

INDIVIDUAL-TREE DIAMETER GROWTH MODELS AND VARIABILITY OF MIXED
NOTHOFAGUS SECOND GROWTH FORESTS IN SOUTHERN CHILE

By

PAULO CESAR MORENO MEYNARD

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2017

© 2017 Paulo Cesar Moreno Meynard

To Elena my daughter
Carmen and Neftalí, my parents

ACKNOWLEDGMENTS

I would like to thank the members of my supervisory committee, Drs. Gezan, Escobedo and Cropper, for sharing with me your time and knowledge through my program. Specially, I am very grateful with Dr. Gezan for giving me the opportunity to realize one of my dreams, to study a Master in University of Florida. In addition, I thank him for his time, patience, guidance, energy, and dedication to help me achieve the goals of this research.

I also want to thank the help of Dr. Ortega from Universidad Austral de Chile, with the database on which this studio is based. Thanks to Sebastian for sharing your knowledge in programming and to be an excellent coworker.

Special thanks to my family, specially to my daughter Elena, who was without her father along this journey; my parents Carmen y Nefalí and my sisters for helping me to realize this dream. Also, I would to thank for their support to Andrés and the Chilean community in Gainesville and so many friends, who became my family during these last two years.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
LIST OF TABLES.....	7
LIST OF FIGURES.....	8
LIST OF ABBREVIATIONS.....	10
ABSTRACT.....	13
CHAPTER	
1 INTRODUCTION.....	15
2 UNDERSTANDING VARIABILITY IN MIXED <i>NOTHOFAGUS</i> SECOND GROWTH FORESTS IN SOUTHERN CHILE: A MULTIVARIATE ANALYSIS.....	18
Background.....	18
Materials and Methods.....	22
Data Description.....	22
Multivariate Methods.....	24
Results.....	26
Discussion.....	30
3 INDIVIDUAL-TREE DIAMETER GROWTH MODELS OF MIXED <i>NOTHOFAGUS</i> SECOND GROWTH FORESTS IN SOUTHERN CHILE.....	41
Background.....	41
Materials and Methods.....	46
Data Description.....	46
Model Fitting.....	48
Results.....	52
Model Fitting.....	52
Model Projection.....	56
Selected Model.....	56
Discussion.....	57
4 CONCLUSIONS.....	72
APPENDIX	
A DOMINANT HEIGHT-SITE MODEL.....	75
B SUPPLEMENTARY TABLES.....	76

C SUPPLEMENTARY FIGURES	77
LIST OF REFERENCES	82
BIOGRAPHICAL SKETCH	88

LIST OF TABLES

<u>Table</u>	<u>page</u>
2-1 Summary statistics of this study dataset for individual and stand variables in each growth zone.	33
2-2 Factor loading for the first five principal component (PC) obtained from the tree and stand subsets together. Values in parenthesis correspond to the proportion of the variance explained by each component.	34
2-3 Factor loading for the first five principal component (PC) obtained from average plot AIDBH and stand subset. Values in parenthesis correspond to the proportion of the variance explained by each component.	35
3-1 Summary statistics of individual and stand attributes for fitting database based on PP and TP networks.	64
3-2 Parameter estimates of the four selected models for estimation of the logarithm of AIDBH (mm).....	65
3-3 Goodness of fit statistics of four models for ln(AIDBH) using test data.....	65
3-4 Projection goodness-of-fit statistics for DBH (cm) in a period of 6 and 12 years using projection database.....	66
3-5 Least square means (standard errors in parentheses) statistics of all pairwise Tukey comparisons between all 11 groups based for Specie-Zone in CV + SpZone model.	67
A-1 Parameter estimates of the dominant height-site model by specie and growth zone.....	75
B-1 P-values with Bonferroni correction obtained from a post-hoc test of all six possible pairwise comparisons for growth zones related to two clusters.....	76
B-2 Mean diameter breast height growth rates (standard deviation in parentheses) by specie and growth zone obtained from the available data (fitting database).....	76
B-3 Goodness-of-fit statistics for factors specie, zone and both in the CV regression model using the test data.....	76

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-1 Non-Metric Multidimensional Analysis (NMDS) using the tree subset, identifying growth zones (a) and species (b).	36
2-2 Non-Metric Multidimensional Analysis (NMDS) using the stand subset, identifying growth zones (a) and dominant species (b).....	37
2-3 Principal coordinates analysis (PCoA) using the tree and stand subsets together, identifying growth zones (a) and species (b).	38
2-4 First two PCA axes with scores and factor loadings obtained using, the tree and stand subsets (a), stand subset together with average plot AIDBH. The two K-means clusters are also identified (b).....	39
2-5 Geographical localization of plots identified according to their two K-means clusters in relation to growth zones.	40
3-1 Plots distribution in the study area (southern Chile) for the permanent plots (PP) and temporary plots (TP) networks.	68
3-2 Statistics C_p , BIC, RSS, R^2_{adj} and MSE_K for different number of predictors with CV regression selection model using fitting database.....	69
3-3 Fit and residuals plots of four selected models using test dataset. CV (a and b), LASSO (c and d), CV + SpZone (e and f), and LASSO + SpZone (g and h).	70
3-4 RMSE% (a) and BIAS% (b) by DBH class for the four selected models evaluated using the test dataset.....	71
3-5 Simulation prediction in future DBH (cm) for projection periods of 6 years (a) and 12 years (b) based on the CV + SpZone model using projection database.....	71
C-1 Scatterplot correlation matrix of predictors (in original units) for the fitting database.....	77
C-2 Scree plot of PCoA with the cumulative variance of each coordinate based on the continuous variables of the tree and stand subsets.....	78
C-3 Scree plot of PCA for continuous variables of tree and stand subset.	78
C-4 PC1 and PC3 with scores and factor loadings, using tree and stand subset.....	79
C-5 PC2 and PC3 with scores and factor loadings, using tree and stand subset.....	79

C-6	Scree plot of PCA for continuous variables of stand subset and average plot of AIDBH.	80
C-7	Scree plot of optimal number of cluster using within sum of squares and average silhouette width for stand-level information.....	80
C-8	Histograms of expected deltas and observed delta obtaining by MRPP test for two K-means cluster using stand-level information.	81

LIST OF ABBREVIATIONS

A	Breast Height Age
As	Chance Corrected Within-group Agreement Statistic
Ad	Dominant Stand Age at Breast Height
AIDBH	Annual Increment in Diameter at Breast Height
BA	Stand Basal Area
BAL	Basal Area of Larger Trees
BALn	Basal Area of Larger Trees of <i>Nothofagus</i>
BALr	Relative Basal Area of Larger trees
BAN	Basal Area of <i>Nothofagus</i>
BIAS%	Relative Bias
BIC	Schwartz's information criterion
CA	Correspondence Analysis
CONAF	National Forest Corporation of Chile
Cp	Mallows' Cp
CV	Cross-Validation
DBH	Diameter at Breast Height
DS	Dominant Species
GVIF	Generalized Variance Inflation Factor
G&Y	Growth and Yield
H	Total Height
Hd	Dominant Height
INFOR	Forest Research Institute of Chile
LASSO	Least Absolute Shrinkage and Selection Operator
In	Natural logarithm

MRPP	Multiple Response Permutation Procedure
MSE	Mean Square Error
MSE _K	Mean Squared Error averaged over the K folds
N	Trees per hectare
NMDS	Non-metric Multidimensional Scaling
PC	Principal Component
PCA	Principal Component Analysis
PCoA	Principal Coordinates Analysis
PP	Permanent Plots
QD	Quadratic mean Diameter
R ² _{adj}	Adjusted r-squared
R ² _{emp}	Empirical coefficient of correlation
RR	Ridge Regression
RS	Relative Spacing Index
RSS	Residual Sum of Squares
RMSE	Root Mean Square Error
RMSE%	Relative Root Mean Square Error
RORACO	Chilean forest type composed by three <i>Nothofagus</i> species: roble (<i>Nothofagus obliqua</i>), raulí (<i>N. alpina</i>), and coihue (<i>N. dombeyi</i>).
SD	Standard Deviation
SDI	Stand Density Index
SI	Site Index
SP	Specie
SS	Sociological Status
TP	Temporary Plots

VIF

Variance Inflation Factor

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

INDIVIDUAL-TREE DIAMETER GROWTH MODELS AND VARIABILITY OF MIXED
NOTHOFAGUS SECOND GROWTH FORESTS IN SOUTHERN CHILE

By

Paulo Cesar Moreno Meynard

August 2017

Chair: Salvador A. Gezan

Major: Forest Resources and Conservation

The second growth forests of *Nothofagus obliqua* (roble), *N. alpina* (raulí), and *N. dombeyi* (coihue), defined locally as RORACO forest type, is one of the most important natural mixed forests of Chile. Several studies have identified a wide range of factors that influence the phenotypic variability found in this forests. Currently there is no individual-tree diameter growth model for the full geographical distribution of the RORACO forest type. This study, contributed to better understanding the underlying variability for these forests through the use of the following multivariate analysis: 1) non-metric multidimensional scaling (NMDS); 2) principal coordinates analysis (PCoA); and 3) principal component analysis (PCA). In addition, an individual-tree diameter growth model was performed, comparing two selection variable procedures: 1) cross-validation (CV); and 2) Least Absolute Shrinkage and Selection Operator (LASSO) regression. The data in this study is based on tree and stand variables typically found in available forest inventories. Findings indicate that site productivity and growth zones do not explain much of the variability of this sampled population. However, stand development stages, tree-to-tree competition, and tree-size attributes are critical variables with a high percentage of variance explained, ranging from 61% to 67%. In addition, the final

diameter growth model used a CV procedure incorporating species and growth zones as a combined factor. A small set of predictors successfully explained a large portion of diameter breast height (DBH) growth, with variables associated to competition, tree- and stand-level. The goodness-of-fit statistics for this final model were $R^2_{\text{emp}} = 0.56$, $\text{RMSE}\% = 44.49\%$ and $\text{BIAS}\% = -1.96\%$ for annual DBH growth prediction, and $R^2_{\text{emp}} = 0.98$ and 0.97 for DBH projection after 6 and 12 years, respectively. This model constitutes a simple and useful tool to support management and decision-making for these ecosystems.

CHAPTER 1 INTRODUCTION

One of the most important forest types in Chile is known locally as RORACO, its name originates from their three main *Nothofagus* species: roble (*Nothofagus obliqua* (Mirb.) Oerst.), raulí (*N. alpina* (Poepp. Endl.) Oerst.), and coihue (*N. dombeyi* (Mirb.) Oerst.). These forests are found in the south of Chile from 36°S to 41°S latitude approximately from the coast to the Andes range, corresponding to 10.8% of the total native forest area of the country (CONAF 2011). In 2015, the saw timber volume from these three *Nothofagus* species was approximately 45% of the national production (INFOR 2016). The volume growth rates for these forests range between 6 and 10 m³ ha⁻¹ year⁻¹ (Donoso et al. 2004), corresponding to some of the highest values for southamerican temperate native forests.

Despite the importance of this resource, forest managers and owners do not have enough information and tools, regarding present conditions and future state under varying silvicultural activities. Some information is available locally, but not for the full geographical range of the RORACO forest type. In addition, another related problem is the high phenotypic variability found in these forests due to different factors, such as large geographical area, varying (or lack of) management and strong anthropic disturbances, together with diverse biophysical variables, ecology, and species (Donoso 1993; Chauchard and Sbrancia 2003; Echeverria and Lara 2004; Luebert and Pliscoff 2006).

A very important tool, to describe and project any stand structure, is individual-tree growth and yield (G&Y) model with the possibility of incorporating management and estimates of timber products (Andreassen and Tomter 2003). These models are

constituted by three components (Burkhardt and Tome 2012): 1) individual-tree diameter at breast height (DBH) or basal area growth equation; 2) individual-tree height growth equation or, alternatively, a height-diameter relationship; and 3) mortality or survival component.

Most G&Y models, particularly individual-tree growth models, are obtained by fitting multiple linear regressions with many potential predictors in their original units or using transformations. These predictors often suffer from multicollinearity due to high correlation between them (Welham et al. 2015), but there are several statistical methods for selecting predictors to avoid multicollinearity, such as stepwise or shrinkage methods (James et al. 2013).

The data for this research originated from a study from the Universidad Austral de Chile (Ortega and Gezan 1998) consisting of two plot networks. One with a total of 128 permanent plots, and another with 60 temporary plots. The data were collected using a stratified population, based on a national forest cadastre (CONAF et al. 1999), for the full geographical distribution of the RORACO forest type (> 14,000 km², including four growth zones, proposed by Gezan and Moreno; 1999). The variables used in this study are based on tree and stand attributes typically obtained from forest inventories, such as site productivity, tree-to-tree competition, individual-tree and whole-stand attributes (Vanclay 1994).

Therefore, the aims of this study were: 1) to understand the underlying phenotypic variability for second growth forests of *N. alpina*, *N. obliqua*, and *N. dombeyi*, through unsupervised multivariate methods, and 2) to develop an individual-

tree growth model to estimate annual increment in DBH for the same species, using a set of tree and stand level predictors.

I will analyze the first aim using several multivariate analysis of ordination and clustering (Chapter 2). I will specifically analyze this objectives: 1) to reduce the dimensionality of the data through ordination techniques, such as non-metric multidimensional scaling (NMDS), principal coordinates analysis (PCoA) and principal components analysis (PCA); 2) to identify those variables or factors that explain high amount of data variability; and 3) to generate K-means clusters to assess the validity of the growth zones.

For the development of an individual-tree growth model (Chapter 3), two selection variable procedures were evaluated: 1) model selection with cross-validation, and 2) LASSO regression. The resulting selected models were later extended using growth zones or species as factors, to evaluate if they contribute to improve the performance of these models, and therefore, better explain dynamics for this forest type. The specific objectives in this chapter were: 1) to compare different multiple linear fitting procedures to obtain the best individual-tree growth model; 2) to generalize diameter growth models to incorporate growth zone and specie, and 3) to evaluate the final diameter growth model using an independent validation dataset.

CHAPTER 2
UNDERSTANDING VARIABILITY IN MIXED *NOTHOFAGUS* SECOND GROWTH
FORESTS IN SOUTHERN CHILE: A MULTIVARIATE ANALYSIS

Background

The structure and dynamics of a mixed natural forest are difficult to understand, where only certain aspects of their variability can be studied in detail (Landres et al. 1999). Forest variability is often expressed as heterogeneity, defined as any form of environmental variation, physical or biotic, occurring in space and/or time (Kuuluvainen et al. 2002). Micro-site or local factors influences individual trees, which are the main source of small-scale environmental heterogeneity in forest ecosystems. These factors include: light quantity and quality due to changes in canopy, interception of precipitation from foliage and branch litter, and root uptake of water and nutrients, among others (Kuuluvainen et al. 2002). Hence, this small-scale heterogeneity is often translated into tree-to-tree competition at many levels. This competition generates a high variability, where complex mixed and natural forests are often more heterogeneous.

One of the most relevant forests in the southern hemisphere are composed by species of the Nothofagaceae family, with *Nothofagus* as their unique genre, which are found in Australia, New Zealand, New Guinea, New Caledonia, Chile, and Argentina. The southamerican *Nothofagus* species dominate temperate and sub-Antarctic forests of Chile and Argentina. In Chile, there are nine species, including three evergreens, and six deciduous (Ormazabal and Benoit 1987; Donoso 1993).

A very important *Nothofagus* forest types in Chile is called RORACO, its name originates from their three main *Nothofagus* species: roble (*Nothofagus obliqua* (Mirb.) Oerst.), raulí (*N. alpina* (Poepp. Endl.) Oerst.), and coihue (*N. dombeyi* (Mirb.) Oerst.). RORACO is the fourth largest forest type in Chile with 10.8% of the total surface

(CONAF 2011), with saw timber production reaching 45% of the total wood market (INFOR 2016).

The second growth forest stands of RORACO are found in the south of Chile from 36°S to 41°S latitude approximately from the coast to the Andes mountain range. This wide distribution leads to a diverse set of ecosystems and growing conditions (Donoso 1993; Donoso et al. 1993). These forests are commonly mixed stands, and are characterized by the presence of two or more vertical strata where *Nothofagus* species are the dominant over story strata and companion are often shade tolerant (Donoso 1981, 1993). In addition, it is common to find some level of hybridization between *Nothofagus* species (Donoso 1987). All of the above elements contribute to a high natural variability of this forest type. Several studies have characterized their different aspects, such as: tree volume, taper equations, growth zones, diameter breast height (DBH) growth models, height-diameter relationships, mortality, etc. (Cubillos 1987; Donoso et al. 1993; Schlatter and Gerding 1995; Ortega and Gezan 1998; Moreno 2001; Chauchard and Sbrancia 2003; Lusk and Ortega 2003; Echeverria and Lara 2004; Salas and Garcia 2006; Donoso and Lusk 2007; Gezan et al. 2007, 2009; Thiers et al. 2008; Esse et al. 2013).

RORACO forest type is a complex ecosystem with many origins, where its classification presents diffuse boundaries overlapping with others forests type, and they are present in a continuum vegetation gradients spanning wide geographical ranges in latitude and altitude (Donoso 2003). These forests belong to the bio-climate temperate hiperoceanic with a large extension from coastal zones, interior valleys, and mountainous areas, containing several microclimates (Luebert and Plissock 2006).

Echeverría and Lara (2004) determined that the variability in diameter growth in *N. alpina* and *N. obliqua* were associated mainly to climate and soils. Another study found that, at a macro region scale, *N. obliqua* productivity depends mainly on climate, but at a regional scale, edaphic and topographic variables are more relevant (Thiers et al. 2008). Esse et al. (2013) found that the natural spatial distribution of *N. dombeyi* is explained, in decreasing order, by annual average rainfall, thermal oscillation and soil drainage.

Two critical factors that affect the variability of RORACO forests are the presence of pure or mixed forest, and their regeneration strategy (Donoso 1993). Presence of pure or mixed forest is explained with the interaction between autecology of each species and biophysical variables, where, for example, extreme climate and soils conditions tends to favor pure forest (Donoso 1993). In addition, high and low forests are the result of different regeneration strategies, including sexual (seeds) and asexual propagation (coppice).

Forest disturbances are an important factor in the ecology of these *Nothofagus* species, contributing to an increase in their variability. Large scale disturbances usually promote a vast regeneration event, which promotes even-aged strata or cohorts (Luebert and Plischoff 2006). However, Donoso (1993) points out differences in the ecology strategies of the *Nothofagus* species, which range from big disturbances to patch dynamics. Here, *N. dombeyi* and *N. obliqua* are more aggressive than *N. alpina*, preferring large disturbances; whereas the latter, due to its shadow tolerance nature, favors patch dynamics. RORACO forest suffer from a high level of anthropic disturbances affecting severely the landscape and their ecology (Luebert and Plischoff

2006). For example, these landscape transformations can influence the development of varieties or even geographic races (Donoso 1993).

Individual- and stand-level growth has been always an important variable used to characterize most forest. Donoso et al. (1993) defined four growth zones for *N. alpina* and *N. obliqua*, where the better growth zones corresponded to those with temperate climate. Esse et al. (2013) proposed for *N. dombeyi* five growth zones, which were defined, in decreasing order, by climate, soil, and topography. Also, Gezan and Moreno (1999), based mainly on climate, soil, site productivity, defined four growth zones and 11 sub-zones. Several studies on volume equations, basal area growth, dominant height-site, and taper equations have found this zoning useful to model fitting (Moreno 2001; Gezan et al. 2009).

Multivariate analysis is a group of methods that help to describe and/or explore several data and they provide with formal inferences about the population (Everitt and Hothorn 2011). There are two methods of multivariate analysis: supervised and unsupervised (James et al. 2013). In supervised methods the goal is predict a desired response from a set of predictor variables, whereas for unsupervised methods the goal is discover relationships among the data from the complete set of variables. For example, with the unsupervised methods of ordination or dimensionality reduction it is possible to understand data variability, or by using cluster analysis unknown similarities between objects can be discovered (James et al. 2013). Some multivariate methods of ordination are: principal component analysis (PCA), principal coordinates analysis (PCoA, also called multidimensional scaling), non-metric multidimensional scaling (NMDS), and correspondence analysis (CA) (Everitt and Hothorn 2011). These

methods help reduce the complexity of the data, visualize multidimensional data and develop research hypotheses (Everitt and Hothorn 2011). Cluster analysis allow finding similar subgroups of objects in a dataset, where some of the approaches commonly used are: agglomerative hierarchical, K-means clustering, and model-based clustering (Everitt and Hothorn 2011).

The main objective of this study is to understand the underlying variability of second growth forests of *N. alpina*, *N. obliqua*, and *N. dombeyi*, through multivariate analysis of ordination (non-metric multidimensional scaling, principal coordinates analysis and principal components analysis) and clustering (K-means) using a group of tree and stand variables typically obtained from forest inventories. In particular 158 plots (1,108 trees with growth data) covering the complete geographical distribution of the RORACO forest type. The specific objectives are: 1) to reduce the dimensionality of the data through ordination techniques; 2) to identify those variables or factors that explain high amount of data variability; and 3) to generate clusters to assess the validity of the growth zones proposed by Gezan and Moreno (1999).

