

Predicting forest productivity using Wet Areas Mapping and other remote sensed environmental data

by

Ivan Bjelanovic

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Forest Biology and Management

Department of Renewable Resources
University of Alberta

©Ivan Bjelanovic, 2016

Abstract

Understanding variability in forest productivity is important to sustainable forest management. The main objective of this thesis was to evaluate efficient and cost-effective ways to predict potential forest productivity (Site Index) using ecological site data obtained from either ground-based sampling or remote sensing. A geocentric approach to estimating SI was selected in order to utilize digital biophysical data for mapping fine scale variability in SI. LiDAR generated Digital Elevation Models (DEM) and Wet Areas Mapping (WAM) provide remotely sensed environmental data at a 1 m resolution for most forested land in Alberta.

Relationships between environmental factors and SI of three major commercial tree species (trembling aspen, lodgepole pine, white spruce) were examined in this study by establishing a network of temporary sample plots in the Lower Foothills Natural Subregion in central Alberta. Data collection involved determination of SI and common ecological field assessment. Strong correlations were found between field determined soil properties and topography of the research area. Six different Flow Initiation Areas (FIA), from 0.5 ha to 10 ha, were tested to reveal optimal FIA for calculation of the Depth-To-Water (DTW) index for SI prediction. Results show that DTW based on smaller FIA was better in estimating aspen SI, the largest size of FIA was best for spruce SI, while the size of FIA did not influence pine SI estimation. A total of 36 species-specific Site Index models were developed for each of three data sources (DEM+WAM, WAM, field assessment) and four modeling methods (MLR, GAM, RT, RF). In terms of best predictors, among remotely sensed variables DTW was selected by each statistical method for each species and in most cases DTW is the strongest predictor in the model, while among ground-based measured variables different variables appeared as the most important for different species according to silvics specifics. In addition to revealing different major drivers, different

strength of relationship was found between species. Prediction accuracy of models obtained is consistent with other similar SI-environment studies. Poorer results for spruce, than for aspen and pine, result from the wide range in ages of sampled spruce stands, the lack of spruce stands across a full range of sites, and the small number of spruce stands with top height trees free of suppression available for sampling. No significant differences in variation explained were observed between DEM+WAM and ground-based models, while WAM data by itself explained most of the total amount of SI variation explained. All four statistical methods could be used in examining SI-environment relationships but with some advantages and disadvantages for each related to data and application specifics.

The study revealed that forest productivity is subject to topographic controls in this study area and variation in productivity could be explained using remotely sensed environmental data. In addition, different tree species respond differently due to contrasting autoecology. SI maps for all species were produced and plausible relationships between terrain attributes and patterns of low and high predicted productivity are generally apparent. This approach appears to adequately portray variation in productivity over short distances and is potentially applicable to forest growth and yield modeling and silviculture planning.

Preface

This thesis is an original work by Ivan Bjelanovic. No part of this thesis has been previously published.

Acknowledgements

First of all, I am very thankful to my supervisor, Prof. Phil Comeau, for the opportunity to work on this challenging project, as well as for guidance, patience, and many opportunities provided throughout whole my graduate studies which helped me to improve my knowledge and skills. Phil's constant positive and supportive feedback made me feel encouraged and self-confident to give my best to complete this studies. I would also like to thank my supervisory Committee members, to Dr. Barry White for support and suggestions during my work on the thesis and providing me opportunity to present results to public at the workshop organized by the Government of Alberta, and to Dr. Mike Bokalo for feedback while evaluating the progress of my thesis project.

I am grateful to Alberta Agriculture and Forestry for providing data and financial support for completion of this studies. I would like to thank Jae Ogilvie for updating me with the latest version of WAM data and Chris Bater for advice and help in obtaining the data from Alberta Agriculture and Forestry. I would also like to acknowledge Alberta Plywood Ltd. and Greenlink Forestry Inc for giving me access to AVI data of the research area. I would like to express many thanks to Craig Cruikshank for providing me great assistance with the field and lab work but also contributed to have great field work experience and to have nice time in Slave Lake during field work. I am so thankful to Gabriel Oltean for amazing help and time spent sharing his experience with field work and data processing. Many thanks to Kirk Johnson and Valerie Krebs for very useful advice and help, and to all my labmates and all great people from the department for creating instructive and friendly environment.

Finally, I would not have been able to get this thesis done without understanding and support of my wife Branka and daughter Lana. And I am also thankful to my sister and parents for their faraway support.

Table of Contents

1. INTRODUCTION AND BACKGROUND.....	1
1.1. Forest site productivity indicators	1
1.2. Relationship between Site Index and environmental factors	6
1.3. Remote sensing in estimating Site Index	13
1.4. Wet Areas Mapping (WAM) explains site productivity	15
1.5. Productivity of western boreal forests.....	17
1.6. Objectives, hypothesis.....	21
2. MATERIALS AND METHODS	23
2.1. Study site selection and description	23
2.2. Sampling design, plot selection and tree selection criteria.....	25
2.3. Data collection	27
2.4. Variable estimation	28
2.5. Analysis.....	35
3. RESULTS	42
3.1. Variation in Site Index and explanatory variables	42
3.2. Correlation analysis and Predictor variable selection for SI modeling	50
3.2.1. Trembling aspen.....	51
3.2.2. Lodgepole pine.....	54
3.2.3. White spruce.....	54
3.3. Depth-To-Water (DTW) index selection	56
3.4. Modeling Site Index variation.....	61
3.4.1. Trembling aspen.....	61
3.4.2. Lodgepole pine.....	71
3.4.3. White spruce.....	79
4. DISCUSSION	88
4.1. Site Index variation and environmental variables	88
4.2. The main productivity drivers	91
4.3. Site Index prediction models - performance and application.....	97
5. CONCLUSIONS	113
REFERENCES.....	115
APPENDIX.....	127

List of Tables

TABLE 2.1. REGRESSION COEFFICIENTS FOR THE GYPSY TOP HEIGHT MODELS (HUANG ET AL., 2009).....	29
TABLE 2.2. SUMMARY OF STAND ATTRIBUTES FOR THE THREE STUDIED TREE SPECIES.	32
TABLE 2.3. SUMMARY OF REMOTELY SENSED VARIABLES CONSIDERED IN MODELING SI BY STUDIED TREE SPECIES. ...	33
TABLE 2.4. SUMMARY OF GROUND-MEASURED VARIABLES CONSIDERED IN MODELING SI BY STUDIED TREE SPECIES.	34
TABLE 3.1. KENDALL’S RANK CORRELATION COEFFICIENT MATRIX BETWEEN REMOTELY SENSED ENVIRONMENTAL AND SOIL ATTRIBUTES. BOLDDED NUMBERS INDICATE SIGNIFICANCE OF RELATIONSHIP AT $p<0.05$	50
TABLE 3.2. KENDALL’S RANK CORRELATION COEFFICIENT MATRIX BETWEEN ASPEN SI AND ENVIRONMENTAL VARIABLES. BOLDDED NUMBERS INDICATE SIGNIFICANCE OF RELATIONSHIP AT $p<0.05$. ITALICIZED NUMBERS INDICATE COEFFICIENT OF DETERMINATION FROM QUADRATIC REGRESSION.	52
TABLE 3.3. KENDALL’S RANK CORRELATION COEFFICIENT MATRIX BETWEEN PINE SI AND ENVIRONMENTAL VARIABLES. BOLDDED NUMBERS INDICATE SIGNIFICANCE OF RELATIONSHIP AT $p<0.05$ (* $p=0.06$). ITALICIZED NUMBERS INDICATE COEFFICIENT OF DETERMINATION FROM QUADRATIC REGRESSION.	53
TABLE 3.4. KENDALL’S RANK CORRELATION COEFFICIENT MATRIX BETWEEN W. SPRUCE SI AND ENVIRONMENTAL VARIABLES. BOLDDED NUMBERS INDICATE SIGNIFICANCE OF RELATIONSHIP AT $p<0.05$. ITALICIZED NUMBERS INDICATE COEFFICIENT OF DETERMINATION FROM QUADRATIC REGRESSION (*EXPONENTIAL TRANSFORMATION).	55
TABLE 3.5. FIT STATISTICS FOR SI MODELS BASED ON DTW CALCULATED AT SIX DIFFERENT WAM'S FLOW INITIATION AREA (FIA) BY SPECIES. MODELS WERE FITTED USING GENERALIZED ADDITIVE MODELS (GAM) AND LINEAR REGRESSION (LR) APPROACHES. ABBREVIATIONS: EDF - EFFECTIVE DEGREES OF FREEDOM; RMSE - ROOT MEAN SQUARED ERROR; R^2 - COEFFICIENT OF DETERMINATION (DEVIANCE EXPLAINED); AIC - AKAIKE INFORMATION CRITERIA; SIGNIFICANCE CODES - *** $p<0.001$, ** $p<0.01$, * $p<0.05$, n.s. $p>0.05$. THE BEST MODELS ARE SHOWN IN BOLD.	56
TABLE 3.6. SELECTED MLR SI PREDICTION MODELS FOR ASPEN USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL).	63
TABLE 3.7. SELECTED GAM SI PREDICTION MODELS FOR ASPEN USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL).	65
TABLE 3.8. FIT STATISTICS AND MODELS RULES IN EXPLAINING ASPEN SI FOR SELECTED RT MODELS USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL). THE TERMINAL NODE NUMBERING REFERS TO CORRESPONDING MODELS IN FIGURE 3.18.	67
TABLE 3.9. FIT STATISTICS AND VARIABLE IMPORTANCE FOR SELECTED RF MODELS EXPLAINING ASPEN SI BASED ON REMOTELY SENSED VARIABLES (RS MODEL), WAM VARIABLES (WAM MODEL), AND GROUND-BASED MEASURED VARIABLES (GB MODEL). ADJ. R^2 - CALCULATED USING "PSEUDO R^2 " (VALIDATION SAMPLES); %INC MSE - VARIABLE IMPORTANCE MEASURE INDICATING AVERAGE PERCENT CHANGE IN THE MEAN SQUARE ERROR WHEN THE PARTICULAR VARIABLE IS PERMUTED WHILE ALL OTHERS ARE RETAINED UNCHANGED.	70
TABLE 3.10. SELECTED MLR SI PREDICTION MODELS FOR PINE USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL). ...	72
TABLE 3.11. SELECTED GAM SI PREDICTION MODELS FOR PINE USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL).	74
TABLE 3.12. FIT STATISTICS AND MODELS RULES IN EXPLAINING PINE SI FOR SELECTED RT MODELS USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-	

BASED MEASURED VARIABLES (GB MODEL). THE TERMINAL NODE NUMBERING REFERS TO CORRESPONDING MODELS IN FIGURE 3.27.....	76
TABLE 3.13. FIT STATISTICS AND VARIABLE IMPORTANCE FOR SELECTED RF MODELS EXPLAINING PINE SI BASED ON REMOTELY SENSED VARIABLES (RS MODEL), WAM VARIABLES (WAM MODEL), AND GROUND-BASED MEASURED VARIABLES (GB MODEL). ADJ.R ² - CALCULATED USING "PSEUDO R ² " (VALIDATION SAMPLES); %INC MSE - VARIABLE IMPORTANCE MEASURE INDICATING AVERAGE PERCENT CHANGE IN THE MEAN SQUARE ERROR WHEN THE PARTICULAR VARIABLE IS PERMUTED WHILE ALL OTHERS ARE RETAINED UNCHANGED.	79
TABLE 3.14. SELECTED MLR SI PREDICTION MODELS FOR SPRUCE USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL).....	81
TABLE 3.15. SELECTED GAM SI PREDICTION MODELS FOR SPRUCE USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL).....	83
TABLE 3.16. FIT STATISTICS AND MODELS RULES IN EXPLAINING SPRUCE SI FOR SELECTED RT MODELS USING REMOTELY SENSED VARIABLES (RS MODEL), USING WAM VARIABLES (WAM MODEL), AND USING GROUND-BASED MEASURED VARIABLES (GB MODEL). THE TERMINAL NODE NUMBERING REFERS TO CORRESPONDING MODELS IN FIGURE 3.34.....	84
TABLE 3.17. FIT STATISTICS AND VARIABLE IMPORTANCE FOR SELECTED RF MODELS EXPLAINING SPRUCE SI BASED ON REMOTELY SENSED VARIABLES (RS MODEL), WAM VARIABLES (WAM MODEL), AND GROUND-BASED MEASURED VARIABLES (GB MODEL). ADJ.R ² - CALCULATED USING "PSEUDO R ² " (VALIDATION SAMPLES); %INC MSE - VARIABLE IMPORTANCE MEASURE INDICATING AVERAGE PERCENT CHANGE IN THE MEAN SQUARE ERROR WHEN THE PARTICULAR VARIABLE IS PERMUTED WHILE ALL OTHERS ARE RETAINED UNCHANGED.	87
TABLE 4.1. VARIABLES INCLUDED AND MODEL PERFORMANCE MEASURES OF SELECTED MODELS BY SPECIES, DATA SOURCES AND MODELING TECHNIQUES. ABBREVIATIONS: RMSE - ROOT MEAN SQUARED ERROR; rRMSE - NORMALIZED RMSE TO THE MEAN OF MEASURED DATA; VARIANCE EXPLAINED - R ² ADJUSTED FOR THE SAMPLE SIZE AND THE NUMBER OF VARIABLES INCLUDED IN THE MODEL; BIC - BAYESIAN INFORMATION CRITERIA. NOTE: PERFORMANCE MEASURES OF RF MODELS WERE COMPUTED FROM CROSS-VALIDATION (OOB) SAMPLES.	93
TABLE 4.2. SUMMARY BY SPECIES OF RESPONSES IN SI ON SITE FACTORS INCLUDED IN FINAL MODELS; ▲ - POSITIVE, ▼ - NEGATIVE, ■ - NO RESPONSE, () - TREND DETECTED ONLY BY SOME STATISTICAL METHODS, * NO SIGNIFICANT RELATIONSHIP BETWEEN SI AND SINGLE PARTICULAR PREDICTOR.	95

