all valid neighbours squared DBHs
# we do it in a loop
for (ct in 1:nrow(p)) {
  nonu = ne$nn.idx[ct,-1]
  nonu = nonu[nonu>0]
  d = p[nonu,"D"]
  d1 = d[d>p$D[ct]]
  d2 = d[d>10 & d <=30]

  BAL[ct] = sum( d1^2 )*pi/40000 / sqha # [ basal area per ha]
  BASAL[ct] = sum( d^2 )*pi/40000 / sqha
  CR13[ct] = sum(d2) / sqha
}
return(data.frame(newID=p$newID, BAL=BAL,BASAL=BASAL, Age =
p$Age, DBH = p$D, DIncrement = p$Di,HtIncrement = p$Hti,
Ht = p$Ht, x = p$x, y = p$y, SPHA = p$TPH0,
flagBALOk=border, CR13=CR13, CZR=rep(CZR,nrow(p)) ))
}

```

```

BALCALL <- BAL(p)
head(BALCALL)
plot(BALCALL$BAL, BALCALL$DIncrement)
plot(BALCALL$BASAL, BALCALL$DIncrement)

```

```
#####Hegyí Index
```

```

p <- BALCALL
head(p)
names(p)[names(p) == "newID"] <- "ID"

```

```

p$newID <- 1:nrow(p)
p$D <- p$DBH

CZR <- plotsize

neighbours.heg = function(x,r,k=100) {
  # returns all neighbours in a distance of r
  # neighbour number is limited to k
  # warning will be given if there might be more neighbours than k
  require(RANN)
  k = min(k,nrow(x))
  neigh = nn2(x[,c("x","y")], x[,c("x","y")], searchtype="radius",
  radius = r-1e-6, k=k)
  if(max(neigh$nn.idx[,k])>0 & k<nrow(x)) {
    warning("neighbours.heg: k might be too less to find all
potential neighbours")
  }
  return(neigh)
}

it.heg.stand = function(p,CZR, k=100) {
  # p are all trees in the plot
  # this data.frame must contain x,y as coordinates
  # and d as diameter [cm]
  # the column newID should contain a unique ID like the row
number
  # if some trees are excluded this ID helps to reconstruct the
  # original data
  require(plyr)

  ne = neighbours.heg(p,CZR, k) # all neighbours for all trees!
  HgCI = rep(NA,nrow(p))
  newID = HgCI
  flagBorder = rep(NA,nrow(p))

  mimax = range(p$x)
  mimay = range(p$y)
  flagBorder = with(p, x<=mimax[2]-CZR & x>=mimax[1]+CZR &
y<=mimay[2]-CZR & y>=mimay[1]+CZR)
  for (ct in 1:nrow(ne$nn.idx)) {
    #print(ct)
    # test if current tree ct is too near the maximum extent of
the plot (B:think this comment is meant for line 53)
    z = c()
    # number of neighbours
    nonna = (ne$nn.idx[ct,]>0)
    nc = length(na.omit(ne$nn.idx[ct,nonna])) # B:na.omit just
removes incomplete cases, i.e. more than number k, or out of CZR

    neighb = ne$nn.idx[ct,nonna][-1] # all valid neighbour
indexes
    neighbd = ne$nn.dist[ct,nonna][-1] # all valid neighbours
distances
    HgCI[ct]=sum( (p$D[neighb]/p$D[ct])/neighbd )
    newID[ct]=p$newID[ct]
  }
}

```

```

}
  return(data.frame(HgCI=HgCI,flagInside=flagBorder,newID=newID))
}

source("J:/B Fury/ci-Hegyi-iterNocea.R")#specifiy which script has
the functions being used

head(p)
str(p)

heg=it.heg.stand(p,CZR,35)
plot(p$x,p$y)
head(heg)
str(heg)
heg.1=data.frame(ID          =          p$ID,          newID=p$newID,
BAL=p$BAL,BASAL=p$BASAL,HgCI=heg$HgCI, Age = p$Age,
                DBH      =  p$DBH,   DbhIncrement =  p$DIncrement,
HtIncrement = p$HtIncrement,
                Ht      =  p$Ht,   x = p$x,   y = p$y,   SPHA = p$SPHA,
CZR=rep(CZR,nrow(p)))

plot(heg.1$HgCI, heg.1$DbhIncrement)
head(heg.1)

str(heg.1)

#####Voronoi
p <- heg.1
head(p)
str(p)
treecoords <- subset(p, select=c(newID,x,y))
require(deldir)

x<-c(treecoords$x)
y<-c(treecoords$y)
plot(x,y)
delresult <- deldir(x,y)

summ<-delresult[["summary"]]
summ
summ8<-summ[8]
summ8merge <- merge(summ8, heg.1, all=TRUE, by="row.names")

plot(delresult)

plot(delresult,add=FALSE,wlines=c("tess"),
wpoints=c("both","real","dummy","none"),
number=FALSE,cex=1,nex=1,col=NULL,lty=NULL,
pch=NULL,xlim=NULL,ylim=NULL,xlab='x',ylab='y',
showrect=FALSE)

plot(delresult,add=FALSE,wlines=c("tess"),

```

```
wpoints=c("none"),
number=FALSE, cex=1, nex=1, col=NULL, lty=1,
pch=NULL, xlim=NULL, ylim=NULL, xlab='x', ylab='y',
showrect=FALSE)

head(summ8merge)
summ8merge$conc <- paste(summ8merge$ID, summ8merge$newID, sep = '.')
str(summ8merge)
plot(summ8merge$dir.area, summ8merge$DbhIncrement)

setwd("F:/B Fury/Tweefontein/hegyi and cvor")
write.csv(summ8merge, "2981.20.42.csv")
```