Materials and Methods

Data Description

A total of 1,108 trees of *N. alpina*, *N. obliqua*, and *N. dombeyi* with diameter growth information were available for this analysis. These originated from a sample of 158 plots from a study of the Universidad Austral de Chile (Ortega and Gezan 1998) consisting on two plot networks. The data were collected under a population stratification based on a national forest cadastre (CONAF et al. 1999). Further details of this dataset can be found in Gezan and Moreno (1999).

This study used the average annual increment in DBH based on the penultimate and antepenultimate growth years, which was obtained from increment cores (864) and stem sections (244) at breast height. The last growth year was discarded as plots were established at different times within the growing season. The dataset includes several tree- and stand-level variables (Table 2-1). The tree-level attributes were: species, diameter breast height (DBH, cm), total height (H, m), breast height age (A, year), basal area of larger trees (BAL, m² ha⁻¹), basal area of *Nothofagus* of larger trees (BALn, m² ha⁻¹), sociological status (SS, defined according to vertical stratification with dominant (1), codominant (2), intermediate (3) or suppressed (4)), relative basal area of larger trees (BALr, with BALr = BAL/BA) and the average annual increment in DBH (AIDBH, mm year⁻¹). The stand-level variables considered were: growth zone (Zone), stand basal area (BA, m² ha⁻¹), density or number of trees per hectare (N, trees ha⁻¹), quadratic mean diameter (QD, cm), dominant height (Hd, m), dominant breast height stand age (Ad, years), dominant specie (DS), site index (SI, m), basal area of *Nothofagus* (BAN, m² ha⁻¹), stand density index (SDI, trees ha⁻¹) and relative spacing index (RS). BAN was calculated as the sum of basal area of *N. alpina*, *N. obliqua* and *N. dombeyi*. Hd and Ad were calculated using the 100 largest diameter trees per hectare, and DS corresponded to the *Nothofagus* specie with the largest proportion of BAN. SI was estimated using available dominant height-site models obtained on the same dataset for an index age of 20 years (see appendix A). SDI was calculated using: $SDI = N \times (25.4/QD)^\beta$ (Avery and Burkhart 2002), where $\beta = -1.4112$, as reported by Gezan et al. (2007) fitted using both the permanent plot (PP) and temporary plot (TP) networks. Finally, RS was calculated according to: $RS = [(10000/N)^{0.5}]/Hd$ (Avery and Burkhart 2002; Contreras et al. 2011).

Multivariate Methods

To understand the underlying variability of the RORACO forest, three methods of ordination were evaluated with the available data: 1) non-metric multidimensional scaling (NMDS), 2) principal coordinates analysis (PCoA), and 3) principal component analysis (PCA). The goal of ordination methods is to represent the main trends of the data through orthogonal axes (linearly independent, uncorrelated), reducing the original number of dimensions. The observed trends can be interpreted graphically or be further evaluated with the use of other multivariate methods such as clustering (Borcard et al. 2011).

NMDS is probably the most flexible ordination method allowing the use of continuous and categorical variables (Everitt and Hothorn 2011), and unlike PCoA and PCA, it is not an eigenvalue-eigenvector-based method. NMDS, for the new data representation, keeps the order of the relationships among the initial set of objects but not the exact distances (Borcard et al. 2011). This method tolerates missing values, as long as there is enough information to identify the position of each object with respect to others (Borcard et al. 2011). For this study, the reduction of dimensionality was established using two axes ($k = 2$), with a modified Gower dissimilarity matrix (Anderson et al. 2006). Two NMDS analyses were performed one with the tree subset and another with the stand subset. Based on the analysis, ordination plots were used to represent the relationships among objects (trees or plots), where additional factors such as, growth zones, tree species, and dominant species were used in the identification of objects to help with the visualization of results. Here, object clouds were delimited using the extreme objects positions in relation to each factor level.

PCoA, in contrast to NMDS, that orders the data in axes according to their contribution to total variability based on eigenvalues. PCoA does allow the use of any similarity/dissimilarity matrix, which represents the relationship between objects or variables. For PCoA in this study, both the tree and stand subsets were analyzed together, where the number of coordinates were selected by examining a scree plot. Here, the same modified Gower dissimilarity matrix and graphical output generated for NMDS were also used.

PCA is the main eigenvalue-eigenvector-based method that describe the variation of correlated quantitative variables, in terms of a new set of uncorrelated variables or axes. As with PCoA, these new axis or principal components (PC) are derived in decreasing order of importance according to their total variance explained (Everitt et al. 2011). PCA uses a similarity/dissimilarity matrix calculated based on the variances and covariances of the variables, or the correlations between them (Borcard et al. 2011); thus PCA uses the Euclidean distance to represent the objects. Representation of the objects (scores) and the original variables (factor loadings) can be done through a PCA Biplot. In this study, an exploratory PCA was used to identify relevant continuous variables. Here, two analyses were performed, one used both the tree and stand subsets, and a second only stand variables with the inclusion of AIDBH plot average. For this method, in order to achieve normality assumptions, some variables were transformed using the inverse, logarithm, square root, or arc-sin functions. After transformations, the variables were standardized. The number of PCA axes was determined by using (Quinn and Keough 2002): 1) a latent root criterion by keeping components with eigenvalues >1 ; 2) through a scree plot; and 3) by having a

percent variance explained greater than 90%. In order to interpret the relationship among objects, distance-Biplot were used.

K-means clustering is a method for dividing a dataset into K different and non-overlapping clusters. In this study, this method was implemented only with the subset of stand variables standardized. The determination of the number of clusters was done through scree plots of: 1) within-cluster sum of squares; and 2) silhouette method (Everitt and Hothorn 2011). In addition, a Multiple Response Permutation Procedure (MRPP) was performed to test the differences between clusters (McCune et al. 2002). Pearson chi-squared test of association was used to establish any association between clusters and growth zones as defined by Gezan and Moreno (1999). If any relationship was found, a post-hoc test for all pairwise comparisons with Bonferroni corrections of the P-values was performed (MacDonald and Gardner 2000).

All statistical analyses were performed using R v. 3.2.2 (R Core Team 2015). For all analyses, where required, a significance level of 5% was used. The R packages used for dissimilarities matrices, NMDS, PCoA, and MRPP was *vegan* version 2.4-1 (Oksanen et al. 2013), and for Silhouette plot *cluster* version 2.0-4 (Maechler 2013).

Results

The variables of tree and stand subset presented, as expected, a high level of correlation between some variables, such as BA-BAN, DH-RS, BA-SDI, BAL-BALn, DH-QD, A-DA, RS-QD, and DBH-H, with values >0.83 (see Figure C-1).

NMDS, using categorical and continuous data for the tree subset, converged with a stress goodness-of-fit value of 0.09, corresponding to a good representation of original distance matrix. For this subset, NMDS ordination with the two selected axes did not identify different groups, only a single group surrounded by isolated data points

(Figure 2-1). For this representation, there are no clear patterns for the different growth zones or species (growth zones represent almost exactly the same tree cloud, and *N. obliqua* had minor differences related to *N. alpina* and *N. dombeyi*), indicating that these factors do not explain the variability of the two selected axes.

Another NMDS was performed only with the stand subset (Figure 2-2), resulting in a stress value of 0.05, which represents a good description of the original distance matrix (Everitt and Hothorn 2011). As with the tree subset, the clouds for growth zones were overlapping; however, the dominant specie had a small difference between *N. obliqua* and *N. dombeyi*, while *N. alpina* is practically contained within the other two dominant species clouds.

The ordination method of PCoA explained ~90% of the cumulative variance using two dimensions (PCoA axes) when both tree and stand variables were evaluated together (Figure C-2). Interestingly, trees belonging to a plot were very close in these new coordinates (Figure 2-3); hence, it appears that the variability associated with the stand variables dominates over the one with individual tree variables. In addition, the results from PCoA confirms the findings from NMDS, with overlapping growth zones, and greater similarities between *N. alpina* and *N. dombeyi*.

The PCA method was also used, based on both subsets, to assist in identification of those variables that better explain the variability of the sample population. To achieve normality, the variables DBH, BA, BALn, and AIDBH were transformed with a square root, and BALr was transformed with the arcsine function. The results of PCA indicated that four or five principal components (PC) explain the majority of the variability (88% and 92%, respectively) found in these *Nothofagus* forests

(Figure C-3). Factor loadings for these five PC are presented in Table 2-2. PC1 with 37% of the variance explained, can be described as “stand development stage” axis; PC2 (24%) a “tree-to-tree competition” axis; PC3 (16%) a “stocking and stand competition” axis; PC4 (11%) a “site productivity” axis; and PC5 (4%) a “tree DBH growth” axis.

PC1 is explained mostly by the stand variables BA, QD, Hd, Ad, BAN, SDI, and RS. The other stand variables N and SI were relevant only on the third and fourth component, respectively. In relation to tree variables, it was observed a high contribution in PC1 for variables related to tree size (DBH, H, and A), while tree-to-tree competition variables (BAL, BALn, and BALr) were important in PC2 with almost identical factor loadings, and AIDBH was also strongly associated with this component but with opposite direction (i.e., negative factor loading). A Biplot for PC1 and PC2 is presented in Figure 2-4a that explains 61% of the variability. Other Biplots are shown in Figure C-4 and C-5.

Recall, PC1 is mainly representing stand development stages, and PC2 tree-to-tree competition. The scores associated with each observation show a wide range of tree-to-tree competition at later stand development stages (left side of Figure 2-4a), and for early stand development stages, this tree-to-tree competition is more uniform (right side of Figure 2-4a). This Biplot also shows clearly the negative association between AIDBH and tree competition variables; where, as expected, high individual competition means low diameter growth. Surprisingly, productivity (represented by SI) had a negligible effect in explaining the variability of this population. From this Biplot, groups of correlated variables can be easily identified: Hd-Ad-QD, BA-BAN-RS, BAL-BALn-BALr-

AIDBH, and A-N. No clear patterns of the distribution of the trees were detected according to growth zone and species.

A second PCA was done with only the variables from stand subset together with average plot AIDBH. Factor loadings of four components are shown in Table 2-3. The meaning of each PC, as expected, were almost identical to the ones described earlier for both subsets, with PC1 (43% variance explained) as “stand development stage” axis; PC2 (24%) a “stocking and stand competition” axis; PC3 (18%) a “site productivity and growth” axis; and PC4 (8%) as “stand competition and growth” axis. Two first PC explained a 67% of the total variance (Figure C-6) and they are represented by a Biplot in Figure 2-4b. Here, the plot observations (scores) present a higher variability in stand competition at later stand development stages (right side of Figure 2-4b), and *vice versa*.

The implementation of K-means clustering with the subset of stand variables identified two clusters that were selected using the within-cluster sum of squares and silhouette methods (Figure C-7). The first cluster was associated with high values of N and RS, while the second with high values of BA, BAN, DA, DH, and QD (Figure 2-4b and Table 2-4). Hence, the main differences among plots were associated to stand development stages, where, for one side, there are young forests with a high tree density, and on the other, older forests with basal area concentrated in few individuals. An MRPP test statistic of ‘chance corrected within-group agreement’ (As) of 0.16 was obtained for these two clusters, indicating strong differences between them. In addition, the statistics of observed delta was 3.51 ($P < 0.001$) versus an expected delta of 4.17 indicating strong significant differences ($P < 0.001$) in terms of stand variables (Figure

C-8). Association between clusters and growth zones was evaluated with a Pearson chi-squared statistics, with a value of 8.69 ($P = 0.034$). The post-hoc pairwise testing only established a small association between clusters and the growth zones 1 and 2 (Table B-1). These two clusters were contrasted in a map with growth zones, where no clear patterns were noted (Figure 2-5).

Discussion

For second growth RORACO forests, high levels of variability were found in the tree and stand subsets, with wide ranges across growth zones in terms of averages and standard deviation (Table 2-1). This variability originates from several factors including biophysics variables, ecology and species characteristics, disturbance, etc. (Donoso 1993; Chauchard and Sbrancia 2003; Echeverria and Lara 2004; Luebert and Plissock 2006). Stand variability is not only found between provenances but also within stands at the individual level (Donoso 1993).

In Chile, the Andes mountain range vegetation has a marked altitudinal gradient, influenced by temperature and rainfall, where higher altitudes are characterized by low temperatures and increased rainfall (Luebert and Plissock 2006). Together with this altitudinal gradient, that in Chile is closely related to longitude, the latitudinal gradient strongly affects the vegetative growth period (Donoso 2003). To simplify the effect of biophysical variables on a forest resource, it is useful to define relatively homogeneous geographical growth zones. The zoning used in this study was defined by combining different sources of information related to (Gezan and Moreno 1999): climate, soil, vegetation distribution, and site productivity. These biophysical variables have been used in many other zoning studies for *Nothofagus* (Novoa and Villaseca 1989; Donoso

et al. 1993; Schlatter and Gerding 1995; Echeverria and Lara 2004; Thiers et al. 2008; Esse et al. 2013).

Variability found within a stand is related to ecological factors and/or forest dynamics. The results from the unsupervised multivariate methods identified important factors at the tree and stand level for the RORACO forest. Both of the PCA presented a high percentage of variance explained related to stand development stages, ranging from 37% to 43%. In addition, two clear clusters were found, which represent two conditions of stand development stages: 1) young forests with a high tree density; and, 2) adult forests with high basal area concentrated in few, large trees. Other authors have reported differences based on stand developmental stage. For example, Donoso (1993) found, for *Nothofagus*, changes in the diameter distribution, from unimodal to bimodal, as the stand ages, reflecting additional heterogeneity that is present in these forests. This heterogeneity is incremented by the interaction with biophysical variables such as climate, altitude, and latitude.

The other multivariate methods, NMDS analysis and PCoA evidenced a differentiation among species, but not for growth zones. Lusk and Ortega (2003) found no differences in productivity for comparisons based on the same zoning. The results of NMDS and PCoA at stand level, showed that in the multivariate space *N. alpina* was intermediate between the others two species, being closer to *N. dombeyi* than *N. obliqua*. Other authors have also found important differences between these species. Chauchard and Sbrancia (2003) in relation to shadow tolerant, indicated that *N. alpina* is the most tolerant, followed by *N. dombeyi*, and then *N. obliqua*. In terms of tree-to-tree competition, the same authors reported that at older ages, *N. dombeyi* stands

present higher density than *N. obliqua*, which defines its sociological status at a juvenile phase and remains in this strata. Donoso (1993) states that *N. obliqua* have a strong natural thinning at 40-50 years of age due to its low ability for tree-to-tree competition. In the present study, tree-to-tree competition, as described by the PC2 explained 24% of the total variability for both subsets (tree and stand). This component was mainly described by the variables of BAL, BALr, and BALn.

Several studies have identify a wide range of factors that influence the variability found in the RORACO forest type. This study, contributed to better understanding the underlying variability for these forest through the use of multivariate analysis (NMDS, PCoA and PCA) based on tree and stand variables typically obtained from forest inventories. The findings indicated that site productivity and growth zones do not explain much the variability of this sampled population. However, stand development stages, tree-to-tree competition, and size-tree attributes are critical variables to explain this variability. Also, growth zones were identified to have relevant importance at the macro region level.

The biophysics variables were not directly included in this study, being represented by growth zones, so that was a limitation to try of explain the phenotypic variability only with tree and stand variables. However, this study was focused to understand the behavior of these variables as first step to development a growth model for RORACO forest type.

Table 2-1. Summary statistics of this study dataset for individual and stand variables in each growth zone.

Variable	Zone 1			Zone 2			Zone 3			Zone 4		
	n	mean (SD)	range									
DBH	321	15.4 (9.3)	5-60	317	18.7 (11.0)	5-62	194	19.5 (11.5)	5-59	276	17.2 (9.6)	5-45
H	321	14.4 (6.0)	5-35	317	17.2 (7.0)	5-38	194	18.0 (7.3)	4-45	276	15.0 (6.7)	4-39
A	321	29.5 (14.8)	8-104	317	34.0 (14.8)	10-83	194	30.8 (14.0)	9-65	276	34.2 (12.9)	10-68
BAL	321	20.3 (15.1)	0-74	317	24.2 (17.2)	0-83	194	21.6 (15.9)	0-64	276	23.9 (18.9)	0-92
BALn	321	16.0 (12.6)	0-56	317	19.8 (15.2)	0-82	194	17.1 (13.3)	0-55	276	18.3 (15.2)	0-63
SS	321	2.2 (1.0)	1-4	317	2.3 (1.1)	1-4	194	2.2 (1.0)	1-4	276	2.3 (1.0)	1-4
BALr	321	0.5 (0.3)	0-1	317	0.6 (0.3)	0-1	194	0.5 (0.3)	0-1	276	0.6 (0.3)	0-1
AIDBH	321	3.0 (2.0)	0-11	317	2.9 (2.1)	0-12	194	3.8 (2.2)	0-11	276	2.7 (1.7)	0-10
BA	42	37.9 (15.9)	10-79	46	42.9 (14.8)	16-84	29	43.2 (14.8)	19-80	41	46.9 (19.8)	11-98
N	42	2,821 (1,325)	760-5,600	46	2,337 (1,245)	320-5,000	29	2,226 (1,104)	640-4,640	41	2,794 (1,135)	760-5,560
QD	42	14.3 (5.7)	7-31	46	17.3 (6.6)	7-33	29	17.3 (6.1)	8-32	41	15.2 (4.5)	8-26
Hd	42	19.1 (6.4)	8-30	46	22.3 (21.2)	11-35	29	22.9 (7.2)	11-42	41	19.8 (6.3)	9-37
Ad	42	36.4 (16.9)	14-87	46	42.4 (14.6)	13-77	29	37.2 (13.3)	16-64	41	40.7 (12.0)	16-68
SI	42	11.8 (3.8)	4-20	46	12.5 (4.1)	4-24	29	13.7 (3.4)	7-23	41	10.2 (3.7)	2-18
BAN	42	31.8 (12.4)	10-62	46	37.2 (14.6)	14-83	29	37.4 (14.0)	8-69	41	40.3 (18.3)	9-90
SDI	42	1,054 (344)	410-1,895	46	1,089 (328)	414-1,906	29	1,081 (304)	717-1,830	41	1,245 (438)	406-2,305
RS	42	0.1 (0.0)	0.1-0.2	46	0.1 (0.0)	0.1-0.2	29	0.1 (0.0)	0.1-0.2	41	0.1 (0.0)	0.1-0.2

Note: n, number of observations; SD, standard deviation; DBH, diameter breast height (cm); H, total height (m); A, breast height age (years); BAL, basal area of larger trees ($\text{m}^2 \text{ha}^{-1}$); BALn, basal area of larger trees for *Nothofagus* ($\text{m}^2 \text{ha}^{-1}$); SS, sociological status; BALr, relative BAL; AIDBH, annual increment in DBH (mm year^{-1}); BA, stand basal area ($\text{m}^2 \text{ha}^{-1}$); N, number of trees (trees ha^{-1}); QD, quadratic diameter (cm); Hd, dominant height (m); Ad, dominant breast height stand age (years); SI, site index (m); BAN, basal area of *Nothofagus* ($\text{m}^2 \text{ha}^{-1}$); SDI, stand density index (trees ha^{-1}); RS, relative spacing index.

Table 2-2. Factor loading for the first five principal component (PC) obtained from the tree and stand subsets together. Values in parenthesis correspond to the proportion of the variance explained by each component.

Variable	Transformation	PC1 (37%)	PC2 (24%)	PC3 (16%)	PC4 (11%)	PC5 (4%)
DBH	square root	-0.271	0.366	-	-	-
H	-	-0.304	0.263	-	-0.151	0.220
A	-	-0.307	0.152	-	0.369	0.144
BAL	-	-0.145	-0.452	-	-	-0.194
BALn	square root	-0.103	-0.438	0.115	-	-
BALr	arcsin	-	-0.454	0.220	-	-
AIDBH	square root	-	0.315	-0.199	-0.277	-0.561
BA	square root	-0.341	-0.127	-0.286	-	-0.165
N	-	0.176	-	-0.509	-	0.220
QD	-	-0.343	-	0.279	-	-0.227
Hd	-	-0.354	-	0.187	-0.231	0.213
Ad	-	-0.327	-	0.148	0.368	-
SI	-	-	-	-	-0.714	0.151
BAN	-	-0.328	-0.109	-0.216	-	-0.359
SDI	-	-0.204	-0.151	-0.494	-	-
RS	-	0.243	0.104	0.336	0.188	-0.493

Note: Factor loadings smaller than 0.1 in absolute value are not shown, and those larger than 0.3, in absolute value, are in bold. DBH, diameter breast height (cm); H, total height (m); A, breast height age (years); BAL, basal area of larger trees ($\text{m}^2 \text{ha}^{-1}$); BALn, basal area of larger trees for *Nothofagus* ($\text{m}^2 \text{ha}^{-1}$); SS, sociological status; BALr, relative BAL; AIDBH, annual increment in DBH (mm year^{-1}); BA, stand basal area ($\text{m}^2 \text{ha}^{-1}$); N, number of trees (trees ha^{-1}); QD, quadratic diameter (cm); Hd, dominant height (m); Ad, dominant breast height stand age (years); SI, site index (m); BAN, basal area of *Nothofagus* ($\text{m}^2 \text{ha}^{-1}$); SDI, stand density index (trees ha^{-1}); RS, relative spacing index.