List of Figures

FIGURE 1.1. SCHEMATIC CROSS SECTION IN LOWER FOOTHILLS NSR REPRESENTED BY ECOSITE PHASE AND SOIL PROFILES (TAKEN FROM BECKINGHAM ET AL., 1996)	17
FIGURE 2.1. STUDY SITE LOCATION AND NATURAL SUBREGIONS (FORESTED) OF ALBERTA (UPPER RIGHT MAP). SAMPLE PLOT LOCATIONS AND ECOSITE (LOWER FOOTHILLS NSR) MAP OF THE STUDY SITE (LOWER MAP).	24
FIGURE 2.2. OOB ERROR IN RELATION TO NUMBER OF DECISION TREES GROWN TO BUILD RANDOM FOREST MODEL. A) SI PREDICTION MODEL FROM REMOTELY SENSED VARIABLES; B) SI PREDICTION MODEL FROM GROUND-BASED VARIABLES.	41
FIGURE 3.1. SI VARIATION BY SPECIES WITH TOTAL BASAL AREA (UPPER PLOTS) AND MAIN SPECIES BA PROPORTION (LOWER PLOTS). COEFFICIENT OF CORRELATION (R) AND ASSOCIATED P-VALUE ARE PRESENTED FOR EACH PLOT, RED LINE SHOWS LINEAR TREND IN DATA.	43
FIGURE 3.2. AGE STRUCTURE BY SPECIES. HISTOGRAMS FOR ASPEN AND PINE SHOW TIME OF REGENERATION OCCURRENCE AFTER THE MOST RECENT FIRE	44
FIGURE 3.3. SAMPLE PLOT DISTRIBUTION BY SPECIES ACROSS ASPECT (POLAR AXIS) AND ALTITUDE (Y-AXIS)	44
FIGURE 3.4. BOXPLOTS WITH JITTER PLOTS OF SI AND ENVIRONMENTAL (REMOTELY SENSED AND GROUND-BASED MEASURED) VARIABLES FOR ASPEN. BLACK DOT REPRESENTS MEAN.	45
FIGURE 3.5. BOXPLOTS WITH JITTER PLOTS OF SI AND ENVIRONMENTAL (REMOTELY SENSED AND GROUND-BASED MEASURED) VARIABLES FOR PINE. BLACK DOT REPRESENTS MEAN.	46
FIGURE 3.6. BOXPLOTS WITH JITTER PLOTS OF SI AND ENVIRONMENTAL (REMOTELY SENSED AND GROUND-BASED MEASURED) VARIABLES FOR SPRUCE. BLACK DOT REPRESENTS MEAN.	46
FIGURE 3.7. SCATTER PLOTS WITH PENALIZED REGRESSION SPLINE (3 DEGREES OF FREEDOM) BETWEEN ENVIRONMENTAL (REMOTELY SENSED AND GROUND-BASED MEASURED) VARIABLES AND SITE INDEX FOR ASPEN.	47
FIGURE 3.8. SCATTER PLOTS WITH PENALIZED REGRESSION SPLINE (3 DEGREES OF FREEDOM) BETWEEN ENVIRONMENTAL (REMOTELY SENSED AND GROUND-BASED MEASURED) VARIABLES AND SITE INDEX FOR PINE.	48
FIGURE 3.9. SCATTER PLOTS WITH PENALIZED REGRESSION SPLINE (3 DEGREES OF FREEDOM) BETWEEN ENVIRONMENTAL (REMOTELY SENSED AND GROUND-BASED MEASURED) VARIABLES AND SITE INDEX FOR SPRUCE.	49
FIGURE 3.10. RELATIONSHIP BETWEEN SI AND SELECTED DTW INDEX FITTED LINEAR FUNCTION FOR ASPEN AND LOGARITHMIC FUNCTION FOR PINE AND SPRUCE. DTW IS CALCULATED AT FIA OF 1 HA FOR ASPEN, 2 HA FOR PINE, AND 10 HA FOR SPRUCE.....	57
FIGURE 3.11. RESPONSE CURVES (GAM METHOD) TO SELECTED DTW INDEX FOR EACH SPECIES. THE Y-AXIS SHOWS EFFECT ON SI WITH VALUES CENTERED ON ZERO. "RUG" LINES ALONG X-AXIS SHOW COVARIATE VALUES, DOTS REPRESENT RESIDUALS. SHADING SHOWS CONFIDENCE BANDS (± 2 SE) OF THE ESTIMATES.	58
FIGURE 3.12. SCATTERPLOTS OF RESIDUALS VS. PREDICTED SI VALUES AND MEASURED VS. PREDICTED SI VALUES WITH 1:1 LINE (RED COLOR) BY SPECIES AND STATISTICAL METHOD OF SI MODELS BASED ON SELECTED DTW INDEX.	60
FIGURE 3.13. DECISION SUPPORT PLOTS FOR ASPEN MLR MODEL SELECTION BASED ON $ADJ.R^2$ AND BIC STATISTICS; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. $ADJ.R^2$ AND BIC ARE GIVEN FOR THE BEST FITTED MODEL FOR PARTICULAR NUMBER OF VARIABLES INCLUDED. DOTTED LINE INDICATES SELECTED MODEL. DASHED LINE MEANS THAT MODEL (AND ALL MODELS TO THE RIGHT) INCLUDES AT LEAST ONE NON SIGNIFICANT VARIABLE.	62
FIGURE 3.14. SELECTED ASPEN SI PREDICTION MODELS BASED ON WAM VARIABLES USING MLR METHOD (LEFT PLOT) AND GAM METHOD (RIGHT PLOT). FOR MODEL DESCRIPTION REFER TO TABLE 3.6 AND TABLE 3.7.	63
FIGURE 3.15. DECISION SUPPORT PLOTS FOR ASPEN GAM MODEL SELECTION BASED ON R^2 AND BIC STATISTICS; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. R^2 AND BIC ARE GIVEN FOR	

THE BEST FITTED MODEL FOR PARTICULAR NUMBER OF VARIABLES INCLUDED. DOTTED LINE INDICATES SELECTED MODEL. DASHED LINE MEANS THAT MODEL (AND ALL MODELS TO THE RIGHT) INCLUDES AT LEAST ONE NON SIGNIFICANT VARIABLE.	64
FIGURE 3.16. PARTIAL RESPONSE CURVES OF ASPEN SI TO SELECTED SMOOTH TERMS IN RS (UPPER PLOTS) AND GB (BOTTOM PLOTS) MODELS FROM TABLE 3.6. THE Y-AXIS SHOWS EFFECT ON SI WITH VALUES CENTERED ON ZERO, WITH LABEL INDICATING SMOOTH TERM AND RESULTING EDF. "RUG" LINES ALONG X-AXIS INDICATING COVARIATE VALUES, DOTS REPRESENT RESIDUALS. SHADING SHOWS CONFIDENCE BANDS (± 2 SE) OF THE ESTIMATES.	65
FIGURE 3.17. DECISION SUPPORT PLOTS FOR DETERMINING THE PRUNING RT FOR ASPEN: A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. SOLID LINE PLOTS THE VARIATION EXPLAINED (R-SQUARED) BY THE SPLIT IN THE MODEL. DASHED LINE PLOTS VARIATION EXPLAINED (R-SQUARED) BY THE SPLIT FROM CROSS-VALIDATION (CV). DOTTED LINE INDICATES THE POINT OF PRUNING (NO OF SPLITS LEFT) TO OPTIMAL RT BASED ON MAXIMIZING CV R-SQUARED.	66
FIGURE 3.18. ASPEN SI REGRESSION TREE MODELS; A) REMOTELY SENSED VARIABLES (RS MODEL), B) ONLY WAM VARIABLES (WAM MODEL), C) GROUND-BASED MEASURED VARIABLES (GB MODEL).	68
FIGURE 3.19. VARIABLE IMPORTANCE PLOTS FOR ASPEN SI RANDOM FOREST MODELS BUILT WITH ALL POTENTIAL EXPLANATORY VARIABLES; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. %INC MSE - VARIABLE IMPORTANCE MEASURE INDICATING AVERAGE PERCENT CHANGE IN THE MEAN SQUARE ERROR WHEN THE PARTICULAR VARIABLE IS PERMUTED WHILE ALL OTHERS ARE RETAINED UNCHANGED. IMPORTANT VARIABLES SHOULD CAUSE RELATIVELY LARGE %INC MSE.	69
FIGURE 3.20. DECISION SUPPORT PLOTS BASED ON R^2 AND RMSE FOR SELECTED SI MODELS FROM REMOTELY SENSED VARIABLES (PLOTS A AND B) AND GROUND-BASED MEASURED VARIABLES (PLOTS C AND D). R^2 AND RMSE ARE GIVEN FOR THE BEST FITTED MODEL FOR PARTICULAR NUMBER OF VARIABLES INCLUDED. DOTTED LINE INDICATES SELECTED MODEL (BLACK COLOR - FOR ASPEN, PINE AND RS SPRUCE; RED COLOR FOR GB SPRUCE).	70
FIGURE 3.21. PARTIAL DEPENDENCE PLOTS SHOWING EFFECT OF SELECTED VARIABLES ON ASPEN SI IN RS (UPPER PLOTS) AND GB (BOTTOM PLOTS) MODELS. HUSH MARKS AT THE BOTTOM OF THE PLOT SHOW THE DECILES OF THE VARIABLE.	71
FIGURE 3.22. DECISION SUPPORT PLOTS FOR PINE MLR MODEL SELECTION BASED ON $ADJ.R^2$ AND BIC STATISTICS; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. $ADJ.R^2$ AND BIC ARE GIVEN FOR THE BEST FITTED MODEL FOR PARTICULAR NUMBER OF VARIABLES INCLUDED. DOTTED LINE INDICATES SELECTED MODEL. DASHED LINE MEANS THAT MODEL (AND ALL MODELS TO THE RIGHT) INCLUDES AT LEAST ONE NON SIGNIFICANT VARIABLE.	72
FIGURE 3.23. SELECTED PINE SI PREDICTION MODELS BASED ON RS VARIABLES USING MLR METHOD (UPPER LEFT PLOT), WAM VARIABLES USING MLR METHOD (UPPER RIGHT PLOT), RS VARIABLES USING GAM METHOD (BOTTOM LEFT PLOT), AND WAM VARIABLES USING GAM METHOD (BOTTOM RIGHT PLOT). FOR MODEL DESCRIPTION REFER TO TABLE 3.10 AND TABLE 3.11.	73
FIGURE 3.24. DECISION SUPPORT PLOTS FOR PINE GAM MODEL SELECTION BASED ON R^2 AND BIC STATISTICS; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. R^2 AND BIC ARE GIVEN FOR THE BEST FITTED MODEL FOR PARTICULAR NUMBER OF VARIABLES INCLUDED. DOTTED LINE INDICATES SELECTED MODEL. DASHED LINE MEANS THAT MODEL (AND ALL MODELS TO THE RIGHT) INCLUDES AT LEAST ONE NON SIGNIFICANT VARIABLE.	74
FIGURE 3.25. PARTIAL RESPONSE CURVES OF PINE SI TO SELECTED SMOOTH TERMS IN GB MODEL FROM TABLE 3.10. THE Y-AXIS SHOWS EFFECT ON SI WITH VALUES CENTERED ON ZERO, WITH LABEL INDICATING SMOOTH TERM AND RESULTING EDF. "RUG" LINES ALONG X-AXIS INDICATING COVARIATE VALUES, DOTS REPRESENT RESIDUALS. SHADING SHOWS CONFIDENCE BANDS (± 2 SE) OF THE ESTIMATES.	75
FIGURE 3.26. DECISION SUPPORT PLOTS FOR DETERMINING THE PRUNING RT FOR LODGEPOLE PINE; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. SOLID LINE PLOTS THE VARIATION EXPLAINED (R-SQUARED) BY THE SPLIT IN THE MODEL. DASHED LINE PLOTS VARIATION EXPLAINED (R-	

SQUARED) BY THE SPLIT FROM CROSS-VALIDATION (CV). DOTTED LINE INDICATES THE POINT OF PRUNING (NO OF SPLITS LEFT) TO OPTIMAL RT BASED ON MAXIMIZING CV R-SQUARED.	76
FIGURE 3.27. PINE SI REGRESSION TREE MODELS; A) REMOTELY SENSED VARIABLES (RS MODEL), B) ONLY WAM VARIABLES (WAM MODEL), C) GROUND-BASED MEASURED VARIABLES (GB MODEL).	77
FIGURE 3.28. VARIABLE IMPORTANCE PLOTS FOR PINE SI RANDOM FOREST MODELS BUILT WITH ALL POTENTIAL EXPLANATORY VARIABLES; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. %INC MSE - VARIABLE IMPORTANCE MEASURE INDICATING AVERAGE PERCENT CHANGE IN THE MEAN SQUARE ERROR WHEN THE PARTICULAR VARIABLE IS PERMUTED WHILE ALL OTHERS ARE RETAINED UNCHANGED. IMPORTANT VARIABLES SHOULD CAUSE RELATIVELY LARGE %INC MSE.	78
FIGURE 3.29. PARTIAL DEPENDENCE PLOTS SHOWING EFFECT OF SELECTED VARIABLES ON PINE SI IN RS (UPPER PLOTS) AND GB (BOTTOM PLOTS) MODELS. HUSH MARKS AT THE BOTTOM OF THE PLOT SHOW THE DECILES OF THE VARIABLE.	79
FIGURE 3.30. DECISION SUPPORT PLOTS FOR SPRUCE MLR MODEL SELECTION BASED ON $ADJ.R^2$ AND BIC STATISTICS; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. $ADJ.R^2$ AND BIC ARE GIVEN FOR THE BEST FITTED MODEL FOR PARTICULAR NUMBER OF VARIABLES INCLUDED. DOTTED LINE INDICATES SELECTED MODEL. DASHED LINE MEANS THAT MODEL (AND ALL MODELS TO THE RIGHT) INCLUDES AT LEAST ONE NON SIGNIFICANT VARIABLE.	80
FIGURE 3.31. SELECTED SPRUCE SI PREDICTION MODELS USING MLR METHOD (UPPER PLOTS) AND GAM METHOD (BOTTOM PLOTS) BASED ON RS (WAM) VARIABLES (LEFT PLOTS) AND GB VARIABLES (RIGHT PLOTS). FOR MODEL DESCRIPTION REFER TO TABLE 3.14 AND TABLE 3.15.	81
FIGURE 3.32. DECISION SUPPORT PLOTS FOR SPRUCE GAM MODEL SELECTION BASED ON R^2 AND BIC STATISTICS; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. R^2 AND BIC ARE GIVEN FOR THE BEST FITTED MODEL FOR PARTICULAR NUMBER OF VARIABLES INCLUDED. DOTTED LINE INDICATES SELECTED MODEL. DASHED LINE MEANS THAT MODEL (AND ALL MODELS TO THE RIGHT) INCLUDES AT LEAST ONE NON SIGNIFICANT VARIABLE.	82
FIGURE 3.33. DECISION SUPPORT PLOTS FOR DETERMINING THE PRUNING RT; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. SOLID LINE PLOTS THE VARIATION EXPLAINED (R-SQUARED) BY THE SPLIT IN THE MODEL. DASHED LINE PLOTS VARIATION EXPLAINED (R-SQUARED) BY THE SPLIT FROM CROSS-VALIDATION (CV). DOTTED LINE INDICATES THE POINT OF PRUNING (NO OF SPLITS LEFT) TO OPTIMAL RT BASED ON MAXIMIZING CV R-SQUARED.	84
FIGURE 3.34. SPRUCE SI REGRESSION TREE MODELS; A) REMOTELY SENSED VARIABLES (RS MODEL), B) ONLY WAM VARIABLES (WAM MODEL), C) GROUND-BASED MEASURED VARIABLES (GB MODEL).	85
FIGURE 3.35. VARIABLE IMPORTANCE PLOTS FOR SPRUCE SI RANDOM FOREST MODELS BUILT WITH ALL POTENTIAL EXPLANATORY VARIABLES; A) REMOTELY SENSED VARIABLES, B) GROUND-BASED MEASURED VARIABLES. %INC MSE - VARIABLE IMPORTANCE MEASURE INDICATING AVERAGE PERCENT CHANGE IN THE MEAN SQUARE ERROR WHEN THE PARTICULAR VARIABLE IS PERMUTED WHILE ALL OTHERS ARE RETAINED UNCHANGED. IMPORTANT VARIABLES SHOULD CAUSE RELATIVELY LARGE %INC MSE.	86
FIGURE 3.36. PARTIAL DEPENDENCE PLOTS SHOWING EFFECT OF SELECTED VARIABLES ON SPRUCE SI IN RS (UPPER PLOTS) AND GB (BOTTOM PLOTS) MODELS. HUSH MARKS AT THE BOTTOM OF THE PLOT SHOWING THE DECILES OF THE VARIABLE.	87
FIGURE 4.1. FLOW INITIATION AREA IN RELATION TO SI VARIATION EXPLAIN (LINEAR REGRESSION) BY DTW WHICH IS CALCULATED BASED ON ASSOCIATED FIA.	90
FIGURE 4.2. MODEL COMPARISON BY SPECIES, DATA SOURCES AND MODELING TECHNIQUES BASED ON VARIANCE EXPLAINED. NUMBERS REPRESENT THE MEAN, ERROR BARS REPRESENT STANDARD ERROR OF THE MEAN.	100
FIGURE 4.3. RESIDUAL AND MEASURED VS. PREDICTED PLOTS FOR ASPEN BY DATA SOURCES AND MODELING TECHNIQUES. THE 1:1 LINE IS SHOWN IN RED COLOR.	104
FIGURE 4.4. RESIDUAL AND MEASURED VS. PREDICTED PLOTS FOR PINE BY DATA SOURCES AND MODELING TECHNIQUES. THE 1:1 LINE IS SHOWN IN RED COLOR.	105

FIGURE 4.5. RESIDUAL AND MEASURED VS. PREDICTED PLOTS FOR SPRUCE BY DATA SOURCES AND MODELING TECHNIQUES. THE 1:1 LINE IS SHOWN IN RED COLOR.	106
FIGURE 4.6. HILLSHADE AND MAPPED ASPEN SI (DETAIL) GENERATED BASED ON SELECTED MLR MODEL (TABLE 3.6), GAM MODEL (TABLE 3.7), RT MODEL (TABLE 3.8) AND RF MODEL (TABLE 3.9). DOTS REPRESENT CALIBRATION SAMPLE PLOTS; BLUE LINES REPRESENT WAM PREDICTED STREAM CHANNELS AT 1 HA FIA. MAP RESOLUTION IS 1 M.	110
FIGURE 4.7. HILLSHADE AND MAPPED WHITE SPRUCE SI (DETAIL) GENERATED BASED ON SELECTED MLR MODEL (TABLE 3.14), GAM MODEL (TABLE 3.15), RT MODEL (TABLE 3.16) AND RF MODEL (TABLE 3.17). DOTS REPRESENT CALIBRATION SAMPLE PLOTS; BLUE LINES REPRESENT WAM PREDICTED STREAM CHANNELS AT 10 HA FIA. MAP RESOLUTION IS 1 M.	111
FIGURE 4.8. HILLSHADE AND MAPPED LODGEPOLE PINE SI (DETAIL) GENERATED BASED ON SELECTED MLR MODEL (TABLE 3.10), GAM MODEL (TABLE 3.11), RT MODEL (TABLE 3.12) AND RF MODEL (TABLE 3.13). DOTS REPRESENT CALIBRATION SAMPLE PLOTS; BLUE LINES REPRESENT WAM PREDICTED STREAM CHANNELS AT 2 HA FIA. MAP RESOLUTION IS 1 M.	112

1. INTRODUCTION AND BACKGROUND

Potential site productivity is a quantitative estimate of site quality and depends on the natural factors inherent to the site and plant species (Skovsgaard and Vanclay, 2008). Environmental factors, particularly temperature, soil moisture and soil nutrient availability are dominant factors controlling potential productivity in forest ecosystems (Smith et al., 1997). However, potential site productivity might be increased or decreased by management practices (Eisenbies et al., 2006; Ryu et al., 2006), disturbances (Thornley and Cannell, 2004), or due to changes in environmental conditions and genetics over time, e.g. many studies have found shifts in tree species productivity (Site Index) related to changes in climate (Albert and Schmidt, 2010; Antón-Fernández et al., 2016; Jiang et al., 2015; Latta et al., 2010; Monserud et al., 2008; Nothdurft et al., 2012; Weiskittel et al., 2011a). That implies that a narrower term of actual (realized) site productivity, as given by Skovsgaard and Vanclay (2008): "*Forest site productivity is the production that can be realized at a certain site with a given genotype and a specified management regime*", would be more appropriate, and that the definition should also include a temporal component.