Table 2-3. Factor loading for the first five principal component (PC) obtained from average plot AIDBH and stand subset. Values in parenthesis correspond to the proportion of the variance explained by each component.

Variable	Transformation	PC1 (43%)	PC2 (24%)	PC3 (18%)	PC4 (8%)
AIDBH	square root	-	-	-0.570	0.555
BA	square root	0.429	0.263	-	0.164
N	-	-0.200	0.559	-	-0.150
QD	-	0.396	-0.324	-	0.171
Hd	-	0.391	-0.275	-0.173	-0.317
Ad	-	0.362	-0.150	0.380	-
SI	-	-	-0.180	-0.668	-0.261
BAN	-	0.406	0.149	-	0.372
SDI	-	0.279	0.514	-	0.101
RS	-	-0.294	-0.308	0.226	0.543

Note: Factor loadings smaller than 0.1 in absolute value are not shown, and those larger than 0.3, in absolute value, are in bold. DBH, diameter breast height (cm); H, total height (m); A, breast height age (years); BAL, basal area of larger trees ($\text{m}^2 \text{ha}^{-1}$); BALn, basal area of larger trees for *Nothofagus* ($\text{m}^2 \text{ha}^{-1}$); SS, sociological status; BALr, relative BAL; AIDBH, annual increment in DBH (mm year^{-1}); BA, stand basal area ($\text{m}^2 \text{ha}^{-1}$); N, number of trees (trees ha^{-1}); QD, quadratic diameter (cm); Hd, dominant height (m); Ad, dominant breast height stand age (years); SI, site index (m); BAN, basal area of *Nothofagus* ($\text{m}^2 \text{ha}^{-1}$); SDI, stand density index (trees ha^{-1}); RS, relative spacing index.

Table 2-4. Mean values of stand subset and average plot AIDBH for each K-means cluster.

Variable	Cluster 1	Cluster 2	Total
AIDBH	4.3	3.4	3.9
BA	33.0	52.8	42.5
N	3226.3	1845.9	2563.0
QD	11.6	20.7	16.0
Hd	16.1	26.1	20.9
Ad	30.2	49.9	39.6
SI	11.6	12.2	11.9
BAN	27.4	46.0	36.3
SDI	1029.4	1209.2	1115.8
RS	0.1	0.1	0.1

Note: AIDBH, annual increment in DBH (mm year^{-1}); BA, stand basal area ($\text{m}^2 \text{ha}^{-1}$); N, number of trees (trees ha^{-1}); QD, quadratic diameter (cm); Hd, dominant height (m); Ad, dominant breast height stand age (years); SI, site index (m); BAN, basal area of *Nothofagus* ($\text{m}^2 \text{ha}^{-1}$); SDI, stand density index (trees ha^{-1}); RS, relative spacing index.

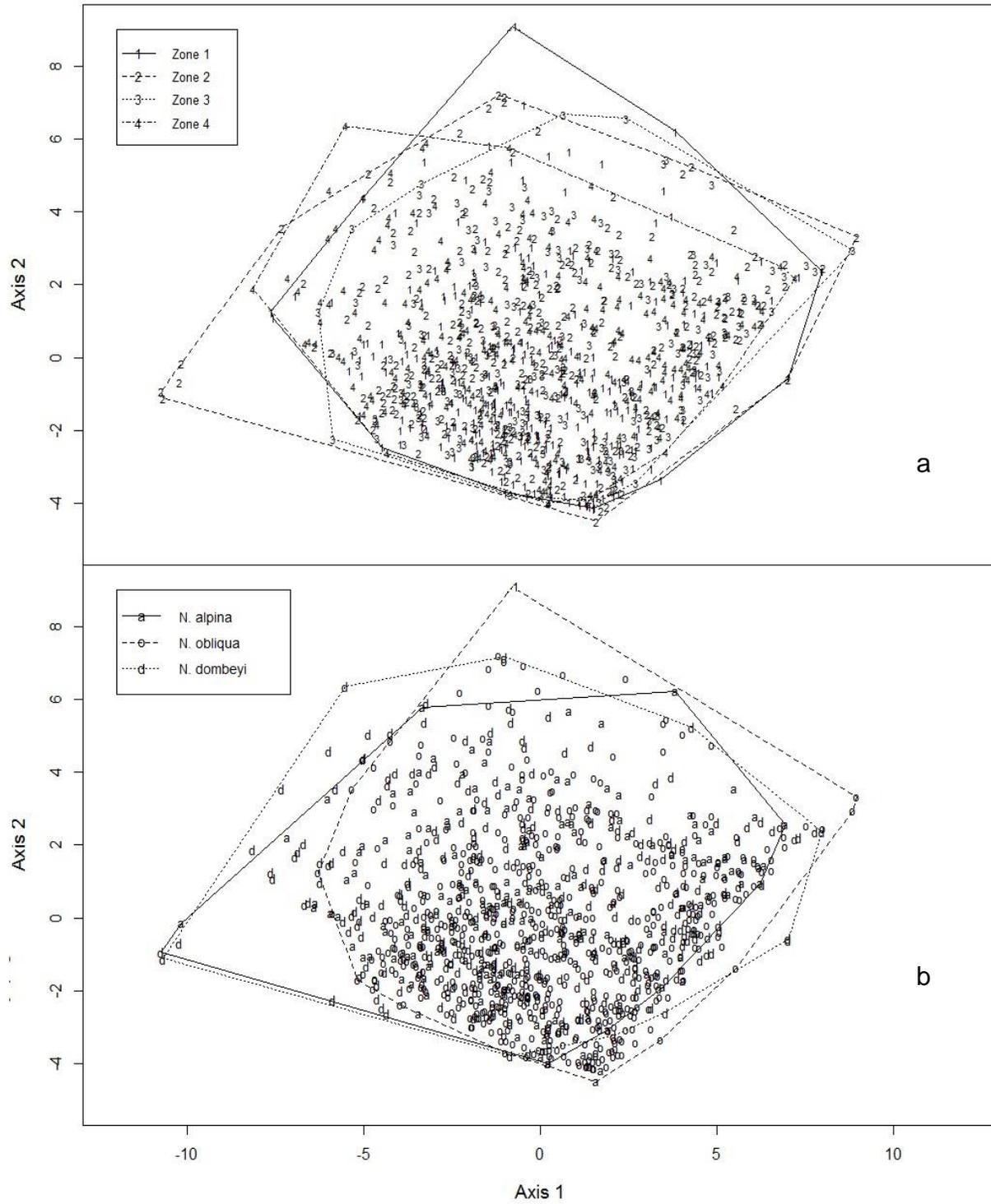


Figure 2-1. Non-Metric Multidimensional Analysis (NMDS) using the tree subset, identifying growth zones (a) and species (b).

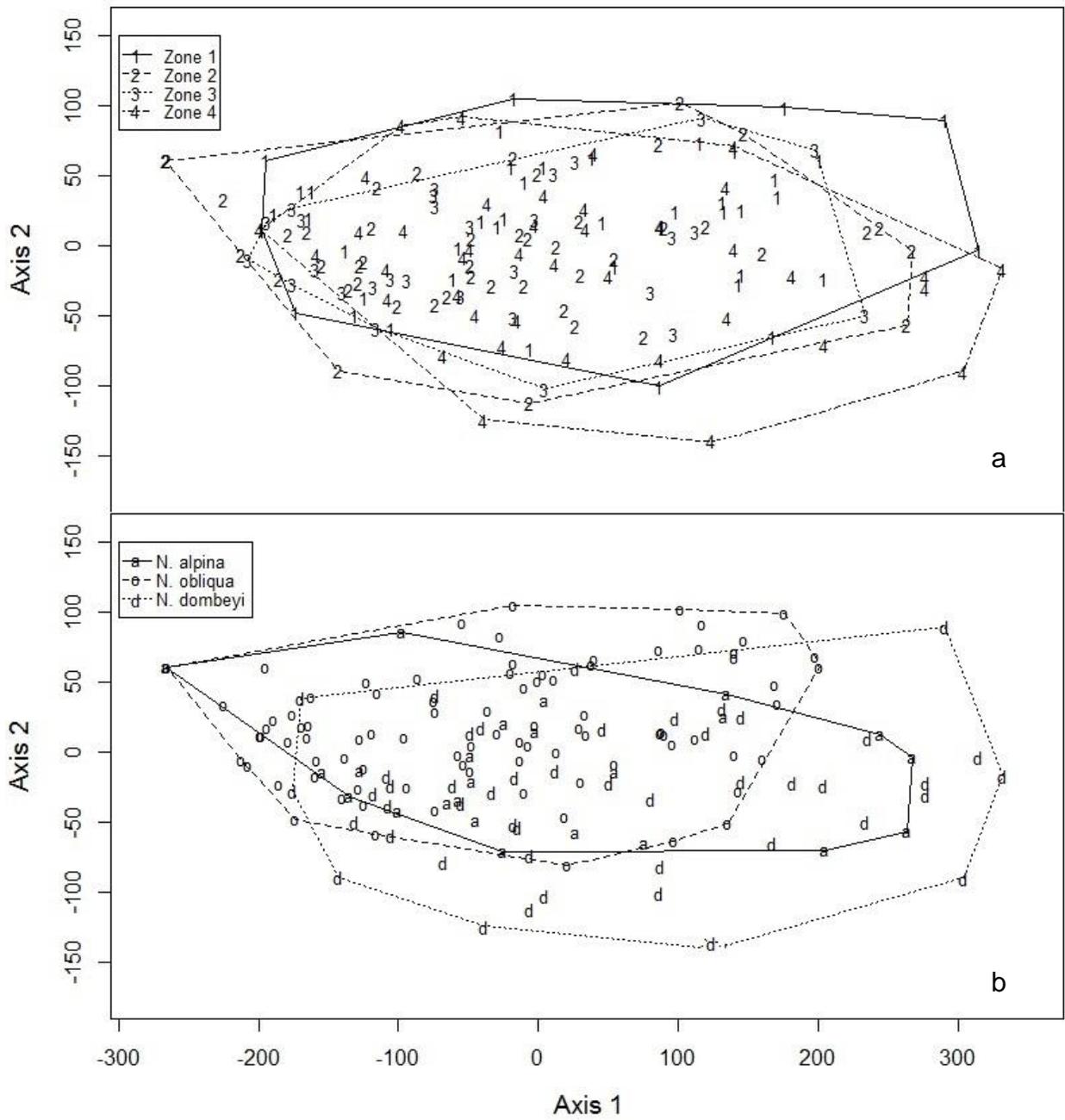


Figure 2-2. Non-Metric Multidimensional Analysis (NMDS) using the stand subset, identifying growth zones (a) and dominant species (b).

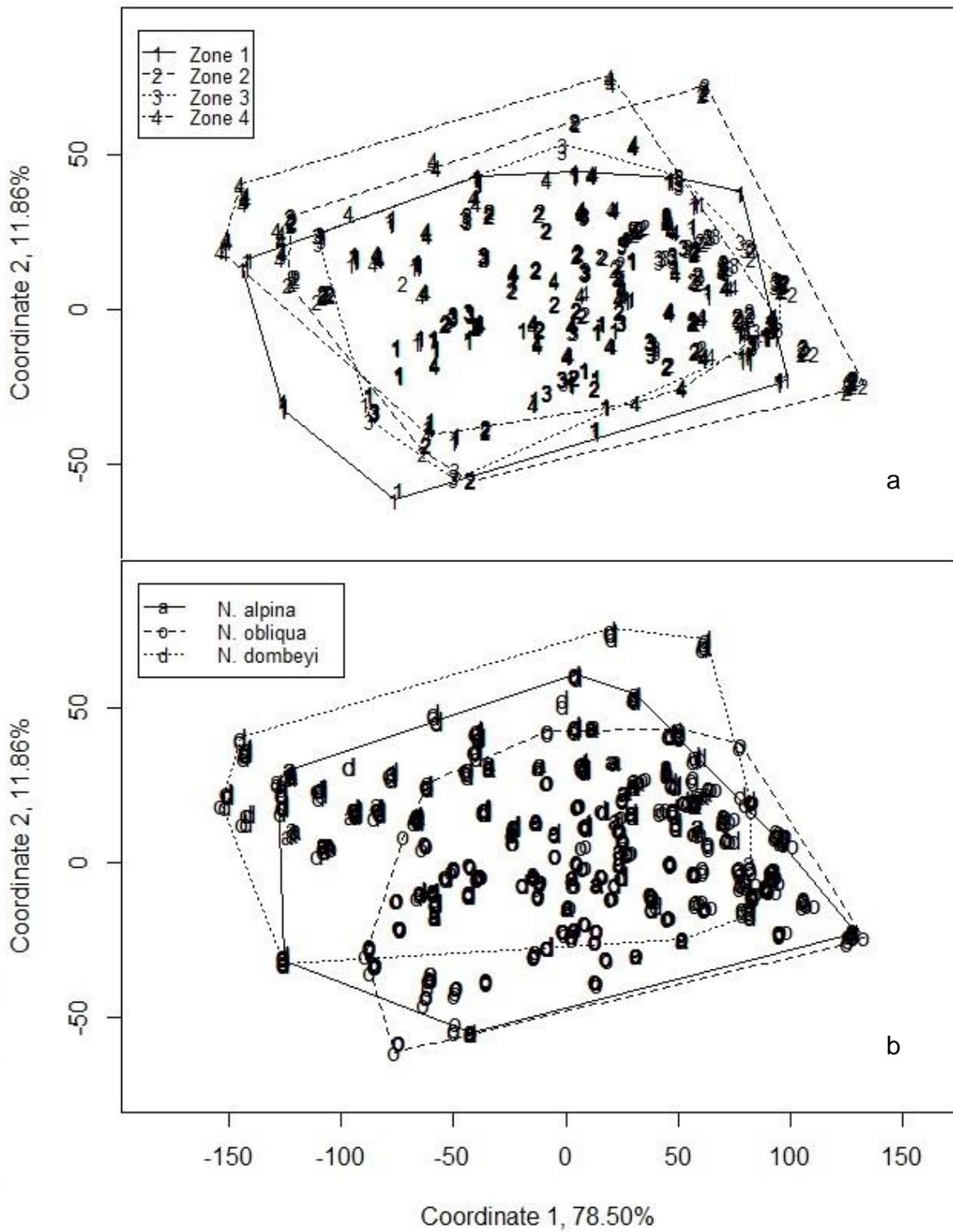


Figure 2-3. Principal coordinates analysis (PCoA) using the tree and stand subsets together, identifying growth zones (a) and species (b).

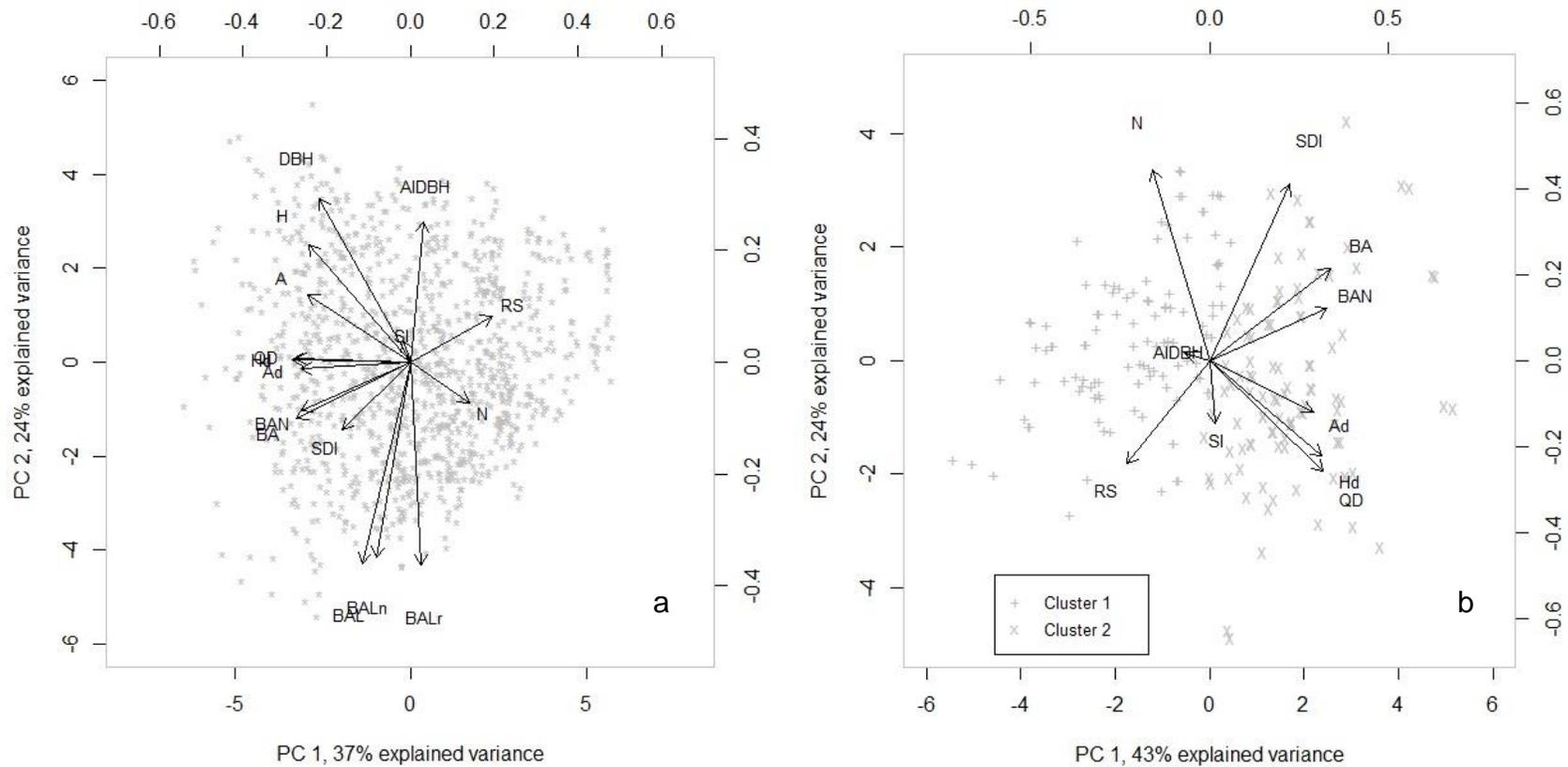


Figure 2-4. First two PCA axes with scores and factor loadings obtained using, the tree and stand subsets (a), stand subset together with average plot AIDBH. The two K-means clusters are also identified (b).

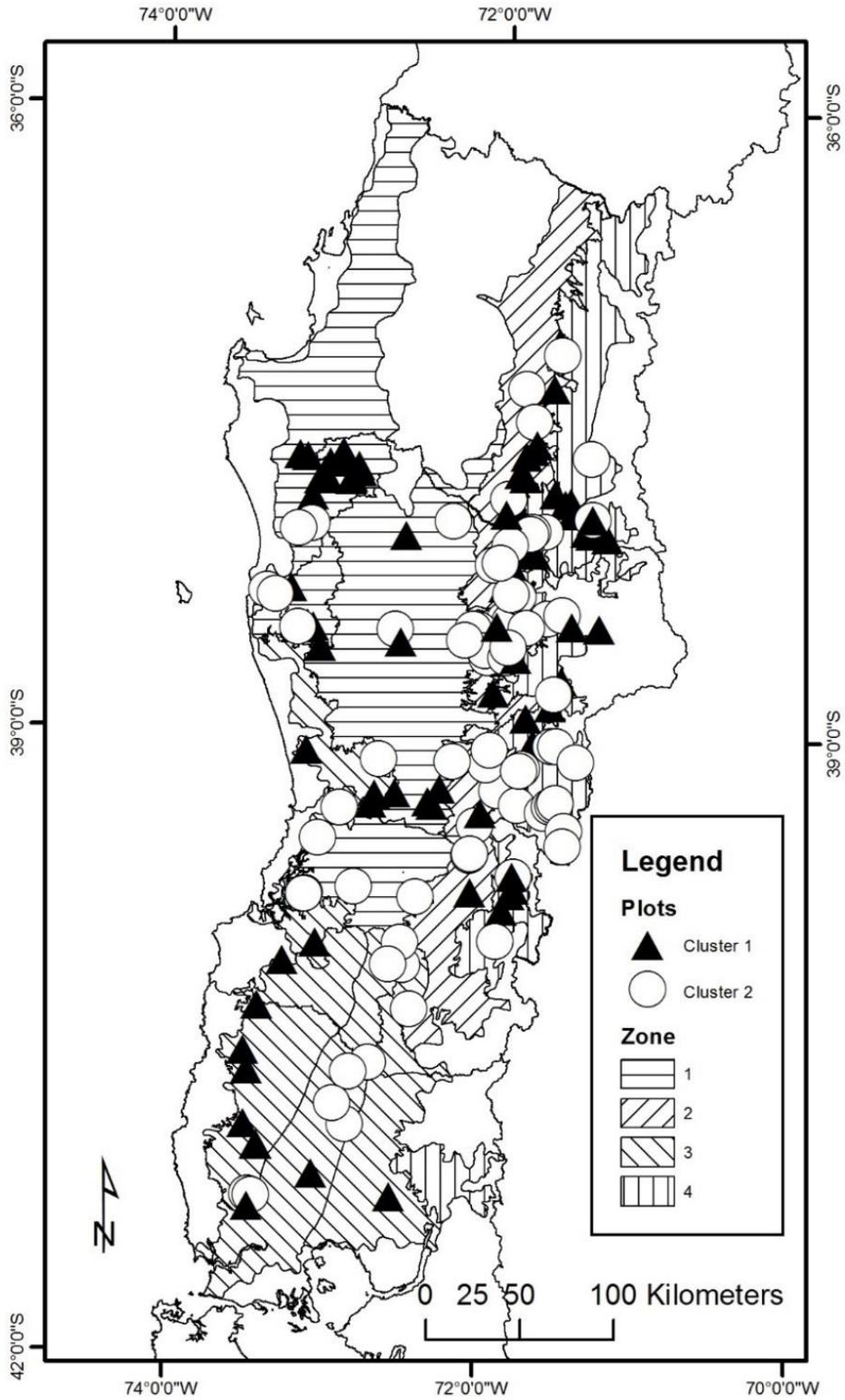


Figure 2-5. Geographical localization of plots identified according to their two K-means clusters in relation to growth zones.

CHAPTER 3 INDIVIDUAL-TREE DIAMETER GROWTH MODELS OF MIXED *NOTHOFAGUS* SECOND GROWTH FORESTS IN SOUTHERN CHILE

Background

Southamerican *Nothofagus* species dominate temperate and sub-Antarctic forests of Chile and Argentina from 33° to 56° south latitude. In Chile, there are nine species of this genera, including three evergreens, and six deciduous (Donoso 1987; Donoso 1993). The second growth forests of roble (*Nothofagus obliqua* (Mirb.) Oerst.), raulí (*N. alpina* (Poepp. & Endl.) Oerst.), and coihue (*N. dombeyi* (Mirb.) Oerst.) known locally as the “RORACO” forest type, is one of the most important natural mixed forests of Chile. According to Chile’s most recent inventory, there are about 1.4 million hectares of these forests, corresponding to 10.8% of the total native forest area of the country (CONAF 2011). In 2015, the saw timber from RORACO forest type was approximately 45% of the national timber production (INFOR 2016). The volume growth rates for these forests range between 6 and 10 m³ ha⁻¹ year⁻¹ (Donoso et al. 2004), corresponding to some of the highest values for southamerican temperate native forests. Given its large geographical area, timber market value and attractive growth rates, this resource presents high economic potential value within the chilean forestry sector (Gezan et al. 2007).

High variability is found in the RORACO forest type due to different factors, such as large geographical area, biological diversity with many companion species, varying (or lack of) management, anthropic disturbances and site productivity (Chapter 2). For example, these three *Nothofagus* species have different altitude ranges: *N. obliqua* is commonly found between 100 and 600 m a.s.l., *N. alpina* between 600 and 900 m a.s.l., and *N. dombeyi* is frequently found over 900 m a.s.l. Additionally, it is possible to find

different ecotones as single or mixed composition forests (INFOR 2013). Echeverria and Lara (2004) found that the main environmental factors that affect growth rates for RORACO forest are: longitude, climatic variables (mean annual rainfall, summer humidity index, frost-free period), and edaphic variables (sand and silt contents) that accounted for 70.4% of the total variance. Definition of growth zones helps to manage this variability, with several reported approaches for the RORACO forest type (Donoso et al. 1993; Gezan and Moreno 1999; Echeverria and Lara 2004); where within a given zone, it is expected to find similar growth patterns, silvicultural treatment responses, and companion species, among others.

In order to obtain economic, social and ecological sustainability of these forest, forest managers and owners need information regarding present conditions, and future state under varying silvicultural activities. Forest inventories gives critical information of the current stand, and growth and yield (G&Y) models provide with future conditions and help with modeling forest behavior (Shortt and Burkhart 1996). Currently, there is a lack of general G&Y models useful for the geographical range of the RORACO forest type. Some specific studies present models; however, these are limited to small geographical areas and specific conditions (Donoso et al. 1993; Gezan and Moreno 1999; Moreno 2001; Echeverria and Lara 2004; Salas and Garcia 2006; Ugalde 2006; Gezan et al. 2007).

Birdsey (1990) states that G&Y models are useful tools for inventory updates and decision making regarding silvicultural management, for short- or long-term targets. G&Y models have been used mainly for plantation or even ages, single species forests as their complexity for mixed or uneven-aged forests increases considerably. Some of

the challenges for the construction of mixed-forest models include: growth rates that differ among species, stands that include shade tolerant and intolerant species, and requirement for selective harvests with a continuous forest cover. All of these, and other factors, increase the uncertainties and rise the complexity and lower prediction accuracy quality of mixed forests G&Y models (Burkhart and Tome 2012).

Numerous G&Y models approaches have been based on: stand initial conditions, species composition, objectives and needs of forest managers, among others (Burkhart and Tome 2012). These approaches range from models that only provide with aggregated stand variables (whole-stand level) to models with information for each tree (individual-tree level) (Burkhart and Tome 2012). Whole-stand level are the most common G&Y models, and these can be extended to produce a stand table through the generation of a diameter distribution with several size classes (Vanclay 1994). Individual-tree models are the highest detail level for modeling growth and yield, and they can be used as feedback for whole-stand models (Burkhart 2003; Burkhart and Tome 2012; Shortt and Burkhart 1996). Individual-tree models have been widely used for different forest ecosystems, including plantations, even-aged mixed hardwood stands, uneven-aged mixed species, tropical forests, boreal forests, and temperate forests (Harrison et al. 1986; Murphy and Graney 1998; Huang and Titus 1999; Andreassen and Tomter 2003; Pukkala et al. 2009; Nunes et al. 2011; Ma and Lei 2015; Salas et al. 2016). For individual-tree models the projection unit is the individual tree, with the distance-dependent and distance-independent model procedures, which differ in the inclusion of the physical coordinates of each tree within a plot. Individual-tree models offer high flexibility to describe and project any stand structure and allow for the

incorporation of more reasonable process-based modules, such as thinning, differentiation of growth patterns and, in the case of mixed forest, responses for different species to competition and mortality, for example. However, whole-stand models often give more accurately estimates and projections of the overall stand variables (Burkhart and Tome 2012; Gonzalez-Benecke et al. 2012), but at a lower level of resolution.

An individual-tree G&Y model consists of three components (Burkhart and Tome 2012): 1) individual-tree diameter breast height (DBH) or basal area growth equation; 2) individual-tree height growth equation or, alternatively, a height-diameter relationship; and 3) mortality or survival component. The DBH or basal area growth is modeled using a function of predictors, tree and stand attributes, including: DBH, total height and age; dominant age, dominant height or site index (as a measure of site productivity); basal area of larger trees and sociological status (as a measure competition index); stand basal area, number of trees per unit area, and quadratic diameter (Huang and Titus 1995; Burkhart and Tome 2012; Ma and Lei 2015). For mixed forests and large areas, these growth equations are tuned-up or calibrated to allow for different species and/or geographical zones (Pukkala et al. 2009). For mortality/survival, often logistic models are used that also consider individual-tree attributes, together with a measure of competition and stand attributes (Murphy and Graney 1998; Nunes et al. 2011; Burkhart and Tome 2012; Ma and Lei 2015).

Most G&Y models, and particularly, individual-tree growth ones, are obtained by fitting multiple linear regressions with many potential predictors in their original units or with transformations. These predictors often suffer from multicollinearity (Welham et al. 2015). There are several statistical methods for selecting predictors including: best

subset, forward, backward, forward stepwise, backward stepwise, and hybrid selection procedures (James et al. 2013). All of these selection methods have the same logic, which is to fit a multiple linear model with the selected subset of predictors (Hastie et al. 2009). Alternatively, some shrinkage or regularization method has been commonly recommended, where a multiple linear model is fitted with all predictors simultaneously incorporating a constraint that shrink the model coefficient toward zero. The advantage of these procedures is related to the reduction of the variance of each coefficient estimate with this regularization (Tibshirani 1996; Hastie et al. 2009; Kyung et al. 2010; James et al. 2013). The two best-known of such procedures are: ridge regression (RR) and least absolute shrinkage and selection operator (LASSO) (Kyung et al. 2010; Hastie et al. 2009).

RR is similar to least squares, except for the incorporating of a ℓ_2 shrinkage penalty associated with a λ tuning parameter (James et al. 2013). The advantage of this method is related to the bias-variance trade-off, where for increases in λ the variance of the coefficient estimates decreases by inducing an, often small, bias in the estimates (Hastie et al. 2009; James et al. 2013). The main disadvantage of RR is that it includes all predictors in the final model, which is, among other issues a problem for model interpretation related to the predictors importance. LASSO, in contrast, uses a ℓ_1 shrinkage penalty that makes some coefficient estimates exactly zero, particularly when the tuning parameter λ is large; thus, LASSO at the same time performs shrinkage and variable selection (Hastie et al. 2009; Kyung et al. 2010; James et al. 2013). One limitation for LASSO, is that under non-Bayesian statistics, due the penalized

estimation, it does not provide P-values, confidence intervals or standard error of the regression coefficients (Tibshirani 1996; Wu et al. 2009; Lockhart et al. 2014).

The lack of growth information for the RORACO forestry type and social and economic importance of this resource suggest the necessity of improvement in growth models for full geographic area of these forests. The main aim of this study is to develop an individual-tree growth model to estimate annual increment in DBH for second growth forest stands of *N. alpina*, *N. obliqua*, and *N. dombeyi*, including a set of tree and stand level predictors to explain diameter growth of these species. The data originates from a study realized by Universidad Austral de Chile (Ortega and Gezan 1998), over an ecologically and geographically diverse area (> 14,000 km², including four growth zones). The specific study objectives are: 1) to compare different multiple linear fitting procedures to obtain the best individual-tree growth model; 2) to generalize diameter growth models to incorporate growth zone and specie, and 3) to evaluate the final diameter growth model using an independent validation dataset.

Materials and Methods

Data Description

The dataset used to fit the individual-tree diameter breast height (DBH) growth model is from a study by the Universidad Austral de Chile (Ortega and Gezan 1998) consisting on two plot networks. One, with a total of 128 permanent plots (PP), and another with 60 temporary plots (TP) (Figure 3-1). The data were collected under a population stratification, based on a national forest cadastre (CONAF et al. 1999), that considered different stand conditions, including: RORACO second growth forest type dominated by their *Nothofagus spp.*, canopy coverage >50%, and with minimum or no stand disturbance.

Each PP had an area of 500 m², but for higher densities (> 4,800 trees ha⁻¹) the plot area was reduced to 250 m². All trees with 5 cm or more in DBH and that were taller than 2 m in height were measured. Total height was obtained in a subsample of 15 trees per plot, and tree increment cores were extracted from this subsample at breast height to obtain breast height age and past diameter growth. The TP were formed by two circular subplots with an area of 125 m² each. All trees with DBH > 5 cm and taller than 2 m were measured. In each subplot, two trees were felled for complete stem analysis. Further details of these datasets is presented in Gezan and Moreno (1999). This study used the annual average DBH growth based on the last years, which originated from the increment cores (PP) and tree sections (TP) at breast height of the penultimate and antepenultimate growth years. The last growth year was discarded as plots were established at different months within a year, without all completing the current growing season.

The dataset included the following tree-level attributes: species (SP), diameter breast height (DBH, cm), height (H, m), breast height age (A, year), basal area of larger trees (BAL, m² ha⁻¹), basal area of *Nothofagus* of larger trees (BAL_N, m² ha⁻¹), sociological status (SS, defined according to vertical stratification with dominant (1), codominant (2), intermediate (3) or suppressed (4)), and relative basal area of larger trees (BAL_r, with BAL_r = BAL/BA). The response variable, as indicated earlier, corresponded to the average annual increment in DBH (AIDBH, mm year⁻¹). Summary statistics of the above attributes are presented in Table 3-1.

The stand-level variables considered that characterized each plot were: growth zone (Zone), stand basal area (BA, m² ha⁻¹), trees per hectare (N, trees ha⁻¹), quadratic

mean diameter (QD, cm), dominant height (Hd, m), dominant breast height stand age (Ad, years), dominant specie (DS), site index (SI, m), basal area of *Nothofagus* (BAN, m² ha⁻¹), stand density index (SDI, trees ha⁻¹) and relative spacing index (RS). BAN was calculated as the sum of basal area of *N. alpina*, *N. obliqua*, and *N. dombeyi*. Hd and Ad were calculated using the 100 thickest trees per hectare, and DS corresponded to the *Nothofagus* specie with the largest proportion of BAN. SI was estimated by using available dominant height-site models obtained on the same dataset for an index age of 20 years (see appendix A). SDI was calculated using: $SDI = N \times (25.4/QD)^\beta$ (Avery and Burkhart 2002), where $\beta = -1.4112$, as reported by Gezan et al. (2007) fitted using both the PP and TP networks. Finally, RS was calculated according to: $RS = [(10000/N)^{0.5}]/Hd$ (Avery and Burkhart 2002; Contreras et al. 2011). Hence, a total of 1,108 trees constituted the final dataset for fitting containing tree and stand attributes for a total of 158 plots with a single measurement.

The validation or projection database was formed by 20 plots belonging to the PP network also established in 2000, re-measured in 2006, and a subset of seven plots was measured a third time in 2012. For 2006 there were a total of 1,568 trees with measurements, and for 2012 this number dropped to 435 trees. Note that the dominant species for these plots are *N. obliqua* (18 plots) and *N. alpina* (2 plots).

Model Fitting

Multiple linear regression, using several potential predictors, were fitted for estimation of AIDBH. Two selection variable procedures were evaluated: 1) model selection with cross-validation (CV), and 2) LASSO regression. The resulting selected models were later extended using regression with groups to identify which of the

factors, growth zone or specie, contributes to improve the performance of these models, and therefore, explain better the biological process.

AIDBH could be influenced by different elements, such as site productivity, tree-to-tree competition, together with individual-tree and whole-stand attributes (Andreassen and Tomter 2003). The fitted growth model form used in this study was:

$$\ln(\text{AIDBH}) = f(\text{productivity, competition, tree, stand}) \quad (3-1)$$

Where, AIDBH is the annual increment in DBH; productivity was represented by Hd and SI; competition predictors were SS, BAL, BALn, BALr, SDI, and RS; tree attributes corresponded to DBH, A, and H; and whole-stand attributes included DA, BA, QD, and BAN. Hence, these corresponded to a total of $p = 15$ predictors.

The original predictors and transformations for each of these 15 predictors were considered, these were: natural logarithm, inverse, inverse square root and quadratic term. Some of these transformations included a constant $c = 10$, particularly for logarithm transformation. The above transformations were selected as these are the most common one reported for individual-tree growth model (Shortt and Burkhart 1996; Murphy and Graney 1998; Huang and Titus 1999; Andreassen and Tomter 2003; Pukkala et al. 2009; Contreras et al. 2011; Nunes et al. 2011; Pukkala et al. 2011; Ma and Lei 2015).

The model selection procedure with CV regression, was setup by using a K-fold cross-validation scheme that divided the fitting database into $K = 10$ random groups (or folds) of similar size (James et al. 2013). As a first step, a model with all predictors and its transformations ($p = 67$), was evaluated using the CV procedure, subsequently, a single expression of each predictor was chosen by backward selection and by doing so,

a temporary model containing $p = 15$ predictors was available for the next step. Then, a new CV procedure was run with these 15 pre-selected predictors to obtain a final candidate model. The best CV model was selected according to the following goodness-of-fit statistics: adjusted r-squared (R^2_{adj}), Mallows' C_p , Schwartz's information criterion (BIC), residual sum of squares (RSS), and the mean squared error averaged over the K folds (MSE_K) (Lumley and Miller 2009; James et al. 2013).

The second procedure used for model selection and model fit was LASSO regression (James et al. 2013). The λ tuning parameter was determined by cross-validation using the fitting database with the mean square error as criteria to obtain the final candidate model.

Once the potential predictors were identified for the CV and LASSO regression procedures, the fitting database was then split 50/50 into training and test dataset. The training dataset was used to obtain the β model parameters. A model correction factor (Baskerville 1972) was incorporated due to the use of the natural logarithmic of the response AIDBH, with $\hat{\beta}_0^* = \hat{\beta}_0 \exp(MSE/2)$ is the adjusted intercept, and MSE is the mean square error for a given model obtained based on the training dataset. To evaluate the predictions performance of the models the following goodness-of-fit statistics were used on the test data (Huang et al. 2003; Welham et al. 2015): empirical coefficient of correlation (R^2_{emp}), root mean square error (RMSE), relative root mean square error (RMSE%), bias (BIAS) and relative bias (BIAS%). Their formulae are:

$$R^2_{emp} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (3-2)$$

$$RMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n - 1}} \quad (3-3)$$

$$RMSE\% = 100 \times (RMSE / \bar{y}) \quad (3-4)$$

$$\text{BIAS} = \sum(y_i - \hat{y}_i) / n \quad (3-5)$$

$$\text{BIAS}\% = 100 \times \text{BIAS} / \bar{y} \quad (3-6)$$

Where, y_i and \hat{y}_i are the i^{th} observation and prediction value, respectively; n is the number of observations, \bar{y} is the mean of the observed values. All of the summations go from 1 to n . These statistics were obtained for the complete test dataset, and additionally, they were calculated by dividing this dataset into three DBH classes, defined as: 5-15 cm, 15-30 cm, and >30 cm.

In order to generalize the candidate diameter growth model from the CV procedure, species and growth zones were incorporated as both independent factors and as interaction among predictors. The species factor was defined with a label for each of the three *Nothofagus* species. The four growth zones correspond to those defined by Gezan and Moreno (1999) for the same population (see Figure 3-1). Each factor was evaluated separately or as a combined specie-zone factor. The fitting of the models was done with the 50/50 training and test dataset, and the goodness-of-fit statistics described before were also calculated. The significance and selection of each factor, or its interaction with a given predictor, were assessed through an F-test as reported in the analysis of variance, and model terms were eliminated using a backward selection procedure. Later, the final selected factors were further evaluated by obtaining their least squares means and all pairwise comparisons using Tukey adjustment. For all analyses, a significance level of 5% was used.

In the case of the LASSO model candidate, dummy variables were defined for each of the levels of specie-zone. A final model was selected by running LASSO regression with a new tuning parameter determined by CV using the fitting database.

In order to verify the levels of multicollinearity between the selected predictors in the CV and LASSO regression models, variance inflation factor (VIF) and generalized variance inflation factor (GVIF) (Fox and Monette 1992) were calculated for the models with and without factors, respectively. All predictors with VIF or GVIF greater than 10 were removed from these four models, as they were considered highly collinear with other predictors.

Performance of the CV and LASSO regression models, with and without groups, was evaluated using an independent validation with the projection database and model simulations of 6 and 12 years since plot establishment. Here, AIDBH estimates and predictors were updated yearly until projection age. Companion species in this forest type, that lack a fitted model, were projected with the same models used for *Nothofagus*, and BALn values for companion species were assumed to be all equal to the minimum *Nothofagus* BALn value.

Normality of residuals and heterogeneity of variances was assessed and no important departures from these assumptions were noted. Database management and statistical analyses were all performed using the statistical software R v. 3.2.2 (R Core Team 2015). The R packages used for the CV regression was *leaps* (Lumley and Miller 2009), and for LASSO regression *glmnet* (Friedman et al. 2016).

Results

Model Fitting

The rates of annual DBH growth for the different species obtained from the fitting database corresponded to averages of 2.6, 3.0 and 3.6 mm for *N. alpina*, *N. obliqua* and *N. dombeyi*, respectively. These rates vary greatly according to growth zone (see Table

B-2), where, zone 3 had the highest annual DBH growth average with 3.8 mm and the zone 4 had the lower growth with 2.7 mm.

The CV regression procedure of variable selection resulted in five predictors of AIDBH, these were selected according to the goodness-of-fit statistics R^2_{adj} , C_p , BIC, RSS, and MSE_k , that consistently chose five predictors (Figure 3-2). All model parameter estimates were significant at $P < 0.001$, and its estimated parameters are presented in Table 3-2. The resulting expression of the model is:

$$\ln(AIDBH) = \beta_0 + \beta_1 \ln(BALn + 10) + \beta_2 SDI + \beta_3 \ln(DBH) + \beta_4 \ln(A) + \beta_5 SS + e \quad (3-7)$$

Where β_0 to β_5 are the parameters, e is the error term, and the other terms were previously described.

The validation of the prediction of CV selection model resulted in a R^2_{emp} of 0.56 with a RMSE% of 44.16% and a BIAS% of -1.96% (see Table 3-3). Thus, the predictions have moderate quality (large RMSE%), but they have negligible bias (Figure 3-3).