Understanding factors that influence forest site productivity is important to sustainable forest management due to the need for reliable prediction of stand growth and timber yield to support decision making, forest management planning, and annual allowable cut determination. Modeling growth and yield requires good estimates of forest site productivity. Moreover, understanding variation in forest productivity is important for sampling, inventory, monitoring, and scientific investigation in forestry because of the need to identify relatively homogeneous areas for stratification of sampling or conducting field trials (Berrill and O'Hara, 2016).

1.1. Forest site productivity indicators

There are several possible ways to assess forest site productivity. Skovsgaard and Vanclay (2008) reviewed approaches to assessing productivity and classified all methods as either geocentric (earth-based) or phytocentric (plant-based) depending on which indicators are used. Geocentric methods are based on site properties such as climate, topography or soil characteristics, while phytocentric methods rely on characteristics of the vegetation that are

effective representatives of site productivity. A special case of the phytocentric indicators classified as dendrocentric or dendrometricis obtained when measurement is based on measurement of the forest stands or individual trees. Methods could also be classified as direct or indirect based on how close the indicator is related to wood volume production. Soil properties and volume measurements are direct while topographic indices and Site Index are indirect. Permanent site variables (climate, topography, drainage, soil physical properties) have strong influences on site productivity, biological features and chemical soil properties, therefore multifactor methods based on four groups of parameters – climate, landform, soil, vegetation – provide the most complete approach and can improve determination of site productivity (Smith et al., 1997).

Method selection for productivity assessment depends on the purpose, scale, and possibility to obtain direct/indirect indicators. It will also depend on required accuracy, among other things, due to strength of correlations between different productivity indicators. However, accurate determination of forest productivity can be difficult due to variability in time and space of indicators used (Skovsgaard and Vanclay, 2013). Productivity is the ability of the site to produce total biomass (net primary production), however in forestry we often refer to realized productivity and ability of a stand to produce aboveground wood volume (Weiskittel et al., 2011b). Several productivity indicators based on parameters more or less directly related to estimation of stand volume production are found. Mean annual volume increment at time of culmination is commonly used as a measure of site productivity (Latta et al., 2009; Skovsgaard and Vanclay, 2008), while using aboveground live biomass (carbon density) (Zald et al., 2016) might be a more complete approach in assessing forest productivity. Stand volume growth per unit of height growth is a measure of site quality based on three dimensional model of the relationship between stem number, quadratic mean diameter and stand basal area, and can be used to adjust height-age based estimates of site productivity (Skovsgaard and Vanclay, 2008).

Tree-ring width analysis is widely used to determine stand productivity. Site average tree basal area increment (BAI) was used to represent aspen growth performances across Canada (Latutrie et al., 2015), growth of four major tree species across Europe (Laubhann et al., 2009), and the growth of Douglas-fir plantations in Austria and Southern Bavaria (Eckhart et al., 2014). Berrill and O'Hara (2014) described application of basal area increment index (BAIi) in characterizing

stand productivity in mixed even-aged and multi-aged stands and conclude that BAI is a strong predictor of volume increment and recommended its use in growth models. Species diversity and structural diversity (Shannon index) also have positive effects on productivity in mixed uneven-aged forests (Danescu et al., 2016). Another approach which suggests usage of a direct indicator of site productivity is the '300 Index' built as a local site parameter in a stand volume growth model for New Zealand radiata pine forests similar to Site Index but sensitive to stand density in addition to age (Watt et al. 2010). However, there are issues with using these volume and related indicators because they are usually difficult to measure, stand age dependant, and there are problems with their determination in young stands (impossible to measure) and in managed stands (affected by silvicultural treatments such as fertilization, site preparation, and thinning).

Huang and Titus (1993) tested the relationship between height and diameter of dominant and codominant trees to represent site productivity for white spruce in boreal mixedwoods but their h:d ratio did not work well due to diameter being strongly influenced by stand density and management activities. Similarly, Wang (1998) found that height of dominant trees at specific diameter is not an adequate productivity measure in sub-boreal white spruce in B.C. Skovsgaard and Vanclay (2008) suggested that stand height might be a good productivity predictor for a given species within a growth region (site type) and therefore of major practical usage in forest management, while using climatic parameters may be more appropriate for general estimation of productivity at larger scales. Thus trade-offs between model accuracy and geographic extent exist (Bontemps and Bouriaud, 2014). However, opportunities to measure stand and site characteristics by remote sensing provides the possibility of using easily measured parameters with satisfactory accuracy for spatial productivity prediction.

The most frequently used indicator of productivity in even-aged forests is rate of height growth of a given species, hence Site Index (SI) seems to be the most commonly used variable for modeling growth and yield for management purposes (Smith et al., 1997) because it is a numeric description of site productivity that is easy to obtain and uses reference age to eliminate age effect. According to Smith et al. (1997) Site Index is defined as the average height of the dominant (and codominant) trees of an even-aged aggregation of trees at a given reference age. The concept is based on correlation of the productive potential of a site and the rate of height growth of free-growing dominant or codominant trees, which means trees that have not been

suppressed in the earlier development. Site Index does not vary with stand density or stocking since the height growth of the largest trees in even-aged stands is roughly independent of stem number (Berrill and O'Hara, 2014). Hence, the main assumption underlying the Site Index approach is that height of dominant trees is independent of competition, stocking, and management practices, and varies only with site productivity. However, this is sometimes not true for extreme density values, as density dependent height repression has been confirmed for conifers species such as jack pine and black spruce in boreal forests at high density-stress levels (Newton, 2015), or for very old stands where height growth does not relate well to site quality (Robichaud and Methven, 1993). Another assumption is that dominant trees were dominant throughout their lives and any damages in the past did not significantly impact height growth. SI is also assumed to be time independent so any environmental and genetic changes over time, such as changes in climate or moisture, have no impact on height development. Lastly, SI theory is based on Eichhorn's rule, the fact that the relationship between height and age corresponds to the relationship between yield and age (Skovsgaard and Vanclay, 2008).

The Site Index, or height-age model, is an indirect phytocentric method that allows estimation of Site Index from height and age. It is species-specific because of different height growth characteristics for each tree species. In practice for estimation of SI it is necessary to measure height and age of dominant trees in a particular stand. Usually, SI is calculated for several trees in a stand and then averaged (Weiskittel et al., 2011b). An advantage of SI is that height is easily measured, even using remote sensing technology (Tompalski et al., 2015b), however determining age may be difficult due to issues such as off-center increment cores (Applequist, 1958), stem decay, or inaccurate counting due to ring boundary visibility (DeRose and Gardner, 2010), "missing rings" and "white rings" caused by severe forest tent caterpillar defoliation on aspen (Hogg et al., 2002) or overstory suppression for white spruce. To use Site Index, regional height-age relationships (SI curves) which describe height development over time for different levels of SI are needed. Stem analysis or monitoring of height on Permanent or Temporary Sample Plots are methods used for development of Site index curves. In Alberta, Huang et al. (2009) provided Site Index models (height-age curves) for four major commercial tree species (trembling aspen, lodgepole pine, white and black spruce) with top height (average height of the 100 largest DBH trees per ha) and total age as input parameters. SI is widely used in forest growth and yield modeling, e.g. both GYPSY (Huang et al., 2009) and MGM (Bokalo et al.,

2013) currently used in Alberta, and their application requires SI estimates for each species that may occur in the stand.

Although Site Index is widely used as a forest productivity measure, it has some disadvantages and limitations. Weiskittel et al. (2011b) listed the most important ones: it is difficult to apply in multi-layered stands; it can change over time for a given stand; it is highly sensitive to dominant tree selection and measurement errors; it is not applicable in afforested areas, or in very young and older stands; it cannot be used to compare productivity potential between species; and it requires a well-constructed height-age equation. However, several options have been proposed to overcome these issues. It may be possible to make adjustments to SI for trees that are released from competition (Osika et al., 2013). Tree ring analysis can be used to confirm no earlier inter-specific impacts for present dominant trees. Also, uncertainty about past growth of selected dominant trees due to abiotic/biotic damages and defoliation might also bias SI estimation, but could be accounted for in the case of forest tent caterpillar or other events (Latutrie et al., 2015). On the other hand, these disadvantages might limit SI application. In terms of SI time independence, Monserud and Rehfeldt (1990) found that genetic variation and environmental conditions influenced SI variation over the long term. SI is also affected by silvicultural practices such as fertilization and site preparation. Finally, to estimate SI we need reliable measurement of height and age of dominant trees and a SI curve. Site Index estimates depend on the quality of the SI curve, and accuracy decreases significantly as stands diverge from the site index reference age (Weiskittel et al., 2011b). Site Index cannot be used in stands where measurements of height and age are unreliable or impossible, where there are no appropriate dominant trees of the relevant species where trees have been suppressed for part of their lives, or where stands are too young or very old.

Several alternatives for estimating SI are available for use when "height-age" models are not applicable. Periodic height increment can be used in multicohort stands where trees experience periods of suppression or in young stands, and that method is known as height growth intercept (HGI) or variable HGI method for species with or without distinct whorls (Nigh, 1995; Skovsgaard and Vanclay, 2008). Old Growth Site Index Adjustments provide a method for estimating SI for a second-growth stand, which is noticeably higher, from adjusted SI of an old-growth stand (Nigh and Love, 1997). Another way to deal with this issue is Species Conversion

Models which estimate Site Index of one species from that of another species in a mixed stand. This method can be applied when the species of interest is not currently present on site or proper dominant trees could not be selected but growth patterns of both species on particular site must be known (Nigh, 2002). However, Hostin and Titus (1996) found that aspen SI explained only 9% of the variation in white spruce SI.

In British Columbia SIBEC (Site Index-Biogeoclimatic Ecosystem Classification) Models relate SI to ecological site classification units termed site series (with a site series representing all areas that should have the same plant association at maturity and have similar soils) in order to estimate the mean Site Index for each given site series/tree species combination (Mah and Nigh, 2015). In Alberta, estimates of Site Index of common species found in each ecosite are provided in the ecosite guides (Beckingham et al., 1996). However, ecosites often include a range of soil nutrient and moisture classes (Kabzems and Klinka, 1987), and a broad range of site productivity. In addition, natural subregions include some range in climate that may result in variability in height growth rates. Site Index adjustments and more precise estimates of average site index for each ecosite for aspen, spruce and lodgepole pine were also developed in three FMUs in north central Alberta (Canfor/WeyCo Report, 2008).

The disadvantages of the "height-age" method indicate limitations for application of Site Index in forest productivity prediction. Alternative approaches to productivity assessment are available based on parameters more directly related to estimation of volume production. In recent years, instead of 'height-age' models (phytogenic SI) many studies advocate a 'site-growth' concept (biophysical SI) where site properties are related to forest growth and quantify the multiple effects of environmental factors on site quality, using physical indicators of the environment in statistical models of site index (Bontemps and Bouriaud, 2014).

1.2. Relationship between Site Index and environmental factors

Vegetation and geocentric methods may be useful in estimating site productivity or might be linked to volume production through Site Index (Berrill and O'Hara, 2014). According to Bontemps and Bouriaud (2014), these models are referred to as 'Site Index prediction models'

and the relationship between biophysical factors and Site Index as ‘SI-environment relationships’.

Indirect Site Index prediction using ecological variables includes two main approaches (Seynave et al., 2005), a synoptic approach which consists of correlating Site Index to site classes; and an analytic approach which consists of modeling SI as a function of various ecological, topographical and soil variables. Analytical approaches have seen increasing use due to increased availability of the necessary data. Information availability and data collection defines the scale of statistical models for prediction of Site Index (Bontemps and Bouriaud, 2014): broad (provincial/national)-scale (spatially environmental information, remotely sensed, satellite or large forest database - NFI) or landscape/regional-scale (field sampling or models of Site Index developed from local environmental factors). Studies carried out in small extents, where climate factors are assumed to be homogeneous, focusing on local site factors, usually show more accurate prediction with fair to high strength of relationship with R-squared ranging between 0.50 and 0.80 (Bontemps and Bouriaud, 2014), and therefore might provide quantitative operational support. Site Index prediction models are scale-dependent, with effective modeling in smaller areas requiring higher resolution data than is required in larger geographic areas (Aertsens et al., 2012b; Chen et al., 2002).

Plant growth is affected by primary environmental and resources factors (e.g. temperature, soil aeration and moisture, acidity etc.). SI-environment relationships are still autoecological because species respond differently to environmental factors, and these responses can be region-specific, making generalizations difficult (Weiskittel et al., 2011b). Edaphic, physiographic and climatic variables are three main groups of geocentric measures (Weiskittel et al., 2011b).

Edaphic variables are important drivers of growth since they directly impact water and nutrient availability for plants. Main predictors of SI include soil moisture and nutrient regimes, physical and chemical properties, parent material, and humus characteristics. Forest floor, soil and humus characteristics were found to explain 67%, 77% and 68% of SI variation for oak, beech and pine respectively, while soil granulometric fractions and litter nitrogen concentrations were the most effective predictors of forest site productivity in Flanders (Aertsens et al., 2012a). In boreal black spruce forests, organic layer thickness was strongly and negatively ($r = -0.48$) correlated with the

Site Index (Laamrani et al., 2014). In Ireland, Sitka spruce SI increased with increasing SNR, and decreased with excess moisture or moisture deficit, with both SNR and SMR explaining 59% of the variation at country level (Farrelly et al., 2011). Several soil physical and chemical properties were found to be significant and strong predictors of productivity but the best model, including positive effects of nitrogen and negative effects of manganese content, explained about 73% of the variance in SI of Douglas-fir stands in central Europe (Eckhart et al., 2014). Piedallu et al. (2011) examined potential productivity using maps of soil water holding capacity at a large scale and found weak relationships with between 10.3% and 14.1% of the variation in Site Index explained for the three studied tree species.

Depending on the type of parent material soil parent material and mineralogy might or might not impact soil properties, hence forest productivity. For instance, trembling aspen SI was not significantly different between fluvial and till parent material types in the Boreal Shield of Quebec (Pinno et al., 2009) and between fluvial, lacustrine and till in Saskatchewan (Pinno and Bélanger, 2011), but different soil and site variables were selected as productivity drivers for each parent material. In contrast, Site Index values appeared to be different on two sandstone parent rock types in mountain Norway spruce forests in Poland (Socha, 2008).