The model selected by LASSO regression, at the beginning had eight predictors; however, after observing the VIF values, the final LASSO model resulted in seven predictors. The tuning parameter λ obtained by cross-validation was 5.5×10^{-4} . In relation to the CV regression model, LASSO incorporated the predictors BALr and DA, and the expression of this model is:

$$\ln(AIDBH) = \beta_0 + \beta_1 \ln(BALn + 10) + \beta_2 SDI + \beta_3 \ln(DBH) + \beta_4 \ln(A) + \beta_5 SS + \beta_6 \ln(DA) + \beta_7 / \sqrt{(BALr + 10)} + e \quad (3-8)$$

Where β_0 to β_7 are parameters to estimate (see Table 3-2). This model showed similar goodness-of-fit statistics as the CV regression, for the predictions based on the validation data, with $R^2_{emp} = 0.57$, $RMSE\% = 44.16$ and $BIAS\% = -2.94$.

To extend the diameter growth model to incorporate the factors of specie and growth zone, the CV model was fitted with these factors and their interactions against all predictors. These factors were added individually (*i.e.*, Zone, SP) and using a combination of both as a unique new factor (SpZone). A backward selection procedure eliminated all of these interactions, resulting in models that only describe different intercepts for each of the levels of these factors. The best model corresponded to the combined factor of SpZone (CV + SpZone, see Table B-3):

$$\begin{aligned} \ln(\text{AIDBH}) = & \alpha_{11} \text{SpZone}_{11} + \alpha_{12} \text{SpZone}_{12} + \alpha_{14} \text{SpZone}_{14} + \alpha_{21} \text{SpZone}_{21} + \alpha_{22} \\ & \text{SpZone}_{22} + \alpha_{23} \text{SpZone}_{23} + \alpha_{24} \text{SpZone}_{24} + \alpha_{31} \text{SpZone}_{31} + \alpha_{32} \text{SpZone}_{32} + \alpha_{33} \\ & \text{SpZone}_{33} + \alpha_{34} \text{SpZone}_{34} + \beta_1 \ln(\text{BALn} + 10) + \beta_2 \ln(\text{DBH}) + \beta_3 \ln(\text{A}) + \beta_4 \text{SS} + e \end{aligned} \quad (3-9)$$

Where α_{ij} is the intercept parameter for the i^{th} specie in the j^{th} zone, and β_1 to β_5 are slope parameters. SpZone_{ij} are the dummy variables associated with the ij^{th} group. The specie were coded as: 1 (*N. alpina*), 2 (*N. obliqua*), and 3 (*N. dombeyi*). The other model terms were described previously. Note that with the incorporations of SpZone_{ij} the predictor SDI left the model as it was not significant ($P = 0.297$).

The goodness-of-fit statistics for this extended model are $R^2_{emp} = 0.56$, $RMSE\% = 44.49\%$, and $BIAS\% = -1.96\%$. This model had similar performance than the CV model (Table 3-3); however, the factor SpZone resulted highly significant ($P < 0.001$).

Equivalently, the LASSO regression model was extended with the factor SpZone (LASSO + SpZone) using a new tuning parameter ($\lambda = 3.9 \times 10^{-3}$). Here, the final model is:

$$\begin{aligned} \ln(\text{AIDBH}) = & \beta_0 + \beta_1 \text{SpZone}_{14} + \beta_2 \text{SpZone}_{21} + \beta_3 \text{SpZone}_{22} + \beta_4 \text{SpZone}_{24} + \beta_5 \\ & \text{SpZone}_{32} + \beta_6 \text{SpZone}_{33} + \beta_7 \ln(\text{BALn} + 10) + \beta_8 \text{SDI} + \beta_9 \ln(\text{DBH}) + \beta_{10} \ln(\text{A}) + \\ & \beta_{11} \text{SS} + \beta_{12} \ln(\text{DA}) + \beta_{13} / \sqrt{(\text{BALr} + 10)} + e \end{aligned} \quad (3-10)$$

Where β_0 to β_{13} are parameters to estimate, and the other terms were described previously. The goodness-of-fit statistics for this model are $R^2_{\text{emp}} = 0.54$, $\text{RMSE}\% = 45\%$ and $\text{BIAS}\% = -4.31\%$, values that are lower in performance with respect to other three previous models evaluated (Table 3-3). Note that standard errors and P-values for LASSO and LASSO + SpZone are not reported because it is not possible to obtain a reliable unbiased estimate of the standard error (Kyung et al. 2010; Lockhart et al 2014). The $\hat{\beta}_0^*$ based on Baskerville (1972) correction of CV, LASSO, CV + SpZone, and LASSO + SpZone models were 2.546, -0.160, 2.829, and -1.239, respectively.

The evaluation of the four predictive models of AIDBH across different DBH classes (5-15, 15-30, >30 cm) provided some interesting differences (Figure 3-4). For $\text{RMSE}\%$, the CV, LASSO, and CV + SpZone models had similar performances with lower values for the intermediate class (15- 30 cm). LASSO + SpZone model had better performance for the other two classes; however, it showed much larger $\text{RMSE}\%$ in the intermediate class. In the case of $\text{BIAS}\%$, the four models resulted in similar performance across all DBH classes. It appears that the best model corresponded to CV + SpZone; however, it presented larger negative $\text{BIAS}\%$ for the class >30 cm, and the LASSO regression models had the smallest bias for the same class.

Model Projection

Model validation using the projection dataset found good results for all models with $R^2_{emp} > 0.97$ (Table 3-4). CV and LASSO regression had better goodness-of-fit statistics, while LASSO + SpZone had the worst performance. A projection of 6 years showed that LASSO regression had a better $R^2_{emp} = 0.99$ and $RMSE\% = 7.26$ but with more $BIAS\% = 0.36$ than the CV model, although all models had excellent statistics. CV and LASSO regression for a simulation of 12 years continued with good prediction performance. CV regression presented $R^2_{emp} = 0.97$, $RMSE\% = 9.66$ and $BIAS\% = 1.75$, at the same projection time, with similar results for LASSO regression with $R^2_{emp} = 0.97$, $RMSE\% = 9.50$ and $BIAS\% = 1.96$. Surprisingly, the extended models with species and growth zone factors were not superior as those without them.

Selected Model

LASSO + SpZone model showed poor performance for fitting and projection, with a negative bias. In contrast, the CV, LASSO, and CV + SpZone models presented similar performance in terms of goodness-of-fit statistics. However, LASSO regression had a larger bias and, in addition, included two more predictors than CV. Hence, CV model could be considered the best model, but the analysis of variance for CV + SpZone showed that the factor SpZone is highly significant with $P < 0.001$. Thus, the selected model for this study correspond to CV + SpZone. Based on this model, the incorporation of species and growth zone confirm the differences reported earlier about DBH growth from the fitting database (see Table B-2 and Chapter 2).

An assessment of the selected model corroborates forest ecological processes and shows biological consistency in the parameter estimates (Table 3-2). For example, BAL_n and SS are competition variables where a higher value indicates a high

competition for a given tree, and their model parameters have both negative sign; hence, with higher competition there are lower rates of growth. In the case of the inverse of BALr, the estimated coefficient value is positive, thus a high competition produces a smaller growth rate, and vice versa. The most important predictor, DBH has a strong effect on growth, where bigger trees have higher diameter increment because due to its positive sign, which is expected as this effect is controlled by the predictor age. Hence, at a fixed age, larger trees tend to have greater growth rates. Finally, the parameter associated with age is negative, thus, for a fixed DBH value, a young tree is capable of more growth than an older tree.

For the CV + SpZone model, the least squares means showed that there are important differences among the levels of this combined factor (Table 3-5). First, these results show similar DBH growth for *N. dombeyi* across all growth zones. Second, *N. obliqua* also has a similar response in all zones with the exception of zone 3, this is probably a result of been the wettest zone with long growing season. Third, *N. alpina* has the largest differences between zones, where growth seems to increase with altitude (e.g., zone 4, Andes mountain ranges, has the largest growth). These results, justify the inclusion of the combined factor SpZone in the final selected model: CV + SpZone. The validation of the projection with independent data for the CV + SpZone model shows high predictive power for estimating future DBH for a period of 6 years; and even if the projection is 12 years results are very consistent (Figure 3-5).

Discussion

Diameter growth models have been recognized as a crucial component for projecting individual-tree growth (Vanclay 1994; Avery and Burkhart 2002; Burkhart and

Tome 2012). In this study, a DBH growth distance-independent model was development for mixed *Nothofagus* second growth forest in southern Chile.

Individual-tree models present good results for short-term projections, with a similar level of accuracy respect to whole-stand models (Buckman 1962; Burkhart 2003). However, these models often have poorer levels of accuracy for long-term projections, where whole-stand models are more reliable. Individual-tree models are more useful to project changes in the forest structure, since the diameter distribution is directly obtained from the input data. Estimation of timber products, stand structure, differentiation by specie and/or strata, and incorporations of silvicultural methods are also improved with individual-tree models (Andreassen and Tomter 2003). In relation to the distance-dependent models, the distance-independent models are simpler to implement given that they do not need the physical coordinates of each tree within a plot; thus, they can be used with the information obtained from a typical forest inventory. Salas et al. (2016) mentioned that both procedures do not have significate differences in growth prediction, even in natural forests; furthermore, spatial dependency is more complex to model. Therefore, it is reasonable to aim at developing a distance-independent individual-tree growth model in this study for *Nothofagus* forests.

The response variable for this study was the annual average DBH growth obtained over a period of two years that originated from increment cores and tree sections from felled trees. Several crucial elements are associated with the use of this response variable. First, a short time period of two years was chosen in order to have better association of observed tree growth with its present tree and stand conditions, a relationship that get weaker as the periods get longer. This is especially present, on

those years with particular conditions (*e.g.*, climatic, diseases, catastrophic event); however, Andreassen and Tomter (2003) found no effects of the prediction accuracy with different growth period lengths. Second, no evidence of wood compression was found in increment cores when compared to sections (data not shown). Finally, the available individual basal area growth was discarded as response as it presented poorer goodness-of-fit statistics than DBH growth.

Individual-tree growth models that include predictors of competition, productivity, tree- and stand-level have been used in many reported studies (Harrison et al. 1986; Murphy and Graney 1998; Huang and Titus 1999; Peng 2000; Andreassen and Tomter 2003; Pukkala et al. 2009; Nunes et al. 2011; Ma and Lei 2015). Hence, there are potentially many predictors that might describe the same process, which will suffer from multicollinearity producing inconsistencies and unexpected results in the fitted models (Welham et al. 2015). Thus, it is necessary to select a subset of these predictors using some variable selection method. In this study, the fitting database contained, as expected, pairs of predictors with high correlation, such as BA-BAN, DH-RS, BA-SDI, BAL-BAL_n, DH-QD, A-DA, RS-QD, and DBH-H, with values >0.83 (Figure C-1), and the variable selection methods implemented here were CV and LASSO regression. The predictors selected by CV and LASSO regression corresponded to competition, tree- and stand-level variables, but they did not include variables associated with productivity. In CV regression DBH, A, SS, SDI, and BAL_n were the selected predictors, while in LASSO regression these were the same predictors with the addition of DA and BAL_r. The tree-level variables selected corresponded to DBH and A, both correspond to a measure of tree size; where for DBH, growth rates increase with larger diameters;

however, age presented a negative parameter, indicating that at a fixed DBH value, older trees present lower growth rates. This agrees with a similar result reported by Cubillos (1987) in *N. alpina*. In this study, competition variables had a high relevance in explaining individual tree growth, as found in other *Nothofagus* studies (Cubillos 1987; Monserud and Sterba 1996; Coomes and Allen 2007; Richardson et al. 2011; Ivancih et al. 2014). Interestingly, the predictor BALn (basal area of larger trees only for *Nothofagus*), was selected instead of BAL (that includes all species), agrees with findings from previous studies that support the theory of additive effects in growth for *Nothofagus* and companion species (Lusk and Ortega 2003; Donoso and Lusk 2007; Donoso and Soto 2016), that translates into independent behavior (and therefore models) for each of these cohorts. In contrast, productivity predictors, such as SI, were not selected for modelling DBH growth in these forests.

The average DBH growth rates for combinations of specie and zone varied (Table B-2), with important differences between species and zone; however, extreme cases were found for the combinations of *N. alpina* - zone 1, and *N. dombeyi* - zone 3, with an 88% larger growth rate for the latter one (*i.e.*, 2.15 mm against 4.05 mm). Other studies also have found differences in growth between *Nothofagus* species and zones (Donoso et al. 1993; Schlatter et al. 1994 and 1995; Donoso et al. 2004; Echeverria and Lara 2004; Gezan et al. 2009; Esse et al. 2013).

The incorporation of species and/or zone factors to growth models has been previously reported (Andreassen and Tomter 2003; Pukkala 2009). In the present study, the combined specie-zone factor (SpZone with 11 levels), explained more variability than the individual factors (with 7 levels for both specie and zone), indicating that a

greater disaggregation of the data is required to model growth rate. This agrees with the average growth differences reported earlier (Table B-2). Interestingly, the LASSO + SpZone model selected as predictors for the combined factor SpZone the extreme groups (Table 3-5), by combining the central five groups into a single group.

The fitting of the four candidate models to predict AIDBH resulted in good overall results, with R^2_{emp} ranging from 0.54 to 0.57, RMSE% of ~44%, and a BIAS% smaller than 3%. Similar goodness-of-fit statistics have been reported for AIDBH model with R^2_{emp} ranging from 0.26 to 0.68 (Wykoff 1990; Monserud and Sterba 1996; Andreassen and Tomter 2003; Nunes et al. 2011). AIDBH projections for 6 and 12 years resulted in R^2_{emp} ranges of 0.23-0.28 and 0.15-0.24, respectively. This drop in the goodness-of-fit statistics is been affected by the poor quality of the projection database, that included, for example, at 6 years 20% of the trees with null or negative DBH increments. This is likely to be associated with field measurement errors and low growth rates for these species.

In contrast, DBH projections for 6 and 12 years result in overall R^2_{emp} values greater than 0.97 for both *Nothofagus* and companion species. Small diameter trees (DBH < 15 cm) presented lower correlations, with values of 0.88 and 0.58 for projections of 6 and 12 years, respectively. As expected, goodness-of-fit statistics were better for the *Nothofagus* cohort than the companion species cohort, because the models were built using only measurements from *Nothofagus*.

The final selected model in this study is CV + SpZone (Table 3-3). The LASSO model was discarded, even that was highly competitive, due to lower prediction performance, and it also included some predictors with high multicollinearity (VIF > 5).

LASSO + SpZone had the worst performance with issues for prediction and projection. The CV model presented better goodness-of-fit statistics than the CV + SpZone model; however, the combined factor SpZone was highly significant ($P < 0.001$). Hence, the inclusion of SpZone is justified given the wide geographical range of this population and its differential diameter growth (see Table B-2). These results has led to many authors to define growth zones (Donoso et al. 1993; Gezan and Moreno 1999; Echeverria and Lara 2004). Alternatively, it is possible to use the CV model that does not require zone or specie to have reliable prediction of DBH growth.

The advantages of CV as a predictor selection procedures is the possibility of providing for a direct estimate of test error (MSE_K) compare to stepwise methods. LASSO is an alternative procedure that regularizes the coefficient estimates, forcing some of them to be exactly equal to zero, performing variable selection. Both procedures have been used commonly due their advantages to identify those variables that better predict a desired response (James et al. 2013). Similarly, the importance of incorporate dummy variables (or factors) in a regression is related to the model specificity, where the factors can correct intercepts or slope from the original regression.

Given the high levels of variability present in these mixed natural forests (Chapter 2), the above results are promising due to high complexity of these ecosystems, including several strata, species, anthropic disturbances, and stage structure (Donoso 1993). However, future research is warranted to better improve the performance of models of individual tree growth for this population. Some of these studies could include: 1) models that do not require age, as this predictor is often difficult to obtain from typical forest inventories; 2) increase the additional validation data from a wider

geographic range and with additional forest conditions; 3) develop an individual-tree growth model for companion species; 4) fully implement an individual-tree model routine that includes a mortality and a height growth module; and 5) develop a compatible system based on both individual-tree and whole-stand models.

The results from this study, on the evaluation of individual-tree DBH growth model indicated that CV + SpZone model is useful for practical applications; although, this model should be used within its frame of inference: RORACO second growth forest, between 37° and 41.5° south latitude, with breast height ages ranging from 20 to 80 years, and with a *Nothofagus* basal area greater than 60%. It is recommended the use of this proposed model for projections as long as 12 years, where this individual-tree DBH growth model showed to be a robust tool with reasonable predictions.

A comparison of two selection predictor procedures to estimate annual increment in DBH was presented using data from a second growth mixed forest conformed by *N. alpina*, *N. obliqua*, and *N. dombeyi*, in southern Chile. A selected small set of predictors successfully explained a large portion of DBH growth, with variables associate competition, tree- and stand-level components; however, no variables associated with productivity were included. The final model constitutes a simple and valuable tool to support management decision for this ecosystems. This study, is the first complete effort to obtain an individual-tree growth model for DBH useful for the full geographical distribution of the RORACO forest type.

Table 3-1. Summary statistics of individual and stand attributes for fitting database based on PP and TP networks.

Tree Specie	<i>N. alpina</i>					<i>N. obliqua</i>					<i>N. dombeyi</i>				
	n	mean	SD	min	max	n	mean	SD	min	max	n	mean	SD	min	max
Tree variables															
DBH	202	16.7	9.8	5.0	42.7	627	17.0	10.0	5.0	62.1	279	19.3	11.3	5.0	60.2
H	202	15.2	6.6	4.2	34.5	627	16.1	7.3	3.5	45.0	279	16.2	6.1	4.2	34.5
A	202	36.7	14.2	11.0	104.0	627	30.6	14.6	8.0	95.0	279	32.5	13.2	9.0	80.0
BAL	202	26.0	16.8	0.0	83.4	627	19.0	14.0	0.0	65.1	279	28.0	20.6	0.0	92.4
BALn	202	20.9	15.0	0.0	82.3	627	15.7	12.3	0.0	54.9	279	20.6	16.6	0.0	82.2
SS	202	2.3	1.1	1.0	4.0	627	2.3	1.0	1.0	4.0	279	2.2	1.0	1.0	4.0
BALr	202	0.6	0.3	0.0	1.0	627	0.6	0.3	0.0	1.0	279	0.6	0.3	0.0	1.0
AIDBH	202	2.6	1.6	0.2	7.7	627	3.0	2.1	0.2	12.1	279	3.6	2.2	0.1	10.2
Dominant Specie															
Stand variables	n	mean	SD	min	max	n	mean	SD	min	max	n	mean	SD	min	max
BA	24	46.7	13.8	10.8	72.3	87	35.7	13.0	9.5	71.5	47	55.2	17.8	23.4	98.4
N	24	2,564	1,184	320	5,000	87	2,378	1,119	320	4,640	47	2,900	1,357	880	5,600
QD	24	16.4	4.8	8.5	30.5	87	15.4	6.2	6.8	33.3	47	16.9	5.4	8.4	30.4
Hd	24	21.2	5.7	10.2	35.1	87	20.8	7.0	7.8	42.4	47	20.8	5.9	9.9	34.1
Ad	24	47.5	10.9	23.0	77.0	87	36.6	14.7	12.7	86.8	47	40.2	13.1	21.3	85.1
SI	24	10.3	3.9	4.1	24.3	87	12.5	3.9	2.0	22.9	47	11.4	3.4	3.9	18.4
BAN	24	37.6	12.4	10.8	59.7	87	30.4	11.5	8.8	57.8	47	46.5	18.7	8.0	89.6
SDI	24	1,212	336	406	1,944	87	956	258	410	1,674	47	1,400	361	640	2,305
RS	24	0.1	0.0	0.1	0.2	87	0.1	0.0	0.1	0.3	47	0.1	0.0	0.1	0.2

Note: n, number of observations; SD, standard deviation; min, minimum; max, maximum; DBH, diameter breast height (cm); H, total height (m); A, breast height age (years); BAL, basal area of larger trees ($m^2 ha^{-1}$); BALn, basal area of larger trees for *Nothofagus* ($m^2 ha^{-1}$); SS, sociological status; BALr, relative BAL; AIDBH, annual increment in DBH ($mm year^{-1}$); BA, stand basal area ($m^2 ha^{-1}$); N, number of trees ($trees ha^{-1}$); QD, quadratic diameter (cm); Hd, dominant height (m); Ad, dominant breast height stand age (years); SI, site index (m); BAN, basal area of *Nothofagus* ($m^2 ha^{-1}$); SDI, stand density index ($trees ha^{-1}$); RS, relative spacing index.

Table 3-2. Parameter estimates of the four selected models for estimation of the logarithm of AIDBH (mm).

Parameter	Estimate	SE	P-value	VIF*	Parameter	Estimate	VIF*
CV (eq. 3-7)					LASSO (eq. 3-8)		
β_0	2.410×10^0	2.173×10^{-1}	< 0.001	-	β_0	-3.098×10^{-1}	-
β_1	-7.062×10^{-3}	2.064×10^{-3}	< 0.001	1.74	β_1	-7.255×10^{-3}	3.2
β_2	2.745×10^{-4}	6.787×10^{-5}	< 0.001	1.22	β_2	2.215×10^{-4}	1.4
β_3	9.046×10^{-1}	6.798×10^{-2}	< 0.001	3.29	β_3	7.892×10^{-1}	5.28
β_4	-1.138×10^0	7.488×10^{-2}	< 0.001	2.26	β_4	-1.073×10^0	6.73
β_5	-1.336×10^{-1}	3.149×10^{-2}	< 0.001	2.22	β_5	-1.325×10^{-1}	2.44
β_6	-	-	-	-	β_6	-4.318×10^{-2}	5.36
β_7	-	-	-	-	β_7	9.792×10^0	6.59
CV + SpZone (eq. 3-9)					LASSO + SpZone (eq. 3-10)		
α_{11}	2.702×10^0	2.562×10^{-1}	< 0.001	2.09	β_0	-1.387×10^0	-
α_{12}	2.908×10^0	2.345×10^{-1}	< 0.001	2.09	β_1	6.790×10^{-2}	2.28
α_{14}	3.065×10^0	2.397×10^{-1}	< 0.001	2.09	β_2	-2.507×10^{-1}	2.28
α_{21}	2.538×10^0	2.120×10^{-1}	< 0.001	2.09	β_3	-1.702×10^{-1}	2.28
α_{22}	2.587×10^0	2.219×10^{-1}	< 0.001	2.09	β_4	-1.955×10^{-1}	2.28
α_{23}	2.841×10^0	2.272×10^{-1}	< 0.001	2.09	β_5	3.455×10^{-2}	2.28
α_{24}	2.678×10^0	2.329×10^{-1}	< 0.001	2.09	β_6	5.922×10^{-2}	2.28
α_{31}	2.946×10^0	2.298×10^{-1}	< 0.001	2.09	β_7	-7.117×10^{-3}	3.30
α_{32}	2.948×10^0	2.680×10^{-1}	< 0.001	2.09	β_8	8.700×10^{-5}	2.07
α_{33}	2.941×10^0	2.353×10^{-1}	< 0.001	2.09	β_9	7.699×10^{-1}	6.21
α_{34}	2.902×10^0	2.358×10^{-1}	< 0.001	2.09	β_{10}	-1.096×10^0	7.30
β_1	-6.517×10^{-3}	2.002×10^{-3}	0.0012	1.82	β_{11}	-1.293×10^{-1}	2.53
β_2	9.307×10^{-1}	6.970×10^{-2}	< 0.001	3.82	β_{12}	-2.250×10^{-2}	5.56
β_3	-1.175×10^0	7.797×10^{-2}	< 0.001	2.61	β_{13}	1.419×10^{-1}	6.91
β_4	-1.401×10^{-1}	3.092×10^{-2}	< 0.001	2.29	β_{14}	-	-

Note: SE, Standard error; VIF*, variance inflation factor or generalized variance inflation factor.

Table 3-3. Goodness of fit statistics of four models for ln(AIDBH) using test data.

Model	N	R ² emp	RMSE	RMSE%	BIAS	BIAS%
CV	551	0.56	1.35	44.16	-0.06	-1.96
LASSO	551	0.57	1.35	44.16	-0.09	-2.94
CV + SpZone	551	0.56	1.36	44.49	-0.06	-1.96
LASSO + SpZone	551	0.54	1.36	45.13	-0.13	-4.31

Note: SpZone, combination of Specie and Zone; R²emp, empirical coefficient of correlation; RMSE, root mean square error; RMSE%, relative root mean square error; BIAS%, relative bias.

Table 3-4. Projection goodness-of-fit statistics for DBH (cm) in a period of 6 and 12 years using projection database.

	Projection = 6 years							
	CV				CV + SpZone			
	n	R ² emp	RMSE%	BIAS%	n	R ² emp	RMSE%	BIAS%
Total	1455	0.98	7.32	0.18	1455	0.98	7.44	0.54
DBH(5-15)	777	0.88	10.06	-0.42	777	0.89	9.75	0.00
DBH(15-30)	510	0.88	6.94	0.94	510	0.87	7.27	1.46
DBH(>30)	168	0.97	4.11	0.46	168	0.97	4.19	0.32
<i>Nothofagus</i>	943	0.99	6.33	-0.26	943	0.99	6.48	0.31
Companion	512	0.96	10.47	1.46	512	0.96	10.47	1.20
	Projection = 12 years							
	CV				CV + SpZone			
	n	R ² emp	RMSE%	BIAS%	n	R ² emp	RMSE%	BIAS%
Total	389	0.97	9.66	1.75	389	0.97	9.82	2.34
DBH(5-15)	177	0.58	17.07	5.03	177	0.59	16.87	6.02
DBH(15-30)	151	0.83	8.39	2.10	151	0.81	8.86	2.70
DBH(>30)	61	0.89	5.55	1.36	61	0.89	5.60	1.00
<i>Nothofagus</i>	295	0.98	7.93	-0.29	295	0.97	8.17	0.53
Companion	94	0.83	18.45	12.49	94	0.83	18.20	11.92
	Projection = 6 years							
	LASSO				LASSO + SpZone			
	n	R ² emp	RMSE%	BIAS%	n	R ² emp	RMSE%	BIAS%
Total	1455	0.99	7.26	0.36	1455	0.98	7.32	0.60
DBH(5-15)	777	0.88	9.96	-0.52	777	0.89	9.85	-0.31
DBH(15-30)	510	0.88	6.94	1.13	510	0.88	7.12	1.51
DBH(>30)	168	0.97	3.96	0.13	168	0.97	4.02	-0.03
<i>Nothofagus</i>	943	0.99	6.28	-0.05	943	0.99	6.38	0.36
Companion	512	0.96	10.29	1.54	512	0.96	10.38	1.29
	Projection = 12 years							
	LASSO				LASSO + SpZone			
	n	R ² emp	RMSE%	BIAS%	n	R ² emp	RMSE%	BIAS%
Total	389	0.97	9.50	1.96	389	0.97	9.50	2.39
DBH(5-15)	177	0.58	17.07	4.74	177	0.59	16.87	5.13
DBH(15-30)	151	0.83	8.30	2.42	151	0.82	8.49	2.80
DBH(>30)	61	0.90	5.30	0.79	61	0.90	5.27	0.37
<i>Nothofagus</i>	295	0.98	7.74	0.05	295	0.98	7.83	0.62
Companion	94	0.83	18.45	12.41	94	0.83	18.04	11.76

Note: CV, cross validation procedure; SpZone, combination of Specie and Zone; DBH, diameter breast height (cm); R²emp, empirical coefficient of correlation; RMSE%, relative root mean square error; BIAS%, relative bias.

Table 3-5. Least square means (standard errors in parentheses) statistics of all pairwise Tukey comparisons between all 11 groups based for Specie-Zone in CV + SpZone model.

	<i>N. alpina</i>	<i>N. obliqua</i>	<i>N. dombeyi</i>
zone 1	0.794 (0.125) ^{abcd}	0.652 (0.057) ^a	1.020 (0.079) ^{cd}
zone 2	0.996 (0.072) ^{cd}	0.697 (0.053) ^{ab}	1.027 (0.154) ^{abcd}
zone 3	-	0.948 (0.063) ^{bcd}	1.034 (0.090) ^{cd}
zone 4	1.153 (0.086) ^d	0.784 (0.069) ^{abc}	0.960 (0.087) ^{abcd}

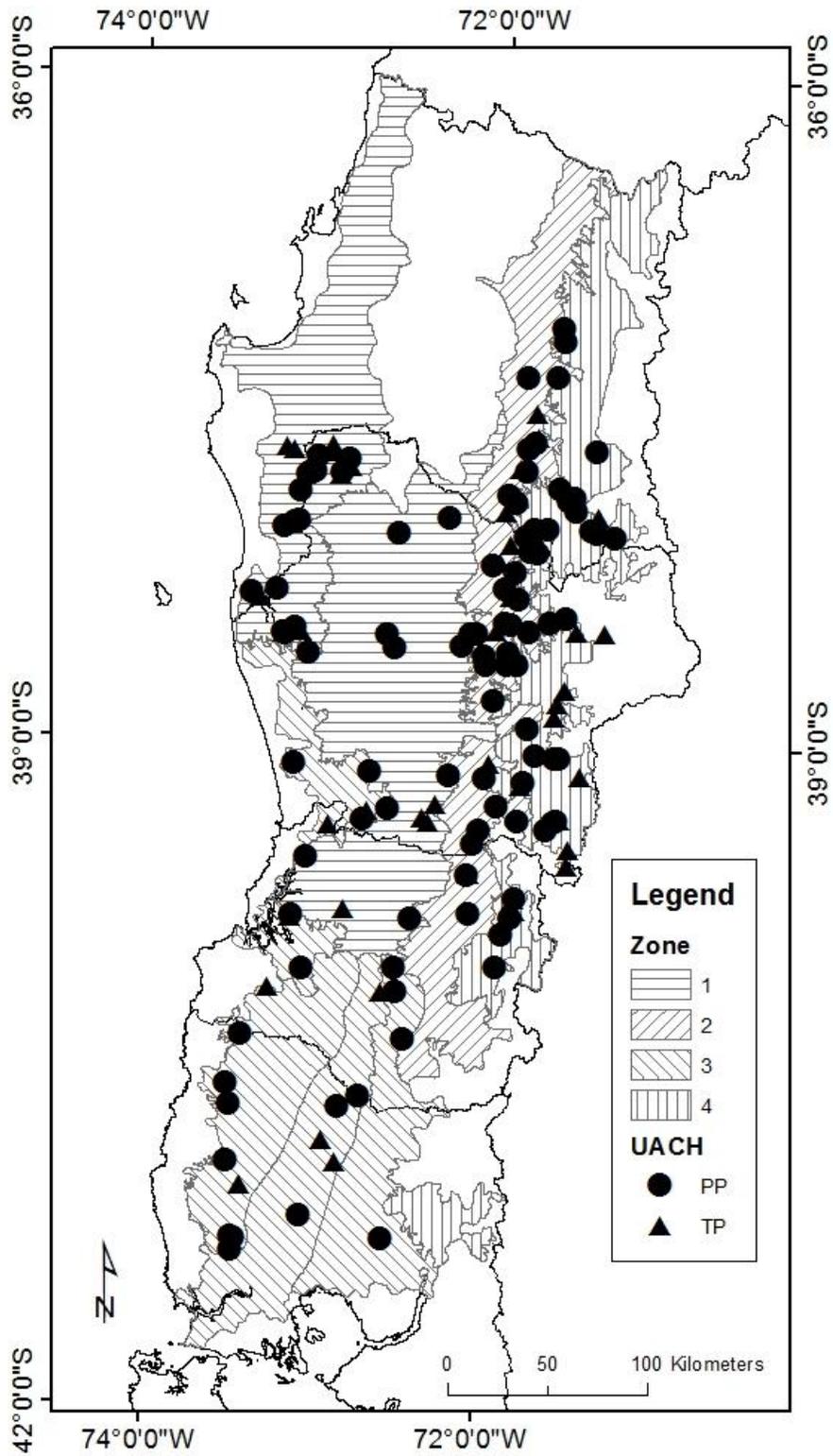


Figure 3-1. Plots distribution in the study area (southern Chile) for the permanent plots (PP) and temporary plots (TP) networks.

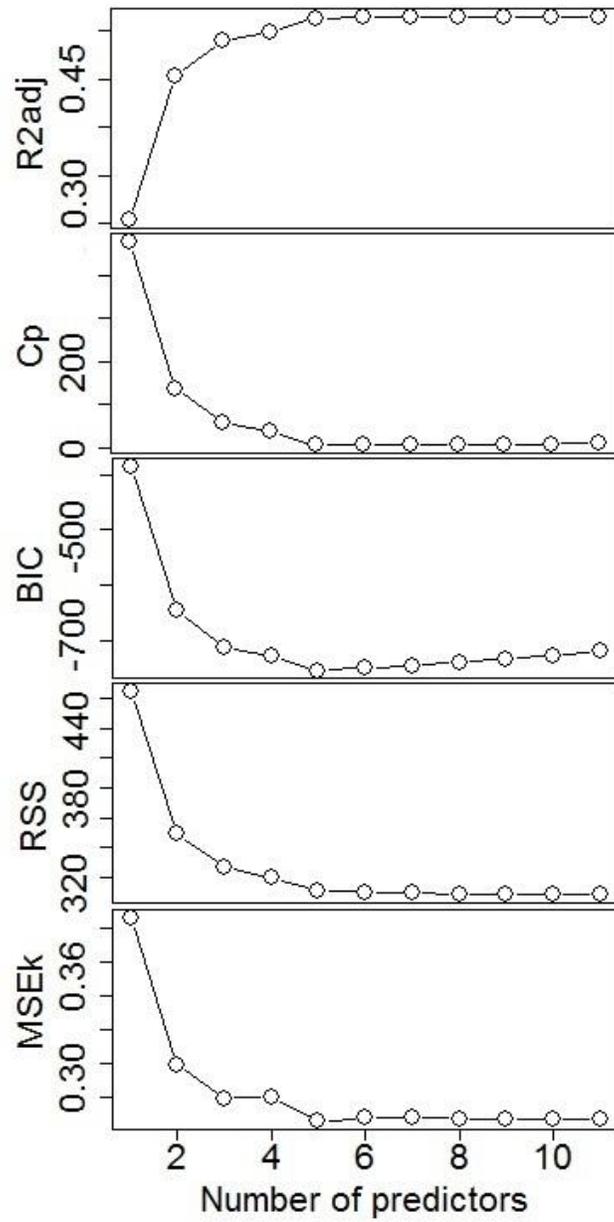


Figure 3-2. Statistics C_p , BIC, RSS, R^2_{adj} and MSE_k for different number of predictors with CV regression selection model using fitting database.

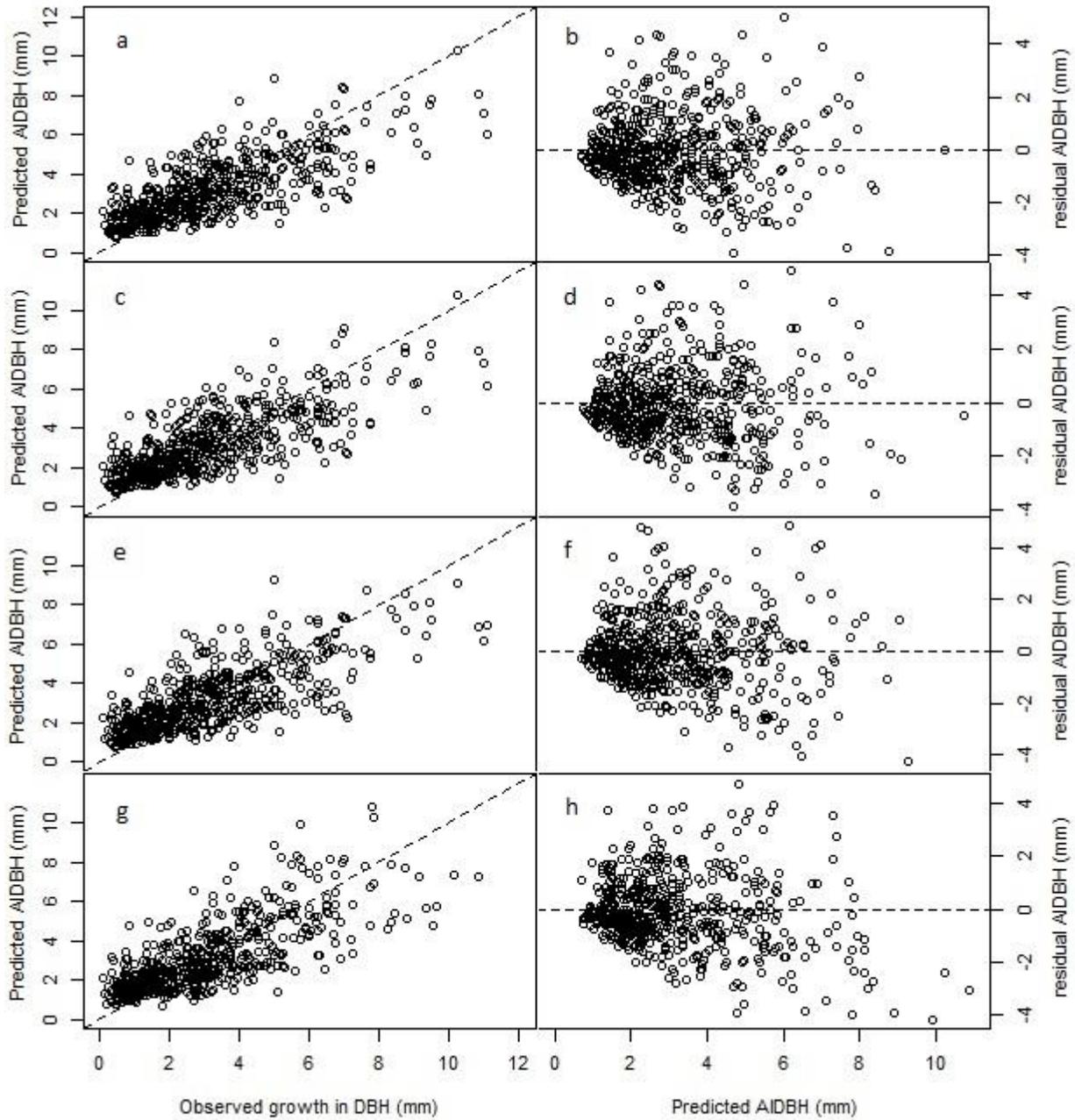


Figure 3-3. Fit and residuals plots of four selected models using test dataset. CV (a and b), LASSO (c and d), CV + SpZone (e and f), and LASSO + SpZone (g and h).

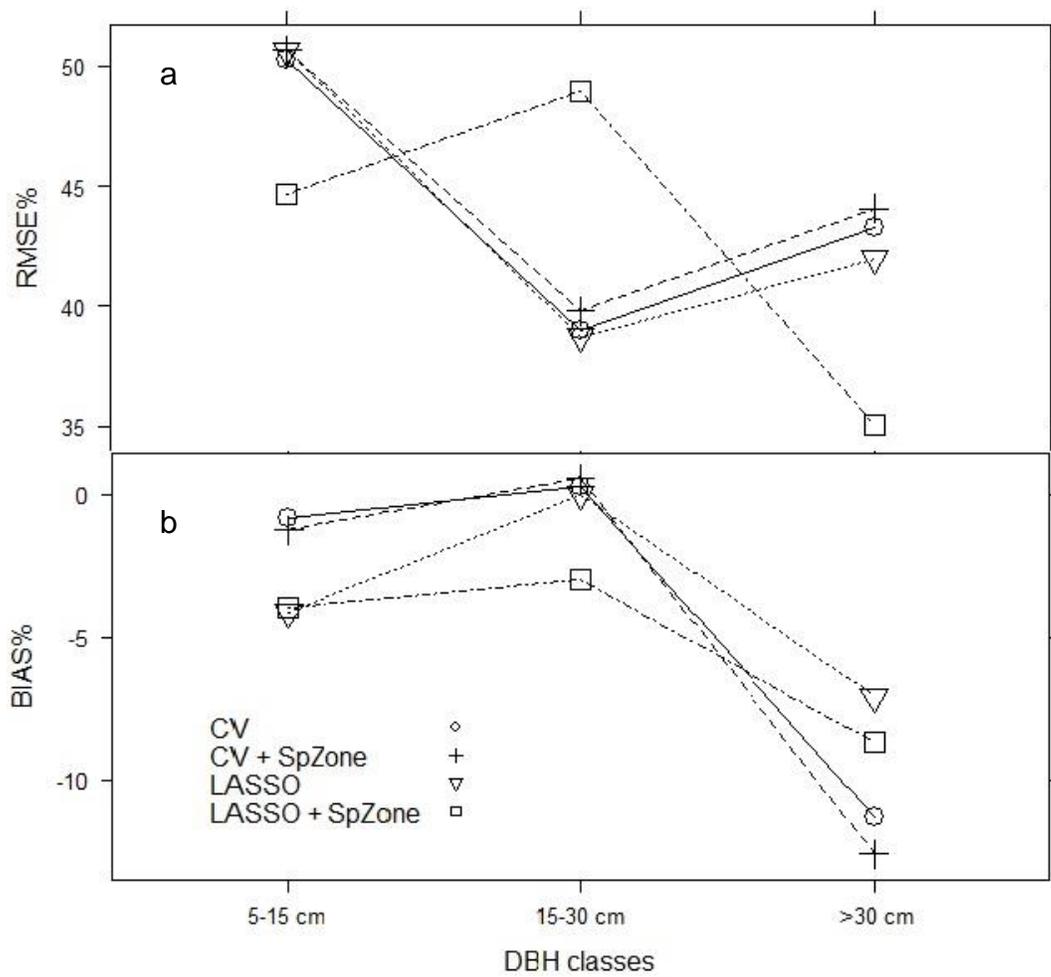


Figure 3-4. RMSE% (a) and BIAS% (b) by DBH class for the four selected models evaluated using the test dataset.