Physiographic and topographic variables indirectly influence forest productivity by altering microclimate and soil formation processes and hence impact primary site factors influencing tree growth. Simple variables such as slope, aspect, elevation, latitude, longitude, surface curvature, and slope position have been found to be useful in predicting SI in numerous studies. Among the nine topographic attributes that were tested against black spruce SI, slope was the strongest predictor with a positive correlation ($r=0.27$ to 0.35) across all four spatial map resolutions (Laamrani et al., 2014). Aspect was found to be the most important SI predictor in Mediterranean mountain forests for all three studied species and was selected by each of five statistical techniques indicating decline in SI from east to west due to maritime influence by humid wind (Aertsen et al., 2010). Altitude was the major explanatory variable in a Norway spruce SI prediction model with a negative relationship and R-squared of 0.61 (Socha, 2008). However, an issue in modeling with these topographic attributes is their interaction with radiation, precipitation, and temperature. For instance, aspect and slope strongly influence potential direct incident radiation and temperature (McCune and Keon, 2002). Some complex topographic

indices, TRMI (Topographic relative moisture index - combining topographic position, profile curvature, aspect and slope steepness) and TOPEX (Topographic exposure scores), were tested against *Sequoia sempervirens* SI in California and found to correlate well with SI (Berrill and O'Hara, 2016). Availability of Digital Elevation Models allows computation of complex topographic indices. Topographic wetness index (TWI) and Topographic Position Index (TPI) appeared to be well related with black spruce SI in Quebec (Laamrani et al., 2014), while SMI (Soil Moisture Index) and MBI (Mass Balance Index) were selected in the best Norway spruce SI model in Bavaria (Brandl et al., 2014). Usually, aspect, slope and slope position are used for productivity prediction at a local level, while longitude and latitude are often used as proxy for climate in large geographic domains. In a study of relationships between SI and topography, soil moisture, and soil nutrients in Alberta, measures of geographic location were found to explain 37% of the variation in SI, where SI decreased with elevation, increased with latitude, and had a quadratic relationship with longitude (Wang et al., 2004). Also, latitude was the strongest physiographic predictor, in addition to soil physical properties and nutrient availability, in explaining *Pinus sylvestris* SI in acidic plateau plantations of northern Spain which was found to be correlated with several climate variables (Bueis et al., 2016).

Numerous studies on modeling Site Index across large geographic regions include climate variables (Hamel et al., 2004; Jiang et al., 2015; McKenney and Pedlar, 2003; Seynave et al., 2005; Sharma et al., 2012; Ung et al., 2001; Watt et al., 2010b; Weiskittel et al., 2011a), but in different regions and climates various variables have been shown to be good predictors of productivity. An interesting relationship was observed in Ireland between Sitka spruce and windspeed with up to 37% of the SI variation being accounted for by windspeed in the wet climate region (Farrelly et al., 2011). In another case, temperature explained 56% of variation in spruce SI, while both temperature and moisture explained 40% of variation in pine SI and 54% for deciduous dominated stands in a study across Norway (Antón-Fernández et al., 2016). For lodgepole pine in Alberta up to 27% of variation in SI was explained using growing degree days $> 5^{\circ}$ (GDD5), the Julian date when GDD5 reaches 100, or July mean temperature (Monserud et al., 2006). Including variables that represent both temperature and moisture availability are generally more effective in productivity prediction (Latta et al., 2009; Weiskittel et al., 2011b). Climate Moisture Index (CMI), which estimates water stress based on precipitation and potential evapotranspiration, has been found to significantly explain forest productivity (Hogg et al., 2008;

Latta et al., 2009; Latutrie et al., 2015). For efficient spatial productivity predictions fine-resolution climate surfaces that are needed can be obtained from climate models. In North America, ClimateNA is widely used to generate spatial climate data and provides 12 annual, 16 seasonal, and 48 monthly climate variables for each km grid cell (Hamann et al., 2013). Although climate variables are often used to explain forest site productivity, some disadvantages of using them are (Weiskittel et al., 2011b): they can be imprecise at a local level since they are often derived from latitude, longitude, and elevation; climate varies strongly, yearly but also due to global warming, therefore extreme weather events and climate changes can have more impact on estimated productivity than long-term climate “normals”; climate affects productivity only at larger geographic scales; and, issues arise in modeling due to high multicollinearity between variables and discovering the most influential variables.

Two main approaches in modeling forest productivity using SI are empirical modeling and process-based modeling (Coops et al., 2011; Swenson et al., 2005; Waring et al., 2006), or by combining them in hybrid models when productivity parameters of process-based models are empirically related to environmental variables (Mitsuda, 2016). Process-based models, such as 3-PG, use detailed knowledge of environmental factors and physiological processes to estimate the effects of climate and site factors on many stand attributes (Weiskittel et al., 2011a). Relationship between Gross Primary Production (derived from MODIS and 3-PG process-based models) and SI (from field plots) were found (R^2 of 0.36-0.56) for western US tree species (Weiskittel et al., 2011a), while relationships between GPP (MODIS) and SI with rank correlation coefficient values of 0.61 and 0.26 were obtained for eastern US conifer and hardwood species, respectively (Jiang et al., 2015). Although process-based approaches provide theoretical improvement in integrating climate and biophysical data in a mechanistic manner, research in the western USA underlined no difference in predictions over empirically-based SI when evaluated using a small subset of climatic variables (Weiskittel et al., 2011a). Therefore, due to complexity of process-based models in terms of obtaining data, parameterisation and sometimes lack of correlation with direct measurements (Aertsens et al., 2012b), SI prediction from environmental data is still preferred for estimation of productivity and evaluation of future climate change impacts (Antón-Fernández et al., 2016; Brandl et al., 2014; Jiang et al., 2015; Nothdurft et al., 2012).

A wide variety of empirical statistical approaches have been used for SI modeling based on parametric or non-parametric regression analysis. Multiple Linear Regression (MLR) (Chen et al., 2002; Pinno et al., 2009; Seynave et al., 2005; Watt et al., 2015; etc.), Multiple Nonlinear Regression (MNLR) (Antón-Fernández et al., 2016; Anyomi et al., 2015; Sabatia and Burkhart, 2014; Ung et al., 2001; Wang et al., 2005), and Nonlinear Mixed-effects Model (NLME) (Wang et al., 2007) are widely used parametric methods. Semi-parametric techniques such as Generalized Additive Models (GAM) (Aertsen et al., 2011, 2010; Albert and Schmidt, 2010; Brandl et al., 2014; Wang et al., 2005) and B-spline regression (BR) (Nothdurft et al., 2012); and non-parametric machine learning techniques such as Regression Trees (RT) (Aertsen et al., 2011, 2010; Laamrani et al., 2014; McKenney and Pedlar, 2003; Wang et al., 2005), Random Forests (RF) (Jiang et al., 2015; Sabatia and Burkhart, 2014; Weiskittel et al., 2011a), Boosted Regression Trees (BRT) (Aertsen et al., 2012a, 2011, 2010), Artificial Neural Networks (ANN) (Aertsen et al., 2011, 2010; Wang et al., 2005), and k-Nearest Neighbour (kNN) (Watt et al., 2015) methods may also be used. Some analyses have also used combinations of statistical methods. For example, the Random Forest (RF) method was applied in variable selection prior to using a non-linear parametric model (Sabatia and Burkhart, 2014); a generalized additive model (GAM) was developed for predicting spruce SI in Bavaria while in a second step residuals were modeled using Boosted Regression Trees (BRT) to detect interaction and nonlinearity (Brandl et al., 2014). Relationships between productivity measures and environmental variables were also explored using both multiple linear regression (MLR) and multivariate canonical correlation analysis (Berrill and O'Hara, 2016). In another study, Principal Component Analysis (PCA) was employed to explore environmental conditions and patterns, Partial Least Squares Regression (PLS) was used for variable selection, and then MLR was used for modeling SI (Bravo-Oviedo et al., 2011); while Bueis et al. (2016) used Principal Component Analysis (PCA) followed by Discriminant Analysis. Generally, the advantage of non-parametric methods over others is the capacity to capture complex relationships, interactions and nonlinearity, which is expected in environmental studies.

Furthermore, several studies compare different imputation methods in modeling SI variation in response to environment factors. Four modeling approaches (NLR, GAM, RT, ANN) were applied in lodgepole pine SI prediction in a case study in the Wapiti region in Alberta, with the main findings: preference should be given to GAM in terms of fit statistics; specification of

functional forms and poorer performance were weakness of NLR; ANN failed on bootstrap validation due to too many parameters being selected in the final model; and, although impressive goodness-of-fit was obtained, poor maps produced from the RT model resulted from creation of too few classes (Wang et al., 2005). Aertsen et al. (2011, 2010) evaluated SI modeling techniques (MLR, RT, GAM, BRT, ANN) in two contrasting ecoregions using two different multicriteria decision analyses by assigning a set of criteria and weights, but came up with the similar conclusions in both regions - nonparametric techniques are preferred in modeling SI over traditional MLR due to their ability to detect complexity of site factor impacts, GAM and BRT are the preferred alternatives in terms of predictive performance and ecological interpretability, RT is preferred when user-friendliness is more important, whereas ANN is poor in most cases. On the other hand, parametric approaches were found to be more precise in comparison of MLR and k-NN methods for building SI prediction models from different data sources in New Zealand's *Pinus radiata* forests. However, results also suggested that the best statistical method may depend on particular region/landscape characteristics and local productivity~environment relationships (Watt et al., 2015). In predicting SI of loblolly pine plantations, the RF model provided better fit than MNLR when using biophysical variables (Sabatia and Burkhart, 2014), while a partial least-squares regression model better addressed multicollinearity than MLR when using soil properties (Subedi and Fox, 2016).

Additionally, to overcome the issue of prediction of unreliable values when used beyond the range of fitting data in standard regression analysis techniques, Antón-Fernández et al. (2016) use a multiplicative potential modifier functional form to develop climate based SI models. Since SI is often modeled based on large datasets with independent samples, which are usually not clustered or do not include repeated measures, mixed modeling approaches are not required. However, sampling environmental and productivity variables might be associated with the issue that observations are more similar to other nearby observations. Hence spatial regression modeling approaches such as Sequential Autoregressive (SAR) is proposed to account for potential spatial autocorrelation in the study of potential productivity, based on SI, of Pacific Northwest forests using climate variables (Latta et al., 2009). They evaluated four methods (Nearest Neighbour, Thin Plate Spline function, MLR, and SAR) and found that the SAR model outperformed all other models in terms of variance explained and in terms of bias, while at the same time dealing with spatial autocorrelation issues.

Random Forest is an artificial learning method which has the advantages of accuracy and efficiency in dealing with many variables, where still being easy to apply. RF has become popular in environmental studies, especially in the application of climate and remotely sensed variables in spatial predictions, such as prediction of: Site Index and its changes under expected climate changes (Jiang et al., 2015; Weiskittel et al., 2011a); differences in SI obtained from field-based forest inventory and LiDAR (Tompalski et al., 2015a); tree species distribution and relative basal area (Dunckel et al., 2015); tree height growth (Travis Swaim et al., 2016); forest aboveground live carbon density (Zald et al., 2016); fire behaviour and forest structure (Kane et al., 2015); future climatic niche (Wang et al., 2015, 2012a); suitable tree species habitat under climate change (Gray and Hamann, 2013); and, species richness (Zellweger et al., 2015).

In summary, geocentrally-based species-specific Site Index models are independent from stand age, density and structure, usually have satisfactory prediction power, and therefore, provide tools that have practical application and that can effectively inform forest management. An additional advantage to using biophysical parameters to predict forest productivity is the possibility to use remotely sensed data.

1.3. Remote sensing in estimating Site Index

Growing availability and resolution of spatial environmental information could increase accuracy in the predictive models of forest site productivity (Bontemps and Bouriaud, 2014). Progress is being made in developing spatialized topographic indices obtained from Digital Elevation Models (DEM) and vegetation structure derived from ALS (Airborne Laser Scanning) technologies, but also further progress should come from extrapolating climate information based on fine-resolution DEM (Bontemps and Bouriaud, 2014). Another advance is in passive optical remote sensing technologies such as Landsat imagery with long temporal image sequences which provide potential indices of landscape productivity using maximum pre-harvest normalized difference vegetation index (NDVI) (Nijland et al., 2015), or prediction of SI from NDVI (Ma et al., 2006).

Light Detection and Ranging (LiDAR) is an active remote sensing technology often used for Airborne Laser Scanning (ALS). It employs a laser to measure distance to a target, then point

clouds of ground or non-ground returns projected in a 3D coordinate system are used to generate Digital Elevation Models (DEM), Canopy Height Models (CHM), individual tree-based and plot-level forest attributes, as well as resource value indicators at the landscape level (Wulder et al., 2008). The advantage of ALS technology over optical imagery is its three-dimensional representation of both topography and vegetation components at high spatial resolution. Also, SI models developed from remotely sensed environmental data allow mapping of forest productivity and avoid issues arising from interpolation of spatially discrete field inventory data. Therefore, ALS enables better use of biophysical and vegetation variables in estimating forest productivity.

Bare-earth Digital Elevation Models (DEM), in addition to computing simple fine resolution topographic parameters, can also be used to calculate hydrological and soil related topographic indices, such as Topographic Position Index (TPI), Depth-To-Water (DTW), Topographic wetness index (TWI), Integrated moisture index (IMI), Topographic relative moisture index (TRMI). Several studies have tested remotely sensed data in prediction of SI. DEM was used to derive IMI (which integrates hillshade, flow accumulation area, curvature, and total water-holding capacity) which explained 64% of the variation in white oak SI (Iverson et al., 1997). In addition to soil properties and chemistry, Berrill and O'Hara (2015) tested DEM produced topographic indices, including TRMI and TOPEX, and found them successful in predicting coast redwood SI. Four different DEM resolutions (100-, 50-, 25-, 12.5-m) were compared for *Cryptomeria japonica* plantation SI prediction with TWI and Solar Radiation Index explaining 65% of the variation in SI (Mitsuda et al., 2007). While in boreal plane forests topographic variables derived from DEM at four scales (20-, 15-, 10-, 5-m) accounted for between 25 and 31% of the variation in black spruce SI (Laamrani et al., 2014).

Furthermore, ALS provides accurate estimates of various stand attributes potentially useful in forest productivity assessment, including height, LAI, and biomass. For instance, Silva et al. (2016) tested 31 metrics calculated from LiDAR's first returns for stem volume (field measurements) modeling of Eucalyptus hybrid clones in Brazil and found that the coefficient of variation for height and 99th percentile of height, explained 84% of the variation. Also, aboveground live biomass (carbon density) was predicted in the western Oregon Cascades forests with adjusted R^2 of 0.75 for a regression model using LiDAR's H95 and INDEX3 as

predictors (Zald et al., 2016). Based on data from a large number of field plots in *Pinus radiata* plantations on the North Island of New Zealand, Watt et al. (2015) evaluated three different data sources (environmental surfaces, LiDAR, RapidEye satellite imagery) as predictors of SI and found that the model with the highest precision included LiDAR H₉₉ and stand age (R² of 0.88 and RMSE of 1.38 m). However, although ALS enables very accurate measures of stand height Tompalski et al. (2015a) demonstrated that ALS data alone cannot be used to estimate Site Index without information on age and species. They computed dominant stand height from ALS point clouds as the weighted mean of maximum height with the average non-ground return count used as weight, and compared their results to height based on SI and age from forest inventory. SI in the forest inventory was found to be underestimated 3.5 m on average with the largest difference in multi-species stands. Further, in addition to dominant height data obtained from ALS Tompalski et al. (2015b) utilized Landsat time series images and disturbance detection algorithms to determine time-since-disturbance (proxy for stand age), which might provide a useful approach for estimating productivity based only on remote sensing. However, they found that this species-independent model overestimated productivity by 0.70 m (RMSE=5.55 m) compared to forest inventory estimates.

1.4. Wet Areas Mapping (WAM) explains site productivity

Topographic Depth-To-Water (DTW) index is a DEM-based hydrologically-related indicator derived following the process of Wet Areas Mapping developed and first applied in New Brunswick (Murphy et al., 2007). Wet Areas Mapping (WAM) in Alberta involves locating and delineating wet and dry areas, flow channels, stream crossings, and associated wetlands from ephemeral to higher order streams based on a 1 metre bare ground DEM derived from LiDAR data (White et al., 2012). It relies on the assumption that topography controls water flow. According to Murphy et al. (2009) the WAM process includes removing artificial depressions in the DEM, determining flow direction, and producing stream accumulation networks based on predetermined flow channel initialization. Flow Initiation Area (catchment area) is an arbitrary threshold which represents the upslope contributing area required to initiate a stream channel. FIA of 4 ha worked well across varying terrains and purposes while representing end-of-summer flow and soil conditions (White et al., 2012). Decreasing the FIA results in extending the stream

network further up-slope and lower Depth-To-Water (DTW) index values. Once the flow channel network is determined the DTW index is computed by assigning a value of zero to flow channel cells and all other surface water features, and then applying the algorithm which calculates the sum of cell slopes along the least slope path to the nearest surface water feature for each raster cell. As a result DTW integrates both slope and distance which means landscape points further from open water bodies, both horizontally and vertically, have higher DTW values and drier soils (Murphy et al., 2009). Flow Accumulation (FA) index is also calculated as the upslope drainage area contributing to the point of interest expressed as number of cells (m²).

Several studies have demonstrated relationships between DTW and soil properties. Murphy et al. (2009) found that DTW could delineate patterns of soil moisture (hydric to subhydric soil moisture regimes) in central Alberta, and suggested usage of DTW for landscapes and climate where belowground flow patterns are driven by surface topography. In addition, detailed physical and chemical soil properties, as well as soil, vegetation and drainage type are found to be subject to topographic controls, and DTW with the smallest FIA (4 ha) effectively estimated these in Foothills Natural Region of Alberta (Murphy et al., 2011). In the Swedish boreal landscape DTW has proven to be the best soil wetness (similar to Alberta's SMR) predictor among several DEM-based topographic indices, and was insensitive to DEM scale but the optimal FIA threshold varied by landform, from 1-2 ha on slowly permeable till deposits to 8-16 ha on coarse-textured deposits where water would drain quickly (Ågren et al., 2014). These studies also demonstrate that DTW works better than TWI in predicting soil and vegetation characteristics. Soil moisture regime, drainage class, and depth-to-mottles were strongly related with DTW at a 2 ha catchment area in young aspen dominated boreal plain forests (Oltean et al., 2016). Hiltz et al. (2012) tested DTW against vegetation index, which is based on moisture-regime classes from xeric (0) to hydric (8), and obtained coefficients of determination of 0.68-0.77 from two boreal sites in Alberta. DTW also appeared to be an important predictor of distribution of 14 understory plant species related to grizzly bear food supply (Nijland et al., 2014). DTW was related to canopy composition in boreal mixedwood forests at the EMEND research site and was considered to be a good moisture and productivity indicator, and can be used to predict forest cover type after variable retention harvesting (Nijland et al., 2015).

Overall, since numerous studies have already confirmed relationships between DTW and various soil and vegetation attributes, which are in turn potential direct productivity indicators, it is ecologically reasonable to expect that DTW could significantly explain SI in forests of central Alberta where soil moisture is driven by topography.

1.5. Productivity of western boreal forests

Forest covers about 348 million ha or half of the landscape in Canada, and represent 9% of the world's forests and 24% of the global boreal forests. Spruce, poplar and pine are the major species groups in terms of wood volume in Canada (Natural Resources Canada, 2015). Western Canadian boreal mixedwood and foothills forests are dominated by trembling aspen, white spruce, and lodgepole pine on mesic sites, and sometimes with mixtures of balsam poplar, white birch, tamarack, and black spruce in wetter conditions. Ecosystem classification in Alberta relies on relationships between topography, soil types, and trees and understory species distribution (Figure 1.1; Beckingham et al., 1996) with climate, soil moisture and nutrient regimes serving as important factors influencing species composition and productivity.

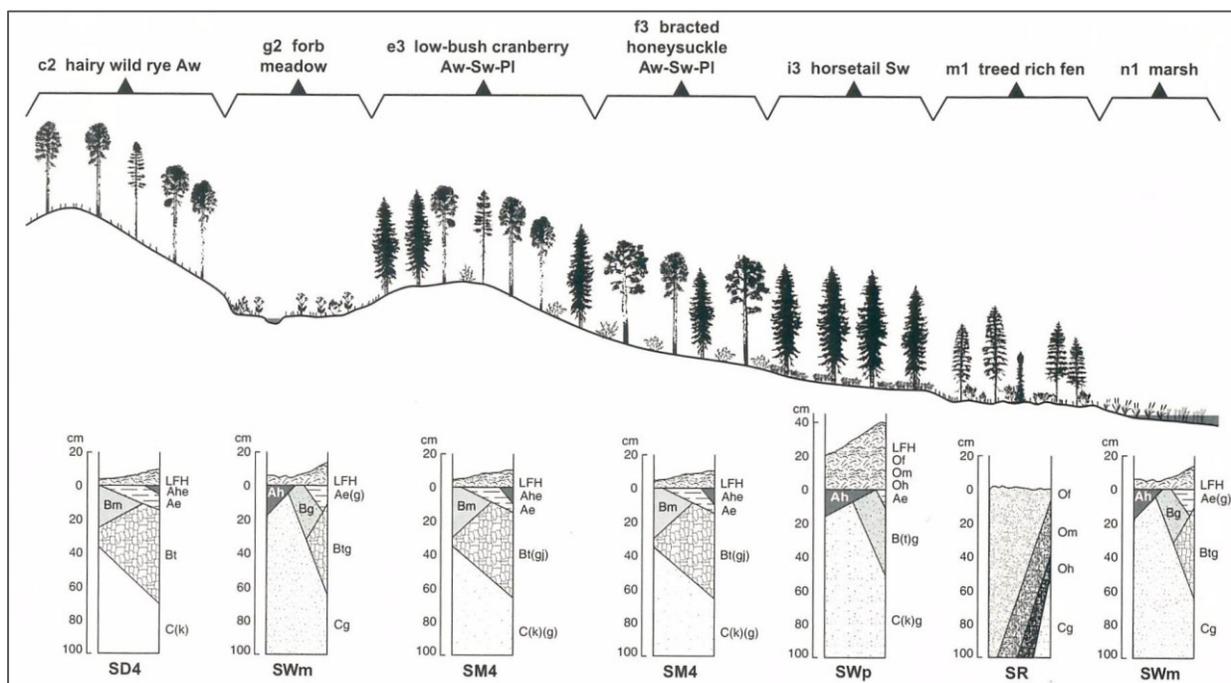


Figure 1.1. Schematic cross section in Lower Foothills NSR represented by ecosite phase and soil profiles (taken from Beckingham et al., 1996)

The wide occurrence of trembling aspen (*Populus tremuloides* Michx.) among tree species in North America reflects its adaptability to many site factors since it grows in a range of climates, habitats, and mixtures. Recently aspen has expanded its range into higher elevations in the Rocky Mountains as a result of interactions between management practices and climate warming (Landhausser et al., 2010). It is also well adapted to disturbance regimes in the boreal forest owing to its ability to regenerate from root suckers. According to Burns and Honkala (1990), aspen is nutrient demanding, grows best on well drained and aerated, loamy soils, and in prairie provinces on cooler and moister north and east slopes and depressions. Aspen growth, however, might be limited due to drainage, when the water table is shallower than 0.6 m or deeper than 2.5 m, or on sandy or heavy clay soils, causing poor soil moisture and/or aeration conditions. Aspen is an early-successional, pioneer, shade intolerant species that regenerates quickly from root suckers following disturbance and has rapid juvenile growth, with MAI culminating between ages 50 and 70 in Alberta, while decay and canopy break-up usually start at 80-100 years. Stand-replacing disturbances are major drivers of classical boreal forest dynamics, however there is no difference in height growth between stands of different origins (fire, clearcut, and tree-fall gap) (Paré et al., 2001). Also, growth of aspen in western Canada is very sensitive to defoliation by insects and drought with possible multi-year lags in growth response (Hogg et al., 2013, 2002). Chen et al. (1998) examined aspen SI in northern B.C. and found that the most productive sites were associated with lower altitude, warm aspects, lower slopes, nutrient-rich and moist soil conditions, thicker forest floor, high nitrogen and less acid concentration. However, in interior B.C. these relations differ between biogeoclimatic zones due to specific local site conditions, such as elevation and aspect, while SI does generally increase with increases in both SMR and SNR (Chen et al., 2002). In a study in mixed aspen-white spruce boreal forests in Lower Foothills of west-central Alberta using LAI as a surrogate for productivity, Little (2001) found that LAI declined moving towards upslope positions, with many soil properties correlating with slope position, while forest floor N, forest floor bulk density, B-horizon silt content and permanent wilting contributed to the best model ($R^2=0.70$). Stand overstory and soil physical and chemical variables accounted for 16% and 62% of aspen SI variation in studies in Saskatchewan and Quebec respectively, but grouping by parent material greatly improved models suggesting strong impacts of parent material on environmental factors (Pinno and Bélanger, 2011; Pinno et al., 2009). Moreover, aspen SI across Quebec was explained ($R^2=0.69$) by degree-days above

5°C, aridity index, and soil water-holding capacity variables, with negligible improvement ($R^2=0.72$) if Shannon index added (Ung et al., 2001). Along an east-west Canadian transect, aspen productivity declines with increasing moisture deficit (aridity index), increases with soil humus type, but stand development (stand age, aspen basal area proportion, Shannon index) dominates variability in productivity with a total of 53% variance explained by all factors together (Anyomi et al., 2015).

White spruce (*Picea glauca* (Moench) Voss) is a widespread and economically important boreal conifer species. Ecologically, it is described as a "plastic" species since it tolerates a diverse range of climate and soil conditions, and is limited only by stagnant water (Burns and Honkala, 1990). Further, it is nutrient demanding but for good yield production requires well balanced water-aeration conditions. In boreal mixedwoods white spruce is not well adapted to wildfire hence, in contrast to aspen, it is considered to be a late-successional shade tolerant species which usually regenerates under the canopy and grows in the understory of aspen stands during the first several decades of development. However, multiple successional pathways exist in the boreal forest so pure even-aged white spruce stands can establish in cases where there is a good seed source and exposed mineral soil seedbed after severe fire (Lieffers et al., 1996). Spruce height growth is affected by overtopping aspen (Filipescu and Comeau, 2007) and therefore might not be reliable as a productivity indicator. Due to evidence of acceleration of height growth after overstory removal Osika et al. (2013) provided SI curve adjustments for white spruce for time after release. A few studies have identified particular site factors important for white spruce productivity in natural stands. Relationships between white spruce SI and soil physical and chemical properties, understory species indicators of soil moisture and nutrients, and foliar nutrients, were investigated in sub-boreal B.C. and revealed that all three groups of predictors were closely related with SI which was best predicted ($R^2=0.83$) using depth of major rooting zone, total N, soil moisture plant indicators, and foliar N, P and S (Wang, 1995). Thereafter, the same sample plots were used but different soil and physiographic variables selected as synoptic climate, soil moisture, nutrient and aeration regimes, and SMR were found to be the strongest predictors explaining 81% of variance in SI. When SNR and SAR were added model performance increased to 90%, while climate represented by zonal vegetation was not significant (Wang and Klinka, 1996). However, validation of this model, which included SMR and SNR, on an independent dataset showed an increase in RMSE from 1.2 m to 3.2 m (Kayahara et al.,

1998). Similarly, effects of climate on white spruce SI in interior B.C. were found to be weak ($R^2=0.08$), with the best model including mean temperature of the warmest month, indicating that spruce seems to be the less sensitive to climate than Douglas-fir ($R^2=0.64$) and lodgepole pine ($R^2=0.27$) (Nigh et al., 2004). These weak correlations are in accordance with findings from eastern Canada that for SI estimation for shade tolerant species, in addition to climate and soil variables, a measure of successional stage needs to be added (Ung et al., 2001). Due to difficulties in white spruce SI prediction, species conversion models might be an option, but while SI species conversion models between white spruce and aspen in north-eastern B.C. were more or less successful with R-squared of 0.36 (Nigh, 2002), a weak relationship (0.09 R-squared) was found in Alberta (Hostin and Titus, 1996).

Lodgepole pine (*Pinus contorta* Dougl. var. *latifolia* Engelm.) is a widely distributed and commercially valuable species in the foothills and mountains of western Canada. However, future climate change conditions appear to be increasingly unfavourable to lodgepole pine and dramatically impact replacement by other species (Coops and Waring, 2011). It is a fire adapted species with serotinous cones, but Mountain pine beetle outbreak is another severe risk for lodgepole pine forests. At its eastern boundary it hybridizes with Jack pine producing trees with intermediate morphological attributes in transition areas. It is a pioneer following fire, very intolerant to shade and competition and has relatively vigorous growth in early development which determine l. pine as an early-successional species. It has probably the widest ecological amplitude among North American conifers and it grows from dry to very wet sites and sites with fluctuating water table, in association with numerous other species (Burns and Honkala, 1990). L. pine tolerates drought and poor soil nutrient conditions but productivity significantly improves with increasing fertility, and it often grows best on sites with mesic moisture conditions (Comeau and Kimmins, 1989). In terms of topography, it grows well on gentle slopes and in basins, but good productivity can occur on rocky soil, steep slopes and ridges (Burns and Honkala, 1990).

Several studies have explored l. pine productivity in the western boreal forests by linking SI and biophysical site parameters. Firstly, Fries et al. (2000) found that lodgepole pine SI responds to increasing temperature and fertility with a correlation coefficient of up to 0.70 in experimental plantations in western Canada and with even stronger responses to temperature in northern Scandinavia. In interior B.C. climate sensitive models were developed where SI increases as

mean annual temperature and precipitation increase but with a stronger impact of temperature (Nigh et al., 2004). According to Kayahara et al. (1998), the SI prediction model based on SMR and SNR from Wang (1992) in interior B.C., originally unbiased with R-squared of 84%, did not perform well after independent dataset validation. Models predicting lodgepole pine SI were calibrated using physical and chemical properties and understory vegetation from 42 stands in south-western Alberta and explained up to 67% of the variation in SI (Szwaluk and Strong, 2003). They also found that humus variables were the best predictors, then plant indicator species, and soil variables. In addition, humus type, SNR and sand content were positively, while humus thickness and %silt negatively correlated with SI. In west-central Alberta up to 75% of SI variation was explained using climate, topography and soil sand fraction (Wang et al., 2005). In province wide studies in Alberta, Wang et al. (2004) evaluated synoptic measures of climate, soil moisture and nutrients to estimate SI with the best model accounting for 41% of the variation in SI including elevation, SNR, latitude, longitude, and natural subregions. Monserud et al. (2006) used climatic data to fit SI models using data from 1145 plots with the resulting models, using only measures of heat as predictors, explaining about one quarter of variation in l. pine SI. Later, Monserud et al. (2008) applied this climate sensitive SI model to evaluate expected climate change impact and predicted steadily increasing SI by 3 m for each 30-year period but with large reduction of 42% in species distribution by the end of 2080.

1.6. Objectives, hypothesis

Covering over 33 million ha Alberta has one of the largest Airborne Laser Scanning (ALS) datasets worldwide (Coops et al., 2016). In addition, the Wet Areas Mapping initiative provides additional hydrological information, and these maps provide valuable information in several areas of natural resource planning, such as forestry, parks and recreation, oil and gas sector, and land reclamation (White et al., 2012), and are potentially useful for estimating Site Index. These highly accurate data, with a resolution of 1 m, provide an excellent source of information for efficient and cost-effective prediction of potential productivity. In addition, Site Index prediction based on physical indicators of the environment (biophysical SI) can help overcome some challenges facing the conventional Site Index method (dendrocentric SI). The main objectives of this study are: (1) to explore the use of remotely sensed environmental parameters generated

from WAM and LiDAR in estimating Site Index; (2) to identify the optimal flow initiation area in terms of computing DTW index; (3) to evaluate the use of remotely sensed data in predicting ecological site and soil characteristics; (4) to evaluate different modeling approaches suitable for biophysical SI prediction; and, (5) to map spatial species-specific site productivity variations based on best selected models from remotely sensed information. Many studies have found soil moisture and vegetation characteristics in western boreal forests to be highly driven by topography hence in this study I hypothesized that forest productivity is also subject to topographic controls. In addition, due to differences in species responses to site factors, different relationships are expected for different species.