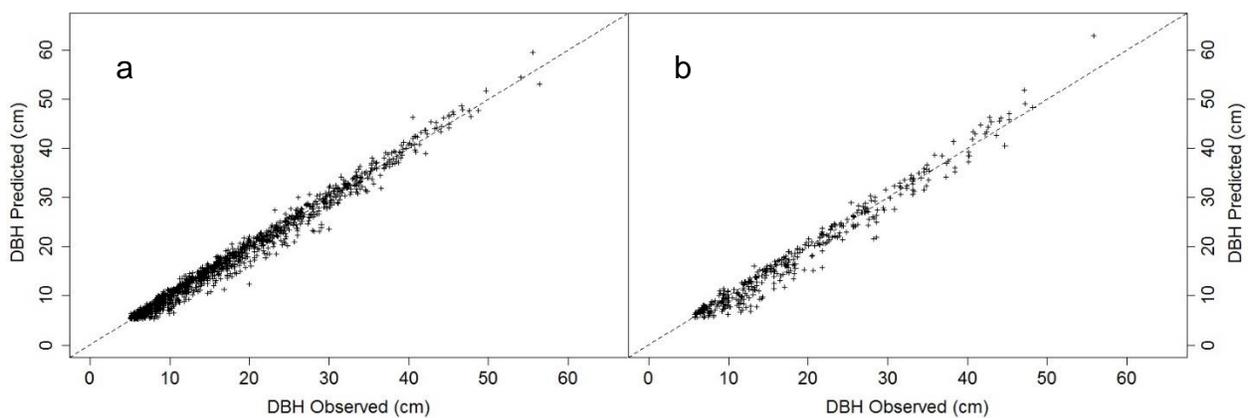


Figure 3-5. Simulation prediction in future DBH (cm) for projection periods of 6 years (a) and 12 years (b) based on the CV + SpZone model using projection database.

CHAPTER 4 CONCLUSIONS

For the second growth mixed RORACO forest type, one of the most important natural forest resources in Chile, currently forest managers and owners do not have enough information and tools to manage this resource ensuring sustainability. This study used data from a total of 128 permanent plots and 60 temporary plots established under a stratification that covers the full geographical range of this forest type in Chile.

The variability found from the unsupervised multivariate methods identified important variables at the tree and stand level for the RORACO forest. The PCA analysis presented a high percentage of variance explained with components related to stand development stages, tree-to-tree competition, and stocking and stand competition. Also, site productivity variables did not explain much of the variability in this dataset.

The other ordination methods used, NMDS and PCoA, provided evidence of important differentiation among species, but not between growth zones. The results with stand level variables, showed that in the multivariate space; *N. alpina* was intermediate between the others two species, being closer to *N. dombeyi* than *N. obliqua*.

In addition, two clusters were enough to separate the sampled population. The main differences among plots were associated to stand development stages, where, there are young forests with a high tree density, and where the other, older forests with basal area concentrated in few individuals. The results from the K-means cluster analysis also identified that growth zones had a relevant importance at the macro region level, particularly zones 1 and 2 that are statistically different.

Supervised methods were successful to model and estimate annual growth in DBH for the combined data, where the selected predictors corresponded to the variables obtained from traditional forest inventories. Both of the selection predictor procedures evaluated, cross-validation (CV) and LASSO regression, showed similar goodness-of-fit statistics with almost identical variables selected. In both cases, a small set of predictors successfully explained a large portion of DBH growth with selected variables associated to competition, tree- and stand-level components; however, no predictors were selected associated with productivity. The latter agrees with the unsupervised methods findings, where productivity variables had a low contribution to explain the variability of these forests. In contrast, as found with unsupervised methods, tree-to-tree competition variables had a high contribution to explain variability and individual tree growth.

The selection of specific predictors associated only with *Nothofagus* (such as basal area of larger trees for *Nothofagus*, BAL_n), instead of those generic predictors that include all species (such as basal area of larger trees, BAL), agrees with others studies that support the theory of additive effects, where *Nothofagus* growth is independent of the companion species growth. Similarly, the incorporation of species and growth zones as factors into the models to estimate annual growth in DBH resulted in an increase of the specificity of these regressions. This is reasonable given the wide geographical range of this sampled population and its differential diameter growth by specie. Also, the final selected model corresponded to a CV with a factor of specie and zone combined, where all estimated parameters showed reasonable biological

interpretation. This model showed moderate goodness-of-fit statistics for both, the prediction and validation datasets.

Further developments are required to provide with a complete set of tools to manage this resource properly. Some of these include: 1) increasing the additional validation data from a wider geographic range and with additional forest conditions; 2) developing models that do not require age, as this predictor is often difficult to obtain from typical forest inventories; 3) constructing an individual-tree growth model for companion species; 4) implementing a full individual-tree model routine that includes mortality and height growth modules; and 5) creating a compatible system based on both individual-tree and whole-stand growth models.

In summary, the final model obtained from this study, constitutes a simple and valuable tool to support management decision for these mixed natural forests that present high levels of variability. This is the first complete effort to develop an individual-tree growth model for DBH that is useful for the full geographical distribution of the RORACO forest type in Chile.

APPENDIX A
DOMINANT HEIGHT-SITE MODEL

The nonlinear models of Chapman-Richards are commonly used in others *Nothofagus* studies (Moreno 2001; Salas and Garcia 2006; Ugalde 2006). The dominant height-site model¹ for this study come from the data originated in the project reported by Ortega and Gezan (1998). The model presented below was fitted with the same database, target species, and growth zones of this study (see Table A-1). The final selected model has the following general equation:

$$Hd = 0.3+a[1-\{1-(SI/a)^{(\beta_0+\beta_1 SI)}\}^{(Ad+0.5)/18}]^{1/(\beta_0+\beta_1 SI)} \quad (A-1)$$

Where, Hd = dominant height (m); Ad = dominant age (years); SI = site index to 20 years (m); and a, β_0 , β_1 are the estimated parameters presented in Table A-1.

Table A-1. Parameter estimates of the dominant height-site model by specie and growth zone.

Specie	Zone	a	β_0	β_1
<i>N. alpina</i>	1	40.108798	0.8756364	-0.0025575
	2	42.716962	0.6814586	0.0235444
	4	35.900857	0.8425262	-0.0034789
<i>N. obliqua</i>	1	50.781108	0.9007745	-0.0055009
	2	36.532756	0.7602626	-0.0036924
	3	52.971650	0.7186384	0.0109518
	4	60.381728	0.8868993	-0.0103325
<i>N. dombeyi</i>	4	42.829543	0.6484692	0.0173443
	all	45.342151	0.6340160	0.0239507

¹ The characteristics of the model used, the source of information and procedures are described by Gezan and Moreno (2000). Curvas de Sitio – Altura Dominante para Renovales de Roble, Raulí y Coigue. Documento Interno Proyecto FONDEF D9711065. Universidad Austral de Chile.

APPENDIX B
SUPPLEMENTARY TABLES

Table B-1. P-values with Bonferroni correction obtained from a post-hoc test of all six possible pairwise comparisons for growth zones related to two clusters.

	Zone 1	Zone 2	Zone 3
Zone 2	0.049	-	-
Zone 3	0.631	0.999	-
Zone 4	0.430	0.999	0.999

Table B-2. Mean diameter breast height growth rates (standard deviation in parentheses) by specie and growth zone obtained from the available data (fitting database).

	<i>N. alpina</i>	<i>N. obliqua</i>	<i>N. dombeyi</i>	Total
Zone 1	2.15 (1.5)	2.92 (2.0)	3.51 (2.3)	3.01 (2.1)
Zone 2	2.66 (1.6)	3.06 (2.3)	3.07 (2.5)	2.94 (2.1)
Zone 3	-	3.69 (2.2)	4.05 (2.3)	3.82 (2.2)
Zone 4	2.71 (1.7)	2.34 (1.5)	3.34 (1.9)	2.75 (1.7)
Total	2.58 (1.6)	3.00 (2.1)	3.56 (2.2)	3.07 (2.1)

Table B-3. Goodness-of-fit statistics for factors specie, zone and both in the CV regression model using the test data.

	n	R ² emp	RMSE	RMSE%	BIAS	BIAS%
Zone	551	0.55	1.37	44.82	-0.07	-2.29
Sp	551	0.56	1.36	44.49	-0.07	-2.29
Zone + Sp	551	0.55	1.38	45.14	-0.07	-2.29
SpZone	551	0.56	1.36	44.49	-0.06	-1.96

Note: Sp, specie; SpZone, combination of Specie and Zone; R²emp, empirical coefficient of correlation; RMSE, root mean square error; RMSE%, relative root mean square error; BIAS%, relative bias.

APPENDIX C SUPPLEMENTARY FIGURES

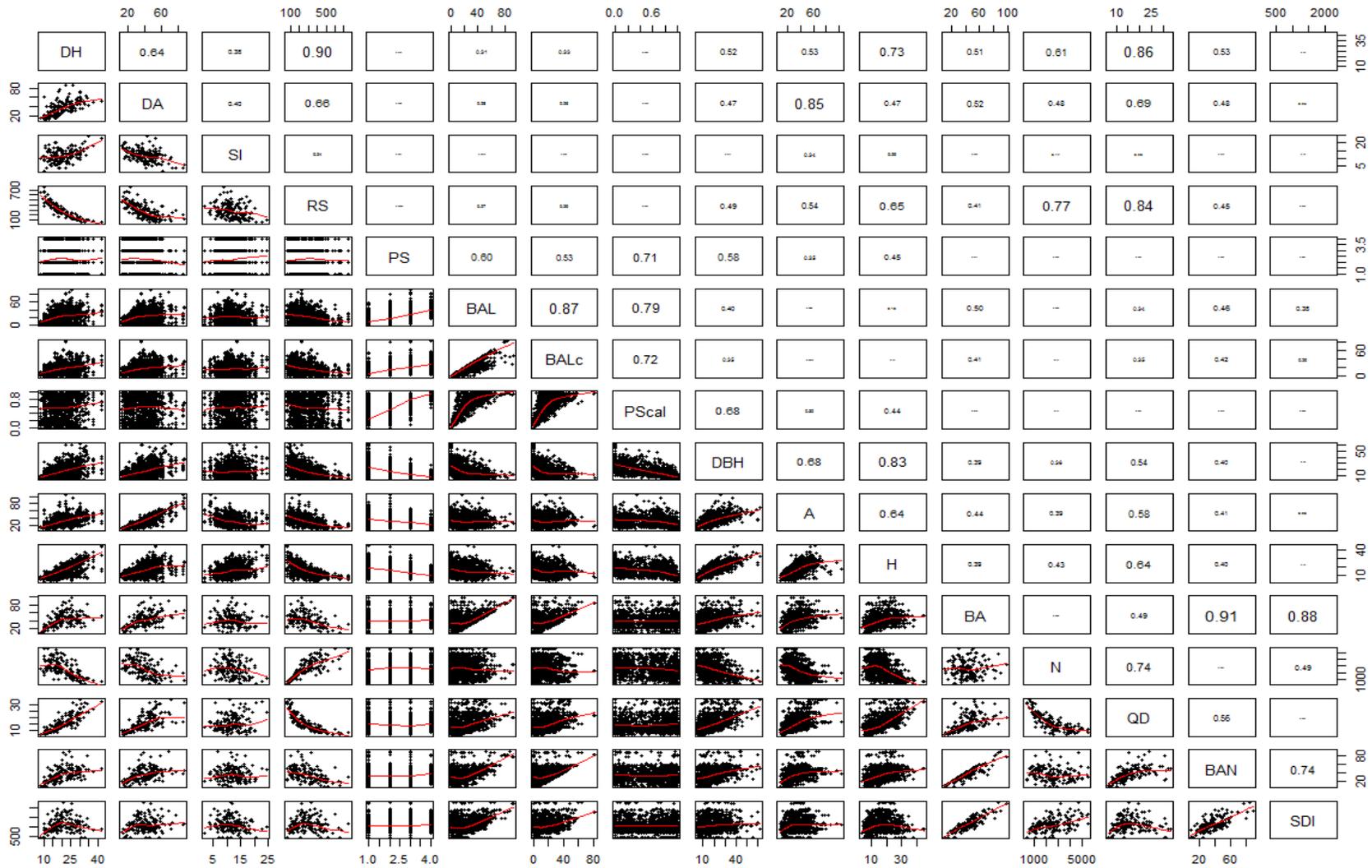


Figure C-1. Scatterplot correlation matrix of predictors (in original units) for the fitting database.

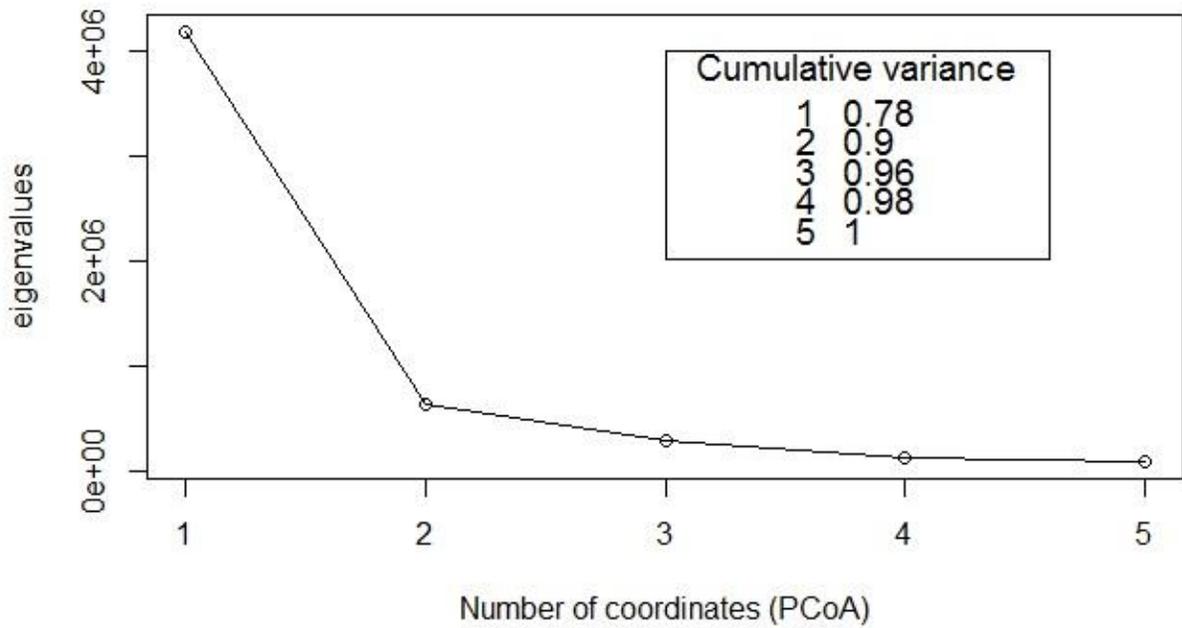


Figure C-2. Scree plot of PCoA with the cumulative variance of each coordinate based on the continuous variables of the tree and stand subsets.

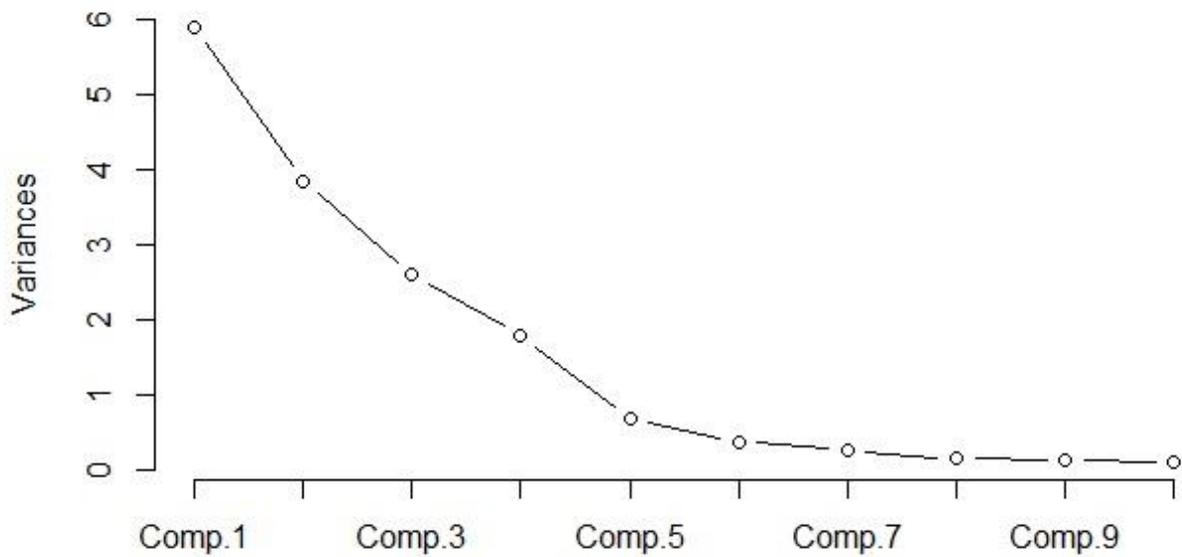


Figure C-3. Scree plot of PCA for continuous variables of tree and stand subset.

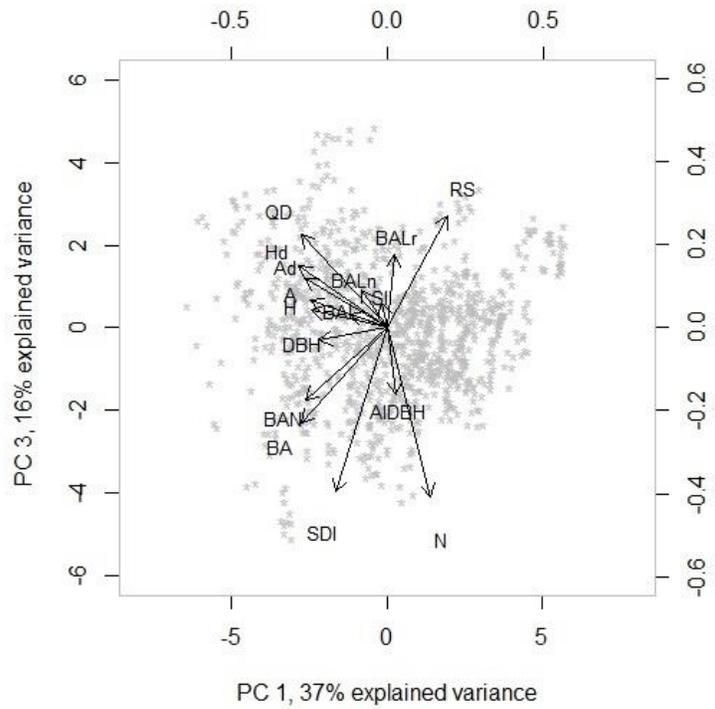


Figure C-4. PC1 and PC3 with scores and factor loadings, using tree and stand subset.

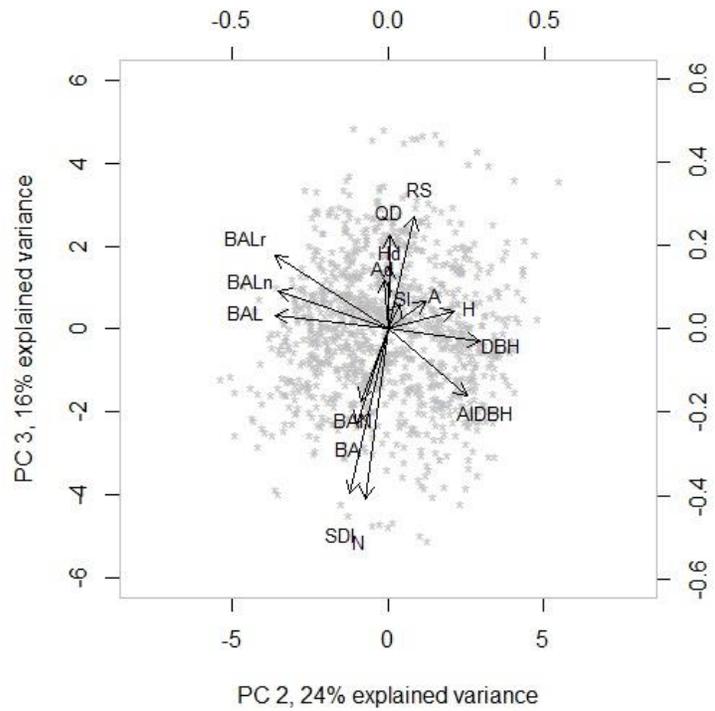


Figure C-5. PC2 and PC3 with scores and factor loadings, using tree and stand subset.