2. MATERIALS AND METHODS

2.1. Study site selection and description

This study was conducted in the area of the 1968 Vega fire (Martens Hills FMA area in Forest Management Unit S17) near Slave Lake in central Alberta, Canada (Figure 2.1). According to Beckingham et al. (1996), the study area is characterized by mixed forests of trembling aspen, white spruce, and lodgepole pine, with an intermix of balsam poplar, white birch, black spruce, and tamarack on moist and wet sites. Topography is generally rolling, soils are predominantly Luvisols and Brunisols with their gleyed variants and organic soils in poorly drained conditions, and climate in this region is generally cooler in the summer and warmer in winter than in boreal forests. Mean annual temperature is 2.4°C, mean temperature during the vegetation period (May-September) is 12.0°C, 158 frost-free days, and mean annual precipitation is 600 mm with 450 mm falling during the vegetation period (extracted from ClimateNA software for the period 1981-2010; Wang et al., 2012b). Most of the study area is located in the Lower Foothills Natural Subregion. However, since the natural subregions of Alberta are separated elevationally (Beckingham et al., 1996), Lower Foothills NSR is actually a transition between aspen-white spruce dominated boreal mixedwood forest (Boreal Forest Natural Region) and lodgepole pine-dominated upper foothills and subalpine forests (Foothills Natural Region) (Beckingham et al., 1996). Since the study area is relatively small in size (15 km in radius), we can say in general that the study site represents Lower Foothills Natural Subregion of Alberta (transition zone between boreal mixedwoods and l. pine foothills forests).

This area has several advantages for a study designed to examine relationships between ecological site factors and Site Index. First, the area has a limited range of stand ages present, with most stands having originated after 1968 fire, and close to SI reference age (50 years). All stands have regenerated naturally with no evidence of management activities. Rolling topography provides a good range of site conditions and tree species. A case study in a geographically restricted area like this, where climate is relatively constant across sites, provides some overall control of macro-climate effects and allows me to focus investigation on effects of local topographic and soil factors. In addition, availability of ecosite mapping helped with

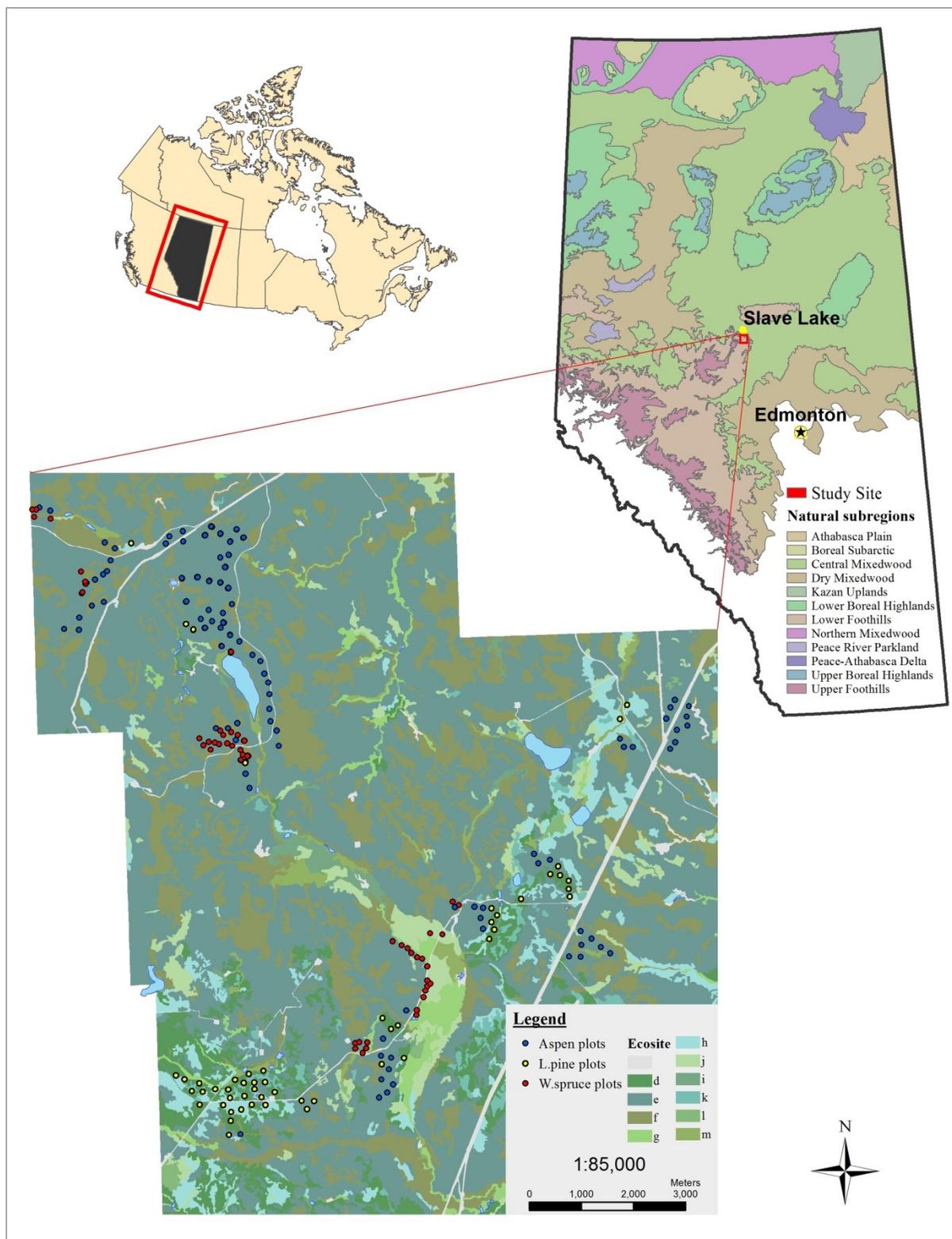


Figure 2.1. Study site location and Natural subregions (forested) of Alberta (upper right map). Sample plot locations and ecosite (Lower Foothills NSR) map of the study site (lower map).

sample site selection. Finally, availability of remote sensing data, and ability to sample three major commercial tree species in the region (trembling aspen, lodgepole pine, and white spruce) make this study of practical value.

2.2. Sampling design, plot selection and tree selection criteria

Site Index was selected as the measure of forest productivity, therefore, according to the main assumption for this approach, trees that have not been suppressed are necessary to determine SI. According to Chen and Popadiouk (2002), aspen and lodgepole pine are shade intolerant and have rapid juvenile growth, regenerate right after the stand-initiating disturbance and tend to dominate in the main canopy layer during the first several decades of boreal mixedwood stand development process (during stand initiation and stem exclusion successional stages). Based on this we can generally assume that dominant trees in these stands at the study site have grown free of competition. However, white spruce usually regenerates together with or in the understory of aspen or pine, and due to their slower growth rates remain in the understory for 70 or more years (Lieffers et al 2003). However, there are multiple successional pathways depending on disturbances, neighbour effects, and stand condition (Bergeron et al., 2014; Chen and Popadiouk, 2002; Lieffers et al., 2003, 1996; Macdonald et al., 2010), hence sometimes white spruce regenerates and grows without suppression. In my study area I found groups of dominant spruce trees. These patches of spruce were usually small in size, older than surrounding deciduous trees, and had no evidence of suppression. These patches presumably originated from a previous fire and remained after the most recent wildfire ("green islands"). Checking the ring width close to pith (early ages) I generally found no evidence of suppression and could therefore assume that these spruce trees satisfy the assumptions for Site Index estimation.

Plot selection and tree selection criteria recommended in SIBEC Sampling and Data Standards (2009) were used. The SIBEC (Site Index/Biogeoclimatic Ecosystem Classification) project was developed by the British Columbia Ministry of Forests to estimate average growth potential of tree species in forested site series, according to biogeoclimatic ecosystem classification in B.C., by using mean Site Index for the main tree species (Mah and Nigh, 2015). Based on these standards, plot selection criteria were: even-aged stand, moderate stand density; no management activities - natural stand conditions (e.g., no partial cuttings, fertilization, site preparation,

pruning); not affected by major disturbances; appropriate age class (in general between 20 and 120 but closer to site index reference age of 50 years the better); and contain suitable top height trees of target species. Selection criteria for top height trees from the same standards were used: largest diameter dominant or co-dominant tree per 100 m²; vigorous with a full crown; straight, disease-free, undamaged stem (minor damage is allowed but that not significantly affects tree growth); free of suppression; not a wolf, open-grown, or veteran tree. This sampling approach allows us to ensure that stand is developed in unmanaged conditions and to assume that height growth is affected only by potential site productivity.

A transect-based sampling design was applied for aspen and lodgepole pine. Firstly, stands were selected for sampling based on: 1) dominant species (ecosite map was used); and 2) the range of environmental conditions (depth-to-water raster was used) to capture a gradient of forest productivity conducive to statistical analysis. Plot centres were located randomly along predetermined transects within the selected blocks to ensure sampling of the full range of DTW conditions that were available for each species. The center of measurement plots was then adjusted from the point on the transect in order to ensure that site conditions were uniform within and around each plot, as required to meet plot and tree selection criteria specified in the SIBEC sampling standards (2009). A buffer area of at least 50 m was left between sample plots and any disturbance or object (e.g. wellsite, road, cutline, etc) to avoid edge effects on trees growth. Neighbouring plots were located a minimum of 150-200 m apart and were situated in different site conditions (different slope, slope position, aspect, depth-to-water, SMR, etc.) in order to minimize correlations due to adjacency and ensure independence of plots.

Due to restricted availability of white spruce stands appropriate for Site Index sampling all possible locations where we were able to select three unsuppressed top height trees were considered regardless of the site conditions but while maintaining a minimum distance of 50 m between other sample plots and artificial objects.

Sample size depended on the range of site factors intended to be used as covariates in modeling so the goal was to sample approximately 100 aspen temporary sample plots, as well as 50 sample plots for lodgepole pine and white spruce.

2.3. Data collection

Data were collected within 300 m² circular sample plots during 2014. A total of 201 temporary sample plots were established. In each plot two types of data were collected: ground-based and remotely sensed. Field (ground-based) sampling in each sample plot consisted of determination of ecological data (topography parameters, soil properties, and stand density), measurements of necessary parameters to determine Site Index (top height tree sampling), and determination of Global Positioning System (GPS) coordinates for the plot centre and sampled top height trees. In each plot a soil pit was excavated at plot centre to represent average soil conditions in the plot. Plot centres were flagged at the nearest tree. Major soil characteristics assessed included humus form, organic matter thickness, soil texture, coarse fragment content, soil drainage class, soil moisture regime, and soil nutrient regime followed by (Beckingham et al., 1996). In addition to soil properties, basic topographic factors determined at the plot centres including topographic position, slope measured with Vertex, aspect with compass, and elevation with Trimble GPS receiver.

Three top height trees were selected in each plot for determination of Site Index for each target species. Measurements of each top height tree included diameter at breast height (DBH) and total height, while an increment borer was used to take "bark to-pith" increment cores at breast height (1.3 m) for aging. DBH of each tree was measured to the nearest 0.1 cm at 1.3 m above ground, while height was measured using a Vertex III hypsometer with an accuracy of 0.1 m. As a measure of stand density, Basal area (BA) was determined for each species by the method of angle-counting of stems using a wedge prism with BAF (Basal Area Factor) 3.

High precision GPS equipment was used to ensure that ground plots accurately corresponded to remotely sensed data sample plots. A handheld Trimble GeoExplorer 6000 series GPS receiver was used to record locations as GPS waypoints with the target of 50 hits for each waypoint to get better precision. To improve precision, post-processing of field recorded GPS point locations was performed by using Slave Lake and Barrhead Cansel base stations to apply differential corrections and to export it in a format suitable for ArcGIS using GPS Pathfinder Office software. Average horizontal precision of mean GPS recordings after corrections for the 201 sample plot centres was 1.28 m, with the majority (75%) of plots under 1.5 m precision.

The LiDAR derived Digital Elevation Model (DEM) and Wet Areas Mapping (WAM) were used to provide remote sensing data. Both datasets were acquired from Alberta Ministry of Agriculture and Forestry. The highly accurate bare ground DEM was generated from LiDAR point clouds obtained from flights during August-September 2008, with a mean return density of 1.6-2.7 hits/m², 45 cm average xy point error, and vertical accuracy of 30 cm. WAM raster models of the cartographic Depth-To-Water (DTW) index and Flow accumulation (FA) index which were applied in analysis were prepared by Jae Ogilvie utilizing WAM algorithms in use during October of 2015. We used DTW rasters created from 0.5 ha, 1 ha, 2 ha, 4 ha, 6 ha, and 10 ha flow initiation areas. Pixel resolution for all raster datasets used in the study is 1x1 m.

Determination of breast height age was done in the lab. Increment cores were glued to a grooved wooden board, and sanded with progressively finer sandpaper until the ring boundaries were clearly visible. Marker, unusually narrow, rings noticed consistently between different trees were used for crossdating (Speer, 2010). Crossdating was also used to determine missing rings, which are usually caused (in aspen) by severe forest tent caterpillar defoliation. Ring counting was done visually (manually) under a binocular microscope. Since sampling occurred during the growing season (May-August 2014 with a break in late June-early July) the newest annual ring (actual growth) was not included for the sampling conducted in May and June, while the newest ring was included for sampling during July and August (since height growth of conifers is completed in mid July and aspen will also have completed most of its height growth by mid-July). Partial inner rings (in the pith) were counted as one year. A pith indicator was used to estimate total breast height age for the cores that missed the pith based on the average size of the inner rings on the core (Speer, 2010).

2.4. Variable estimation

Site Index (SI) is estimated using species specific top height sub-models (height - age curves) of the GYPSY (Growth and Yield Projection System) forest growth and yield model (Huang et al., 2009) which is currently in use in Alberta. GYPSY's SI curves for the investigated species are as follows:

[Eq. 1] Aspen -
$$H_{top} = SI_t \times \left(\frac{1 + \exp(b_1 + b_2 \sqrt{\ln(1+50)} + b_3 [\ln(SI_t)]^2 + b_4 \sqrt{50})}{1 + \exp(b_1 + b_2 \sqrt{\ln(1+totalage)} + b_3 [\ln(SI_t)]^2 + b_4 \sqrt{50})} \right);$$

$$[\text{Eq. 2}] \quad \text{Lodgepole pine - } H_{top} = SI_t \times \left(\frac{1 + \exp(b_1 + b_2 \sqrt{\ln(1+50)} + b_3 \ln(SI_t) + b_4 \sqrt{50})}{1 + \exp(b_1 + b_2 \sqrt{\ln(1+totage)} + b_3 \ln(SI_t) + b_4 \sqrt{50})} \right);$$

$$[\text{Eq. 3}] \quad \text{White spruce - } H_{top} = SI_t \times \left(\frac{1 + \exp(b_1 + b_2 \sqrt{\ln(1+50^2)} + b_3 [\ln(SI_t)]^2 + b_4 \sqrt{50})}{1 + \exp(b_1 + b_2 \sqrt{\ln(1+totage^2)} + b_3 [\ln(SI_t)]^2 + b_4 \sqrt{50})} \right);$$

where H_{top} (m) is average height of the 100 largest DBH trees per ha, SI_t (m) is Site Index (top height) at 50 years total age, $totage$ (years) is total age from the point of germination, and b_1 , b_2 , b_3 , and b_4 are species specific model parameters (Table 2.1).

Table 2.1. Regression coefficients for the GYPSY top height models (Huang et al., 2009).

	Aspen	Lodgepole pine	White spruce
b_1	9.908888	12.84571	12.14943
b_2	-3.92451	-5.73936	-3.77051
b_3	-0.32778	-0.91312	-0.28534
b_4	0.134376	0.150668	0.165483

Given that GYPSY requires total age, while I measured breast height age, it was necessary to estimate the number of years for each trees to grow to breast height. For this I used GYPSY's generalized program that predicts Site Index from total age and height as initial parameters, and also provides years needed for a tree to reach breast height (Y2BH) from the point of germination (Appendix 1. in Huang et al. (2009)). To do this, I first computed total age by adding an estimate of Y2BH for the given site and species to the counted breast height age, then ran the program. If the Y2BH output (number of full years - rounddown) matched with the assumed value it was accepted, otherwise I used Y2BH output (number of full years - rounddown) as a new assumption and repeated the procedure. SAS software ver. 9.4 for windows (SAS Institute Inc, 2013) was employed in this procedure for age correction. Fire history was used to determine maximum age for aspen and lodgepole pine. When total age exceeded the maximum age after the recent fire, total age was corrected to the fire age assuming that errors resulted from application of SI curves during young ages.

For each selected top height tree, Site Index calculation used GYPSY's generalized program for predicting SI from top height measurements and total age (Appendix 1. in Huang et al. (2009)).

Site Index at the plot level was calculated as the average of the three estimates of SI on the plot, except for one spruce plot where only two acceptable top height trees were sampled (one additional tree was rejected due to indications of suppression when the core was examined in the lab). In addition to TH tree selection criteria, by sampling and averaging three TH trees at the plot level we control for within-site variation to get the SI value that best reflects site productivity at the plot level.

Determination of soil and topographic variables followed methods described in the Field Guide to ecosites of West-central Alberta (Beckingham et al., 1996) and is similar to the approach currently in use in forest inventory in Alberta and in SIBEC. Soil characteristics, considered as direct indicators of forest productivity, have strong influences on tree growth since they control availability of moisture and nutrients. Classification of soil characteristics consists of five levels for the humus form, 14 levels for soil texture, four levels for effective soil texture, seven levels for drainage, nine levels for soil moisture regime, five levels for soil nutrient regime, while topographic position determination is slightly modified to six levels. Course fragment content (particles >2 mm in size) was estimated in %. Average surface organic matter thickness was measured to the nearest cm. Slope was determined in %, while aspect was measured as degrees of the maximum slope direction.

All digital terrain covariates were obtained from the 1x1 m resolution bare-ground digital elevation model (DEM) and Wet Areas Mapping indexes (DTW and FA). Built-in surface tools (slope, curvature, aspect) in the spatial analyst toolbox in ArcGIS 10.1 were used to create corresponding surfaces. Beers aspect, Slope Position Index, and Topographic Convergence (Wetness) Index rasters were generated using tools in the Topography Toolbox for ArcGIS 10.1 (Dilts, 2015). Slope indicates the rate of elevation change (percent rise) in the direction of maximum slope. Surface curvature, or "the slope-of-the-slope", is calculated in ArcGIS as three indices: profile, planform, and mean curvature. Profile curvature explains how the surface is curved in the direction of the maximum slope and therefore reflects acceleration or deceleration of the flow across the cell; planform curvature is perpendicular to the direction of maximum slope and indicates convergence and divergence of the flow; mean curvature combines the previous two to identify local high or low points (Jenness, 2007). A curvature value of zero

represents flat terrain, while positive and negative values represents upwardly concave or convex surfaces, respectively.

Aspect refers to compass direction of the downslope but due to its polar nature (0-360° values), and issues related with averaging it, untransformed aspect is a poor variable for quantitative analysis (McCune and Keon, 2002) such as modeling variation in SI by environmental parameters. Instead I calculated Aspect Index. Aspect Index provides a transformation (ie. it "folds") of aspect to index values ranging between 0 and 180 so the coolest site has a value of zero and the warmest site has value of 180. Aspect Index was calculated in two ways following McCune and Keon (2002): folded aspect for incident radiation on the north-south line (so that N becomes zero and S becomes 180) using the equation $\text{Aspect Index} = 180 - |\text{aspect} - 180|$; and folded aspect for heat load using a northeast-southwest line (so that NE becomes zero and SW becomes 180) using the equation $\text{Aspect Index} = |180 - |\text{aspect} - 225||$.

Topographic Wetness Index (TWI) is a widely used topographically based soil wetness model and represents the natural logarithm of the ratio of upslope flow accumulation area and slope at the cell. Consequently, higher TWI values indicate accumulated moisture conditions and low slope, and low TWI means steep slope and dry sites. The algorithm $\text{TWI} = \ln(\text{flow accum} + 1) / (\tan(((\text{slope deg})^{3.141593} / 180))$ was used together with a conversion factor of 10000 to standardize for the 1m raster resolution.

Slope Position Index (SPI) is classified based on Topographic Position Index (TPI) and slope (Jenness, 2007; Weiss, 2001). TPI compares elevation of the standing (central) point and average elevation of the surrounding neighborhood providing positive (central point is higher - like upslope or ridges) or negative (central point is lower - like lower slope or valley) values while a value of zero means it is a midslope or flat area. TPI is scale dependant so for this study I tested circular neighbourhood areas with radius of 5, 9.772, 15, 20, 30, 50, and 100 m, with the central point defined as a 1x1 m cell, as well as a circular area 2 and 3 m in radius. The method described by Weiss (2001) was used to classify TPI into six Slope Position categories based on standard deviation (SD) of TPI and slope: crest ($\text{SD} > 1$); upper slope ($0.5 < \text{SD} \leq 1$); middle slope ($-0.5 < \text{SD} < 0.5$, $\text{slope} > 5^\circ$); flat ($-0.5 \leq \text{SD} \leq 0.5$, $\text{slope} \leq 5^\circ$); lower slope ($-1.0 \leq \text{SD} < -0.5$); valley ($\text{SD} < -1$). This approach to classifying slope position corresponds to the classification used in

Beckingham et al. (1996) for field surveys in Alberta, and allows comparison between SPI maps derived from DEM data with field observations. Buffer areas of 15 and 20 m best corresponded to the field observations and therefore were selected as the most appropriate for the moderate (hilly) relief of the study area. Using different sizes of standing points (cell size, 2 and 3 m in radius) was found to have no effect so we proceeded with the raster cell elevation value as the standing point. All derived remotely sensed variables were calculated at the plot level except for Slope Position Index (SPI) which already uses a buffer area in calculation. Since pixel size in all rasters produced is 1 m², focal statistics tools in the spatial analyst toolbox in ArcGIS 10.1 were used to get values for circular plot with a radius of 9.772 m.

From the original dataset, aspen plots not affected (older than) the recent fire, aspen in mixed stands, and spruce and pine plots with extreme DTW values (resulting in a large gap in the dtw range) were excluded from the further analysis. Aspen measured in mixed stands with lodgepole pine experience lower SI values than in the pure stands with similar site conditions, which may result from suppression due to interspecies competition in earlier development. A summary of the final dataset used for the analysis is presented in the Tables 2.2.-2.4.

Table 2.2. Summary of stand attributes for the three studied tree species.

Variables	Abbrev.	Trembling aspen (n=97)		Lodgepole pine (n=50)		White spruce (n=45)	
		Mean (S.D.)	Min/Max	Mean (S.D.)	Min/Max	Mean (S.D.)	Min/Max
Site Index (m)	SI	20.97 (2.78)	14.21/25.52	14.93 (3.02)	5.44/19.43	17.40 (2.67)	11.99/22.99
Mean top height tree age (years)	AGE	45.6 (0.7)	42.3/47.0	43.4 (3.0)	33.3/47.0	83.2 (13.8)	31.7/114.3
Mean height of top height trees (m)	TH	20.0 (2.7)	13.4/24.7	13.4 (2.8)	5.1/20.2	24.3 (3.6)	11.2/29.9
Mean DBH of top height trees (cm)	DBH	20.5 (3.0)	14.1/26.7	16.2 (3.2)	8.8/23.3	38.2 (8.2)	15.2/59.0
Basal Area (m ² /ha)	BA	39.62 (11.16)	21/84	33.78 (12.91)	6/72	41.87 (11.60)	12/63
Basal Area participation (%)	BA%	94.8 (10.7)	50/100	78.2 (21.7)	23.1/100	76.9 (19.3)	42.9/100

Table 2.3. Summary of remotely sensed variables considered in modeling SI by studied tree species.

Variables	Abbrev.	Trembling aspen (n=97)		Lodgepole pine (n=50)		White spruce (n=45)	
		Mean (S.D.)	Min/Max	Mean (S.D.)	Min/Max	Mean (S.D.)	Min/Max
Depth-to-water at 0.5 ha c.a.(m)	DTW_0.5	4.51 (5.76)	0.02/33.17	1.00 (1.13)	0.00/5.19	1.97 (2.08)	0.02/7.11
Depth-to-water at 1 ha c.a. (m)	DTW_1	5.80 (7.65)	0.10/51.18	1.14 (1.33)	0.00/6.69	2.35 (2.47)	0.02/9.42
Depth-to-water at 2 ha c.a.(m)	DTW_2	7.47 (8.97)	0.17/55.70	1.22 (1.47)	0.00/7.50	2.67 (2.66)	0.02/10.33
Depth-to-water at 4 ha c.a. (m)	DTW_4	10.01 (10.78)	0.29/57.15	1.56 (1.85)	0.00/7.87	2.82 (2.82)	0.02/10.33
Depth-to-water at 6 ha c.a.(m)	DTW_6	11.35 (11.92)	0.36/57.26	1.86 (2.31)	0.00/8.21	3.28 (3.29)	0.03/11.78
Depth-to-water at 10 ha c.a. (m)	DTW_10	12.68 (12.57)	0.45/58.02	1.96 (2.37)	0.00/8.75	3.51 (3.33)	0.03/11.78
Flow accumulation	FA	15.7* (169.1)	3.7/1399.9	12.3* (393.2)	3.9/2515.2	42.7* (21994.9)	4.5/127651.0
Altitude (m)	ALT	810.9 (38.5)	742.8/929.8	840.6 (42.9)	755.5/908.2	802.3 (43.1)	730.7/878.3
Aspect index N-S folded	AI0	95.1 (37.5)	10.0/162.9	96.7 (34.1)	16.0/160.2	84.4 (41.5)	10.9/165.6
Aspect index NE-SW folded	AI45	88.9 (40.2)	12.6/168.2	86.0 (36.7)	18.1/161.9	72.0 (40.1)	17.8/167.5
Slope (%)	SLO	18.1 (13.5)	3.5/62.8	4.5 (9.8)	3.7/23.9	16.2 (9.0)	3.1/35.3
Mean curvature	MCUR	0.9 (1.64)	-3.09/8.93	0.64 (0.84)	-0.56/3.06	0.14 (1.25)	-2.49/3.45
Profile curvature	PRCUR	-0.59 (1.10)	-5.81/2.18	-0.39 (0.46)	-1.72/0.30	0.06 (0.84)	-1.73/2.34
Plan curvature	PLCUR	0.30 (0.68)	-1.00/3.12	0.25 (0.49)	-0.50/1.45	0.20 (0.79)	-0.87/2.95
Topographic Wetness Index	TWI	70.1* (2288.9)	16.8/ 22552.5	97.7* (2704.8)	36.6/ 18950.0	69.5* (529.5)	28.4/3619.1
Slope Position Index - 15 m buffer	SPI15	Level, lower slope, middle slope, upper slope, crest		Level, middle slope, upper slope, crest		Level, lower slope, middle slope, upper slope, crest	
Slope Position Index - 20 m buffer	SPI20	Depression, level, lower slope, middle slope, upper slope, crest		Level, middle slope, upper slope, crest		Level, lower slope, middle slope, upper slope, crest	

* median

Table 2.4. Summary of ground-measured variables considered in modeling SI by studied tree species.

Variables	Abbrev.	Trembling aspen (n=97)		Lodgepole pine (n=50)		White spruce (n=45)	
		Mean (S.D.)	Min/Max	Mean (S.D.)	Min/Max	Mean (S.D.)	Min/Max
Soil organic thickness (cm)	SOT	10.8 (5.3)	3/50	11.8 (14.0)	4/70	15.9 (10.7)	5/70
Coarse fragments (%)	CF	7.1 (13.0)	0/70	11.1 (15.2)	0/75	5.4 (12.7)	0/60
Humus form	HUM	Moder, raw moder, mor		Moder, raw moder, mor, peatymor		Moder, raw moder, mor, peatymor	
Texture	TEXT	Loamy sand, sandy loam, silt loam, loam, sandy clay loam, clay loam, silty clay loam, sandy clay, clay		Sandy loam, loam, sandy clay loam, clay loam, silty clay loam		Sandy loam, loam, sandy clay loam, clay loam, silty clay loam, sandy clay, silty clay	
Effective texture class	ETEXT	Coarse, moderately coarse, medium, fine		Moderately coarse, medium, fine		Moderately coarse, medium, fine	
Drainage	DRN	Rapidly drained, well-drained, moderately well-drained, imperfectly, poorly, very poorly		Well-drained, moderately well-drained, imperfectly, poorly, very poorly		Well-drained, moderately well-drained, imperfectly, poorly, very poorly	
Soil moisture regime	SMR	Subxeric, submesic, mesic, subhygric, hygric		Submesic, mesic, subhygric, hygric, subhydric		Submesic, mesic, subhygric, hygric, subhydric	
Soil nutrient regime	SNR	poor, medium, rich		Very poor, poor, medium, rich		poor, medium, rich	
Altitude (m)	ALTT	797.2 (38.2)	728.7/914.7	824.5 (42.8)	739.7/894.6	786.5 (43.5)	711.5/864.6
Aspect index N-S folded	AI0K	96.5 (49.1)	0/180	50.0 (104.0)	0/176	92.2 (50.7)	0/180
Aspect index NE-SW folded	AI45K	86.1 (52.8)	0/180	89.9 (53.7)	1/180	80.3 (56.0)	7/179
Slope (%)	SLOV	18.5 (17.6)	0/80	9.6 (6.3)	0/27	15.8 (11.9)	0/45
Slope Position	SPI	Depression, level, lower slope, middle slope, upper slope, crest		Depression, level, lower slope, middle slope, upper slope, crest		Level, lower slope, middle slope, upper slope, crest	

2.5. Analysis

Graphical data exploration was done prior to analysis to investigate the form of relationships between Site Index and individual predictor variables. Scatterplots with smoothing spline fits were used in exploring linearity, while boxplots were used in examining normality of distribution of variables. Statistical analysis then involved two major steps in terms of building species-specific SI prediction models: (1) Correlation analysis was used in selection of potential explanatory variables; (2) variables most closely related to SI were used in modeling SI variation. Dummy variables were assigned to all categorical variables. Log-transformed SI was tested in order to improve model predictive performance (R^2 value), however since this did not result in significant improvements, untransformed Site Index was used as the dependant variable. Finally, goodness-of-fit parameters were determined to compare between the best selected models for each species, modeling approach and data source. Model evaluation and comparison was based on absolute and relative measures of the fit (Quinn and Keough, 2002): the root mean squared error (RMSE), relative RMSE, coefficient of determination (R^2), adjusted R^2 (R^2_{adj}), Akaike information criteria (AIC), and Bayesian information criteria (BIC). The general rule for model selection was that a simpler model (with fewer parameters) was accepted if difference in $adj.R^2$ was less than 2%. Residual plots and plots of measured versus predicted SI were also examined. In addition, partial response curves (PRC), plotting particular variables against Site Index while holding others constant, were examined to evaluate behaviour of each covariate included in the model in terms of ecologically interpretability. All statistical analysis was completed using the R studio statistical software Version 0.98.1091 (R version 3.1.0; R Core Team, 2014). Graphics were created using the packages *ggplot2* (version 1.0.0; Wickham 2014) for explanatory data analysis, *plotmo* (version 3.1.4; Milborrow 2015) for 3-D plots, and *raster* (version 2.5-8; Hijmans et al., 2016) for spatial model predictions.

Correlation analysis was performed to explore strength of the relationship between Site Index and environmental variables as the first step in variable selection for modeling. Kendall's rank correlation coefficient was applied since assumptions of normality and/or linearity were not satisfied for the majority of bivariate relationships, and to accommodate use of dummy variables in the datasets. In cases where a non-monotonic relationship was observed among variables, a quadratic relationship was tested. Only variables that were significantly ($p < 0.05$) correlated with

SI were selected for modeling. Kendall's correlation coefficient was also used to select between collinear variables for particular site factor, e.g. Slope Position Index with different buffer areas, Aspect Index, surface curvature, drainage or SMR, soil texture etc.

The following modeling techniques, which are widely applied in similar studies, were used:

- i. Parametric - Multiple Linear Regression (Aertsen et al., 2010; Bravo-Oviedo et al., 2011; Chen et al., 2002, 1998; Fries et al., 2000; Hamel et al., 2004; Mitsuda et al., 2007; Monserud et al., 2006; Pinno and Bélanger, 2011; Pinno et al., 2009; Seynave et al., 2005; Sharma et al., 2012; Socha, 2008; Szwaluk and Strong, 2003; Wang and Klinka, 1996; Wang, 1995; Wang et al., 2004; Watt et al., 2015, 2010a);
- ii. Semi-parametric - Generalized Additive Models (Aertsen et al., 2011, 2010; Albert and Schmidt, 2010; Brandl et al., 2014; Wang et al., 2005);
- iii. Non-parametric (single regression tree-based method) - Classification and Regression Trees (Aertsen et al., 2011, 2010; Laamrani et al., 2014; McKenney and Pedlar, 2003; Wang et al., 2005);
- iv. Non-parametric (multiple regression tree-based method) - Random Forest (Jiang et al., 2015; Sabatia and Burkhart, 2014; Weiskittel et al., 2011a).