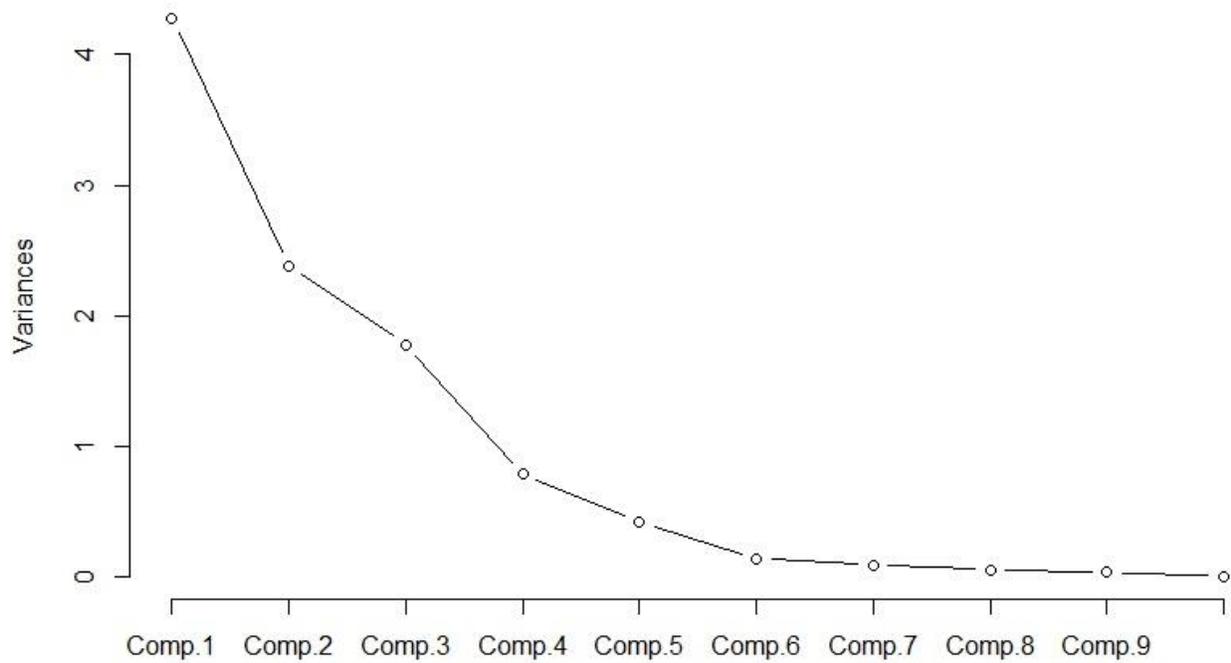


Figure C-6. Scree plot of PCA for continuous variables of stand subset and average plot of AIDBH.

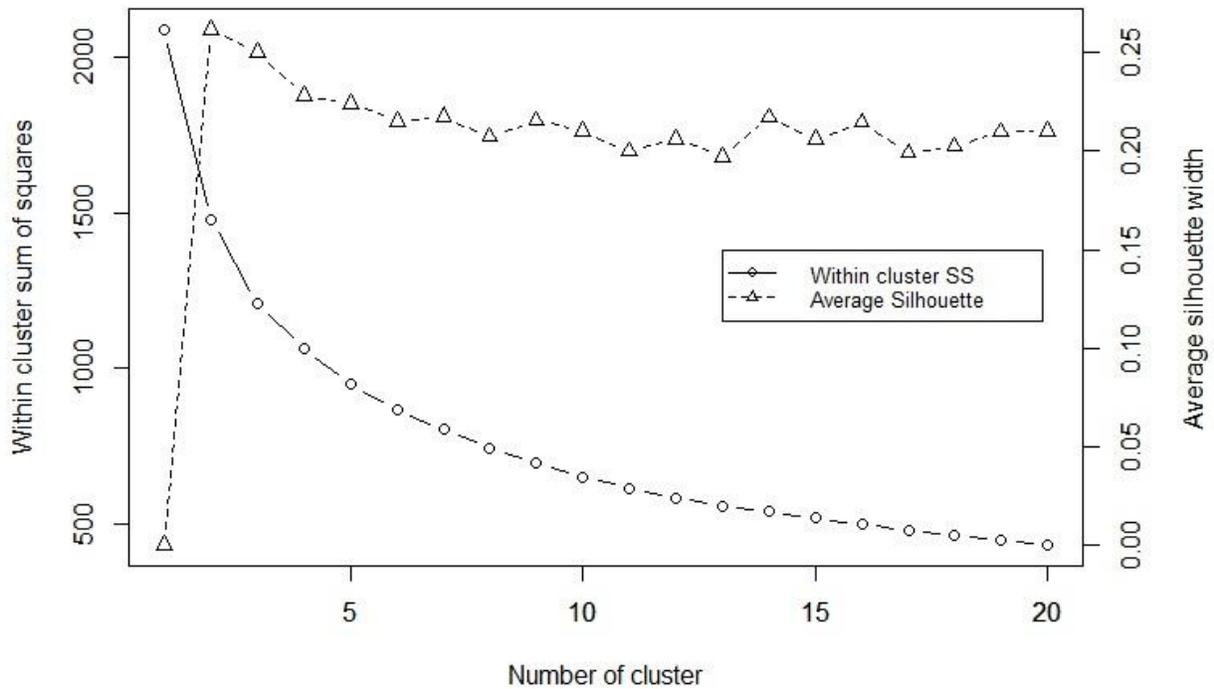


Figure C-7. Scree plot of optimal number of cluster using within sum of squares and average silhouette width for stand-level information.

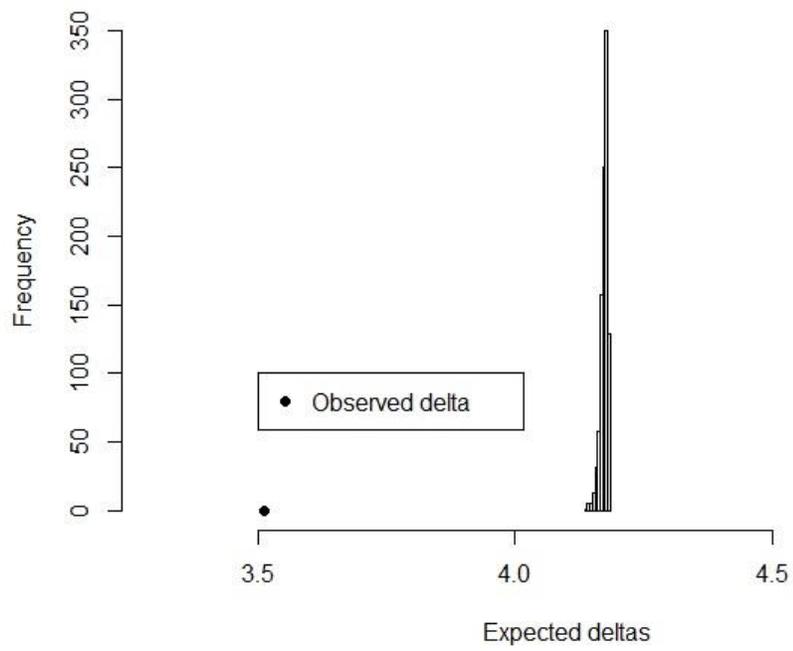


Figure C-8. Histograms of expected deltas and observed delta obtaining by MRPP test for two K-means cluster using stand-level information.

LIST OF REFERENCES

- Anderson MJ, Ellingsen KE, McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters* 9:683–693
- Andreassen K, Tomter SM (2003) Basal area growth models for individual trees of Norway spruce, Scots pine, birch and other broadleaves in Norway. *For Ecol Manage* 180:11–24
- Avery TE, Burkhart HE (2002) *Forest measurements*, 5th edn. McGraw-Hill, New York
- Baskerville GL (1972) Use of logarithmic regression in the estimation of plant biomass. *Can J For Res* 2:49–53
- Birdsey R (1990) Updating methods for forest inventories: an overview. In: LaBau VJ, Cunia T (eds) *State of the art methodology of forest inventory: a symposium proceedings*. USDA For Serv Gen Tech, pp 429–435
- Borcard D, Gillet F, Legendre P (2011) *Numerical ecology with R*. Springer, New York
- Buckman RE (1962) Growth and yield of red pine in Minnesota. USDA Forest Service, Technical Bulletin 1272, Washington, DC
- Burkhart H (2003) Suggestions for choosing an appropriate level for modeling forest stands. In: Amaro A, Reed D, Soares P (eds.) *Modelling forest systems*. CAB International, Wallingford, pp 3–10
- Burkhart HE, Tome M (2012) *Modeling Forest Trees and Stands*. Springer, Netherlands
- Chauchard L, Sbrancia R (2003) Diameter growth models for *Nothofagus obliqua*. *Bosque* 24(3):3–16
- CONAF (Corporación Nacional Forestal, CL), CONAMA (Comisión Nacional del Medio Ambiente, CL), BIRF (Banco Internacional de Reconstrucción y Fomento, USA), Universidad Austral de Chile, Pontificia Universidad Católica de Chile, Universidad Católica de Temuco (1999) *Proyecto catastro y evaluación de los recursos vegetacionales nativos de Chile*. Ministerio de agricultura, Santiago, Chile
- CONAF (2011) *Catastro de los recursos vegetacionales nativos de Chile. Monitoreo de cambios y actualizaciones. Periodo 1997 - 2011*. Ministerio de agricultura, Santiago, Chile
- Contreras M, Affleck D, Chung W (2011) Evaluating tree competition indices as predictors of basal area increment in western Montana forests. *For Ecol Manage* 262:1939–1949

- Coomes DA, Allen RB (2007) Effects of size, competition and altitude on tree growth. *Journal of Ecology* 95:1084–1097
- Cubillos V (1987) Modelos de crecimiento diametral para algunos renovales de Raulí. *Cienc invest for* 1(1):67–76.
- Donoso C (1981) *Ecología Forestal*. Editorial Universitaria, Santiago, Chile
- Donoso C (1987) Variación natural en especies de *Nothofagus* en Chile. *Bosque* 8(2):85–97
- Donoso C (1993) *Bosques templados de Chile y Argentina. Variación, estructura y dinámica*. Editorial Universitaria, Santiago, Chile
- Donoso C, Premoli A, Gallo L, Ipinza R (2004) Variación intraespecífica en las especies arbóreas de los bosques templados de Chile y Argentina. Editorial Universitaria, Santiago, Chile
- Donoso PJ, Donoso C, Sandoval V (1993) Proposición de zonas de crecimiento de renovales de roble (*Nothofagus obliqua*) y raulí (*Nothofagus alpina*) en su rango de distribución natural. *Bosque* 14(2):37–55
- Donoso PJ, Lusk CH (2007) Differential effects of emergent *Nothofagus dombeyi* on growth and basal area of canopy species in an old-growth temperate rainforest. *J Veg Sci* 18(5):675–684
- Donoso PJ, Soto D (2016) Does site quality affect the additive basal area phenomenon? Results from Chilean old-growth temperate rainforests. *Can J For Res* 00:1330–1336
- Echeverria C, Lara A (2004) Growth patterns of secondary *Nothofagus obliqua*–*N. alpina* forests in southern Chile. *For Ecol Manage* 195:29–43
- Esse C, Donoso PJ, Gerding V, Encina-Montoya F (2013) Determination of homogeneous edaphoclimatic zones for the secondary forests of *Nothofagus dombeyi* in central-southern Chile. *Cien Inv Agr* 40(2):351–360
- Everitt B, Hothorn T (2011) *An introduction to applied multivariate analysis with R*. Springer, New York
- Fox J, Monette G (1992) Generalized collinearity diagnostics. *JASA* 87:178–183
- Friedman J, Hastie T, Simon N, Tibshirani R (2016) glmnet: Lasso and elastic-net regularized generalized linear models. R package version 2.0-5

- Gezan SA, Moreno P (1999). Establecimiento y medición de una red de parcelas permanentes para renovales de roble (*Nothofagus obliqua*), raulí (*Nothofagus alpina*) y coigüe (*Nothofagus dombeyi*). *Bosque Nativo (Valdivia)* 22:9–11
- Gezan SA, Moreno P, Ortega A (2009) Taper models for roble, raulí and coigüe second growth forests in Chile. *Bosque* 30(2):61–69
- Gezan SA, Ortega A, Andenmatten E (2007) Diagramas de manejo de densidad para renovales de roble, raulí y coigüe en Chile. *Bosque* 28(2):97–105
- Gonzalez-Benecke CA, Gezan SA, Leduc DJ, Martin TA, Cropper Jr. WP, Samuelson LJ (2012) Modeling survival, yield, volume partitioning and their response to thinning for longleaf pine plantations. *Forests* 3(4):1104–1132
- Harrison WC, Burk TE, Beck DE (1986) Individual tree basal area increment and total height equations for Appalachian mixed hardwoods after thinning. *South J Appl For* 10:99–104
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning* (2nd edition): data mining, inference, and prediction. Springer-Verlag, Berlín
- Huang S, Titus SJ (1995) An individual tree diameter increment model for white spruce in Alberta. *Can J For Res* 25:1455–1465
- Huang S, Titus SJ (1999) Estimating a system of nonlinear simultaneous individual tree models for white spruce in boreal mixed-species stands. *Can J For Res* 29:1805–1811
- Huang S, Yang Y, Wang Y (2003) A critical look at procedures for validating growth and yield models. In: Amaro A, Reed D, Soares P (eds.) *Modelling forest systems*. CAB International, Wallingford, pp 271–293
- INFOR (2013) Rentabilidad económica de distintas propuestas silvícolas para renovales de *Nothofagus* en el centro sur de Chile. Informe Técnico 193. <http://biblioteca.infor.cl/DataFiles/31429.pdf>
- INFOR (2016) Bosque nativo 12. <http://wef.infor.cl/publicaciones/br/2016/12/BN201612.pdf> . Accessed 13 March 2017
- Ivancich H, Martínez-Pastur G, Lencinas MV, Cellini JM, Peri PL (2014) Proposals for *Nothofagus antarctica* diameter growth estimation: simple vs. global models. *J For Sci* 60:307–317
- James G, Witten D, Hastie T, Tibshirani R (2013) *An introduction to statistical learning with applications in R*. Springer, New York

- Kuuluvainen T (2002) Natural variability of forests as a reference for restoring and managing biological diversity in boreal Fennoscandia. *Silva Fennica* 36(1):97–125
- Kyung M, Gill J, Ghosh M, Casells G (2010) Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* 5(2):369–412
- Landres PB, Morgan P, Swanson FJ (1999) Overview of the use of natural variability concepts in managing ecological systems. *Ecological Applications* 9:1179–1188
- Lockhart R, Taylor J, Tibshirani RJ, Tibshirani R (2014) A significance test for the lasso. *Ann. Stat.* 42(2):413–468
- Luebert P, Pliscoff P (2006) Sinopsis bioclimática y vegetacional de Chile. Editorial Universitaria, Santiago, Chile
- Lumley T, Miller A (2009) Leaps: regression subset selection. R package version 2.9.
- Lusk CH, Ortega A (2003) Vertical structure and basal area development in second-growth *Nothofagus* stands in Chile. *J Appl Ecol* 50:639–645
- Ma W, Lei X (2015) Nonlinear simultaneous equations for individual-tree diameter growth and mortality model of natural Mongolian oak forests in northeast China. *Forests* 6:2261–2280
- MacDonald PL, Gardner RC (2000) Type I error rate comparisons of post hoc procedures for $I \times J$ chi-square tables. *Educ Psychol Meas* 60(5):735–754
- Maechler M (2013) Cluster analysis extended Rousseeuw et al. R CRAN
- McCune B, Grace JB, Urban, DL (2002). Analysis of ecological communities. MjM Software Design, Glenden Beach
- Monserud RA, Sterba H (1996) A basal area increment model for individual trees growing in even- and uneven-aged forest stands in Austria. *For Ecol Manage* 80:57–80
- Moreno P (2001) Proposición preliminar de curvas de índice de sitio para renovales de raulí. Tesis Ingeniero Forestal. Universidad de Chile, Chile
- Murphy P, Graney D (1998) Individual-tree basal area growth, survival, and total height models for upland hardwoods in the Boston Mountains of Arkansas. *South J Appl For* 22:184–92
- Novoa R, Villaseca C (1989). Mapa agroclimático de Chile. Instituto de Investigaciones Agropecuarias (INIA), Ministerio de Agricultura, Santiago, Chile

- Nunes L, Tome J, Tome M (2011) Prediction of annual tree growth and survival for thinned and unthinned even-aged maritime pine stands in Portugal from data with different time measurement intervals. For *Ecol Manage* 262:1491–1499
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R B, ... , Oksanen M J (2013) Package 'vegan'. Community ecology package
- Ormazabal C, Benoit I (1987) Present status of conservation of the genus *Nothofagus* in Chile. *Bosque* 8(2):109–120
- Ortega A, Gezan S (1998) Cuantificación de crecimiento y proyección de calidad en *Nothofagus*. *Bosque* 19(1):123–126
- Peng C (2000) Growth and yield models for uneven-aged stands: past, present and future. For *Ecol Manage* 132(2):259–279
- Pukkala T, Lähde E, Laiho O (2009) Growth and yield models for uneven-sized forest stands in Finland. For *Ecol Manage* 258:207–216
- Pukkala T, Lähde E, Laiho O (2011) Using optimization for fitting individual-tree growth models for uneven-aged stands. *Eur J Forest Res* 130:829–839
- Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press. Cambridge
- R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Richardson SJ, Hurst JM, Easdale TA, Wiser SK, Griffiths AD, Allen RB (2011) Diameter growth rates of beech (*Nothofagus*) trees around New Zealand. *N Z J For* 56(1):3–11
- Salas C, Garcia O (2006) Modelling height development of mature *Nothofagus obliqua*. For *Ecol Manage* 229:1–6
- Salas C, Gregoire T, Craven D, Gilabert H (2016) Forest growth modelling: the state of the art. *Bosque* 37(1):3–12
- Schlatter J, Gerding V (1995) A method of site classification for forestry production: A Chilean example. *Bosque* 16(2):13–20
- Schlatter J, Gerding V, Adiazola J (1994) Sistema de ordenamiento de la tierra. Herramienta para la planificación forestal, aplicada a las regiones VII, VIII y IX. Serie técnica, Facultad Ciencias Forestales, Universidad Austral de Chile. Valdivia, Chile.

- Schlatter J, Gerding V, Huber H (1995) Sistema de ordenamiento de la tierra. Herramienta para la planificación forestal, aplicada para la X región. Serie técnica, Facultad Ciencias Forestales, Universidad Austral de Chile. Valdivia, Chile
- Shortt JS, Burkhart HE (1996) A comparison of loblolly pine plantation growth and yield models for inventory updating. *South J Appl For* 20:15–22
- Thiers O, Gerding V, Hildebrand E (2008) Renovales de *Nothofagus obliqua* en centro y sur de Chile: Factores de sitio relevantes para su productividad. In: Bava J, Picco OA, Pildaín MB, López P, Orellana I (eds) Libro de actas de Eco Reuniones: segunda reunión sobre los *Nothofagus* en la Patagonia. Esquel, Argentina, pp 255–260
- Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *J R Statist Soc B* 58(1):267–288
- Ugalde G (2006) Crecimiento en altura de renovales de lenga (*Nothofagus pumilio* (Poepp. Et Endl.) Krasser) en Monte Alto (XII región) en función de la calidad de sitio. Tesis Ingeniero Forestal. Universidad de Chile, Chile
- Vanclay JK (1994) Modelling forest growth and yield: applications to mixed tropical forests. CAB International, Wallingford, UK
- Welham SJ, Gezan SA, Clark SJ, Mead A (2015) Statistical methods in biology: design and analysis of experiments and regression. CRC Press, Boca Raton
- Wu TT, Chen YF, Hastie T, Sobel E, Lange K (2009) Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* 25(6):714–721
- Wykoff WR (1990) A basal area increment model for individual conifers in the northern Rocky Mountains. *For Sci* 36(4):1077–1104

BIOGRAPHICAL SKETCH

Paulo C. Moreno Meynard has experience in a broad range of aspects related to forest resources. He received a title of Forestry Engineer (as US Bachelor) from the Universidad de Chile, Chile, a certificate in Geographic Information Systems and Remote Sensing in Agroforestry Management from Universidad Católica de Temuco, Chile, a certificate in Geospatial Analysis from University of Florida, and completed his Master of Science in the School of Forest Resources and Conservation at the University of Florida. After his undergraduate experience, Paulo was worked as forest researcher at Universidad Austral de Chile, Valdivia, where he started to work with forest inventories, GIS, biometry and statistics focused in the species *Nothofagus alpina*, *N. obliqua*, and *N. dombeyi*. After this, he worked as regional manager at Forest Research Institute of Chile, in Coyhaique, small city located in Chilean Patagonia. Here, Paulo developed research related to management and genetic of *N. pumilio*, biometry and genetic of *Pinus ponderosa*, environmental restoration, and CDM projects of Kyoto's Protocol. In 2009, Paulo was employed as project engineer at Forestal Mininco SA, in Coyhaique, where he contributed to obtain the FSC certification, additionally he worked in production of native species by cuttings, management for *Pinus ponderosa*, *Pinus contorta*, and *Pseudotsuga menziesii*, biometry, inventories, and GIS. Paulo in 2014, also was a short consultant for World Bank related to mechanisms to reduce carbon emissions, REDD+.