Multiple Linear Regression (MLR) is a parametric approach that uses the ordinary least squares method in determining SI-environmental relationships. Prior to the MLR analysis the form of the relationship between each selected environmental variable and SI was examined graphically, and in cases when non-linear relationships were evident data were linearized using quadratic, logarithmic, or exponential functions. The backward selection procedure was used retaining only variables with a significant slope parameter ($p < 0.05$). Best-subsets regression from the *leaps* package in R (version 2.9; Lumley 2009) was used in the backward selection procedure to screen the model and associated explanatory variables in order to determine the most important and optimal number of variables in the model. The best model, among the set of candidate models, was selected based on maximizing adjusted R^2 and minimizing BIC. Interactions between the selected variables were also tested to evaluate their behaviour in the models. Multicollinearity and potential overfitting were assessed using Variance inflation factors (VIF) assuming $VIF < 5$ overcomes the issues. Regression diagnostics were used in evaluating model precision and bias.

Assumptions of normality and homogeneity of variance were tested on residuals visually on graphs and using the Shapiro-Wilk test.

Generalized Additive Models (GAM) are a non-parametric (generalized) extension of Generalized Linear Models which involve fitting a smoothing (non-linear) curve for multiple explanatory variables. GAM's are useful when the non-linear relationship between response and explanatory variables is unknown (Zuur et al., 2009). Alternatives to additive modelling are to transform (linearize) particular variables, or to apply a chosen parametric non-linear model. But the power of additive modeling in ecological research, even if it is a data- rather than a model-driven method, is its ability to determine the nature of the relationship between the response and predictors and, consequently, increases our understanding of ecological relationships (Guisan et al., 2002). For each of the independent variables additive modeling allows application of different smoothing functions and smoothing coefficients (Logan, 2010), thus GAM is referred to as a semi-parametric method when linear and smooth terms are combined in the model. Since each non-parametric term is associated with a smoothing function, response is modeled as a sum of the functions, not as a single equation, hence GAMs are termed additive (Quinn and Keough, 2002). The advantage of GAM is that it provides better fits for highly non-linear and non-monotonic relationships. However, selection of the optimal amount of smoothness, the balance between the less and more smoothed curve, is a matter of bias-variance trade-off. Thus complexity of fitting the GAM model is in determining the appropriate smoothing function and degree of smoothing for each term. Instead of applying the back-fitting method used in some older GAM algorithms, the *mgcv* R package uses generalized cross-validation techniques to optimize the level of smoothing (effective degrees of freedom - edf) and resulting smoothing (penalized) spline (Zuur et al., 2009). It also provides GAMM (Generalized Additive Mixed Models) modeling if there is more than one source of variation.

GAM modeling was done using *mgcv* package Version 1.8-5 (2015) in the R programming environment (Wood and Wood, 2013). The following smoothness determination was applied: scatter plots with smoothing splines were examined for all variables to distinguish between linear and non-linear trends with SI; different smoothing functions and associated basis dimension were then selected in order to achieve optimal fitting of the data; a penalized cubic regression spline function with shrinkage applied with an upper limit of degrees of freedom was then applied to

guard against over-smoothing but still capture non-linear trends in the data; and, finally, smooth terms resulting in straight lines (edf=1) were switched to linear terms. According to Zuur et al. (2009), a cubic regression spline (fitting third order polynomials on intervals of data and ensuring a smooth connection at the knots) is one of the most commonly used smooth function, while shrinkage options allow a smoothing spline to have zero edf meaning that variable has negligible effect on the response. Backward variable selection was used: poor predictors determined by shrinkage (edf close to zero) were dropped simultaneously; then the least significant term (p-value > 0.05) was dropped; and then BIC was used to select the optimal model. Model validation plots were examined using the built-in *gam.check* function output. Adj.R², RMSE, Partial response plots, residual vs. predicted, and actual vs predicted plots are reported for the model evaluation. Detailed explanation of fitting GAM models with R can be found in Wood (2006), application to SI prediction under climate changes in Albert and Schmidt (2010), and GAMM with environmental predictors of tree species distributions in Andraž et al. (2015).

Classification and Regression Trees (CART) provide a non-parametric method for building descriptive and predictive models from data. This method is effective for discovering complex hierarchical interactions among predictors and non-linear relationships between predictor and response variables. CART has no assumptions relating to the distribution of the response or predictor variables (Kane et al., 2015). Predictors can be any combination of categorical, binary, or continuous variables. Regression trees are a subset of CART and deal with dependent variables that take continuous or ordered discrete values, with squared difference between the observed and predicted values typically used as prediction error (Loh, 2011). The CART algorithm consists of three general steps: building, stopping, and pruning the tree (Olden et al., 2008). It works by evaluating all independent variables for all possible splits of observations into two groups, binary recursively partitioning so that one parent node has two child nodes, and determining predictors that maximize the homogeneity of the two resulting groups with respect to the response variable (the best explains deviance in the response variable). Partitioning continues in each resulting group (terminal nodes) until either group is homogeneous or there are not enough observations for splitting. Measures of homogeneity and splitting criteria used to decide on the split should be set and can vary with the algorithm used, while stopping criteria are usually set based on the minimum number of observations in terminal nodes or minimum

number of observations for a split attempt. To deal with the issue of overfitting more strict splitting and/or stopping rules can be defined or applied in the process of pruning the tree. The 10-fold cross-validation technique (Maindonald and Braun, 2003) is usually used to prune the overfitted tree to optimal size. Predicted value is the mean of observations in the terminal node (Quinn and Keough, 2002) hence CART actually predicts classes because it is not fitted with the smooth function. Some CART advantages, if the decision tree is not too large, are being simple to visualize and interpret, robust to outliers, simple to implement, and simple to run. A weakness of CART is that it is unstable (Ziegler and König, 2014), with results in these models being sensitive to small changes in inputs. In addition, like other tree-based methods, CART is not appropriate for analyzing data with a complex error structure (Maindonald and Braun, 2003).

To grow regression trees I used the *rpart* package Version 4.1-10 in R (Therneau et al. (2015) based on Breiman's CART algorithm) which uses the "anova" method and complexity parameter (cp) as the splitting criteria. A complexity parameter of 0.01 (default) was used, what means the split is accepted if it explains 1% in response variation. Stopping criteria were set to a minimum of 10 observations for each split attempt, with a minimum 5 observations in any terminal node, and 10-fold cross-validation (default). Overall, splits were selected based on the predictor that resulted in maximizing between groups sum-of squares, accepted if overall R^2 was improved by at least 1%, and the partitioning process was repeated for each new group (node) until the stopping criteria was reached. Pruning was applied to avoid overfitting. Graphs of R-squared (fitted and from cross-validation) and cross-validation relative error versus number of splits were examined to determine optimal number of splits (to minimize the cross-validation error). The optimal regression tree was obtained following the *prune* function with applied cp associated with determined optimal tree size. More detail about CART with *rpart* is given in Maindonald and Braun (2003), while more about ecological modeling applications of CART in SI predictions can be found in McKenney and Pedlar (2003) and application of *rpart* package in using GIS environmental variables in habitat prediction for rare forest herb is provided by Bourg et al. (2005).

Random forest (RF) is a non-parametric method widely and successfully applied for modeling in environmental (Prasad et al., 2006) and life sciences (Touw et al., 2013), or in combination with parametric methods as a variable filter prior to using other modeling techniques (Sabatia and

Burkhardt, 2014). RF is an extension of Classification and Regression trees (CART) which grows many decision trees, first introduced by Breiman (2001). RF is a machine learning method appropriate for nonlinear multiple regression. It overcomes the issue of stability of CART models (Prasad et al., 2006), with RF models being relatively stable since they are a combination of many trees. Other advantages are accurate prediction power, ability to detect variable interaction, ability to measure variable importance, internal cross-validation, insensitivity to outliers, ability to overcome collinearity, low tendency to overfitting (hence no need for pruning), efficiency with large datasets with many variables, and the fact that it is relatively easy to apply. RF can be applied to regression problems as well as to classification problems, estimation of probabilities, and analysis of survival data. The approach consists of four general steps (Ziegler and König, 2014): dataset is first bootstrapped with replacement; for every bootstrap sample decide on the independent variables to be used for growing the tree; a decision tree is grown for every bootstrap sample; and, summarize (average) over all grown trees to build a forest (model). I explored regression relationships between environmental variables and SI, therefore random bootstrapping was used and Mean Square Error (MSE) was used as the stopping criteria for node splitting. Random bootstrapping of samples and variables is first done to generate many single datasets from the original dataset. Randomly bootstrapping samples with replacement resulted in a training ('bootstrap') set, containing about two-thirds of the samples in the original dataset, while the remaining samples are the 'out-of-bag' (OOB) samples, thus cross-validation is actually built-into the method resulting in no need for additional validation. Each decision tree is grown by binary recursively partitioning (splitting) data into subsets. Number of variables to be included for splitting the data at each node (*mtry*) limits correlation between decision trees and is set by default in the RF algorithm (\sqrt{p} for classification, $p/3$ for regression; where p is number of variables) or can be tuned so that the error rate is minimal. Variables are randomly evaluated to be included for every tree 'node', and the best in terms of minimizing the MSE as splitting criteria are selected. The regression decision tree is grown to the greatest extent, however restriction of minimum terminal node size used in a training ('bootstrap') is commonly applied. Finally, results are summarized after each set of decision trees is created to build a model (Ziegler and König, 2014). Model prediction error, called out-of-bag (OOB) error (which is actually RF's MSE), is calculated for each decision tree using corresponding OOB samples. Averaged MSE for all trees and variance explained (an estimate of R^2 using overall MSE) are

given in terms of model validation. In addition, a variable importance measure of each predictor is given as the average percent change (before and after permutation) in MSE when the variable is permuted while all others are retained unchanged, therefore permuting an important predictor should cause a relatively large increase in MSE and vice-versa (Weiskittel et al., 2011a).

The ‘*randomForest*’ package Version 4.6-10 in R statistical software (Liaw et al., 2015) was used. I applied function *tuneRF* to determine optimal *mtry* (number of predictors selected at each node). However, due to the small total number of predictors the effect of tuning the number of predictors selected at each node was negligible, so the default ($p/3$) was retained. After preliminary analysis the number of regression trees to grow was fixed to 3000 trees to ensure minimizing the OOB error (Figure 2.2). Other parameters were assigned default values in the *randomForest* function (min nodesize=5, etc.). Square root of MSE, variance explained output (pseudo R^2), residuals vs predicted plots, observed vs predicted plots with 1:1 lines, and partial dependence plots (effect of each predictor on Site Index while the other predictors are held constant) were used for model evaluation. In addition, we computed variable importance (% increase MSE) to determine optimal number and identify the best predictors. The backward variable selection approach (Wang et al., 2015; Weiskittel et al., 2011a) was used. Starting from the "full" model, variables with the lowest importance were dropped until only two variables were left in the model, each time computing RMSE and pseudo R^2 . Finally, R-squared was plotted against number of variables and the model with the highest R^2 was chosen.

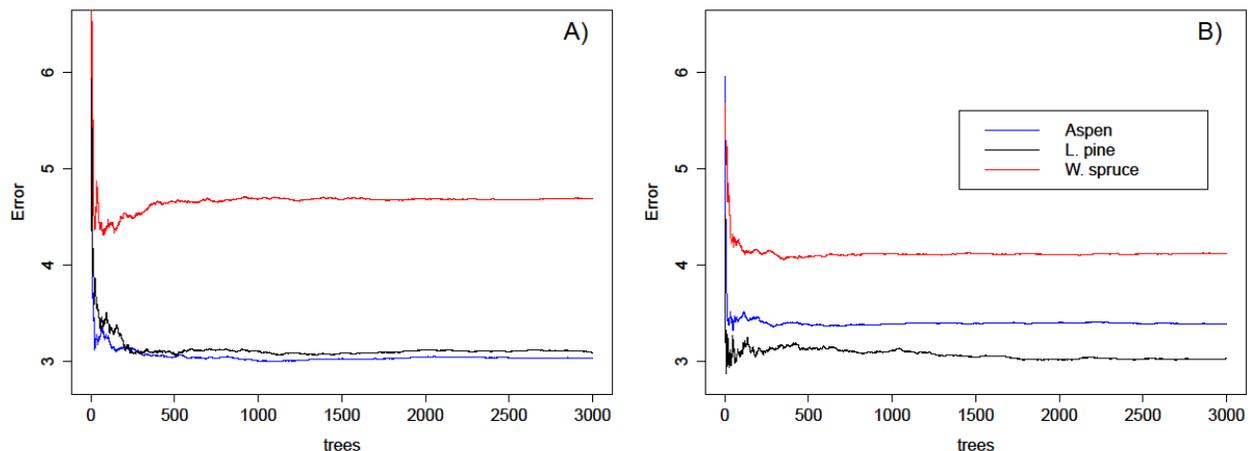


Figure 2.2. OOB Error in relation to number of decision trees grown to build random forest model. A) SI prediction model from remotely sensed variables; B) SI prediction model from ground-based variables.

3. RESULTS

3.1. Variation in Site Index and explanatory variables

Spatial distribution of sample plots is shown in Figure 2.1 and summary statistics for all variables determined are presented by species in Tables 2.2-2.4. Relatively uniform spatial distribution resulted across selected aspen and pine blocks excluding extreme pine sites observed on hill tops, while spruce plots were grouped in a few locations where spruce remained after the last fire. Stand characteristics ranged widely in terms of Site Index and Basal Area for all three species. The highest average SI of 21.0 m was observed for aspen, then 17.4 m for spruce, and 14.9 m for lodgepole pine, while the widest SI range was recorded for pine, then aspen, and spruce. Total BA, used as a stand density measure, averaged 39.62 m²/ha in aspen stands, 33.78 m²/ha in pine, 41.78 m²/ha in spruce, which are common values for associated ages, with a maximum of 84 m²/ha in the most productive pure aspen stand, and a minimum of 6 m²/ha in a pure pine stand with an understory dominated by lichens and Labrador tea and a poor soil nutrient regime. Aspen and pine SI significantly correlated with total BA (Figure 3.1), but no significant relationship was observed for spruce due to the wide age range of spruce stands. Aspen stands were mainly pure with an average 94.8% of BA being composed of aspen, but sometimes with a minor component of other broadleaves, usually balsam poplar. Pine dominated (78.2%) in the mixture with aspen, birch or black spruce, while white spruce (76.9%) was commonly mixed with balsam poplar and aspen. Although other studies have observed increases in SI with increase of main species BA proportion (Anyomi et al., 2015, 2013; Brandl et al., 2014; Pinno et al., 2009), no significant relationship were found for the three species in this study (Figure 3.1) confirming the assumption of no impact of inter-species competition on height growth for the trees sampled in this study.

The Vega fire of 1968 resulted in a narrow age range with averages of 45.6 years in aspen stands and 43.4 years in pine stands. Only two pine plots originated from 1980 after another fire which occurred in a restricted area. Delays in regeneration after fire occurred mainly on more productive sites, however 91.8% of aspen plots were recruited in the first two years since fire and 91.7% of pine plots regenerated within six years after the fire (Figure 3.2). Blue-joint grass and

other competing vegetation may cause regeneration delay on more productive sites. Spruce plots, except two plots, were not affected by the Vega fire and ranged between 75 and 115 years with the mode in the 80-90 age class (Figure 3.2). Similarly to some other studies, an age trend in SI was noted for all three species. However, due to the narrow age range in aspen and pine, this trend is negligible and associated with the GYPSY model, which appears to overpredict Y2BH on lower productivity sites and underpredict on more productive sites. This is probably due the anamorphic form assumed for these SI curves (Weiskittel et al., 2011b). In addition, potential causes of declines in spruce SI with tree age might result from early suppression effects of taller competition, or due to changes in site conditions and productivity over time (climate, fertility) which was explained as a common cause of these trends in similar studies (Albert and Schmidt, 2010; Anyomi et al., 2015, 2013).

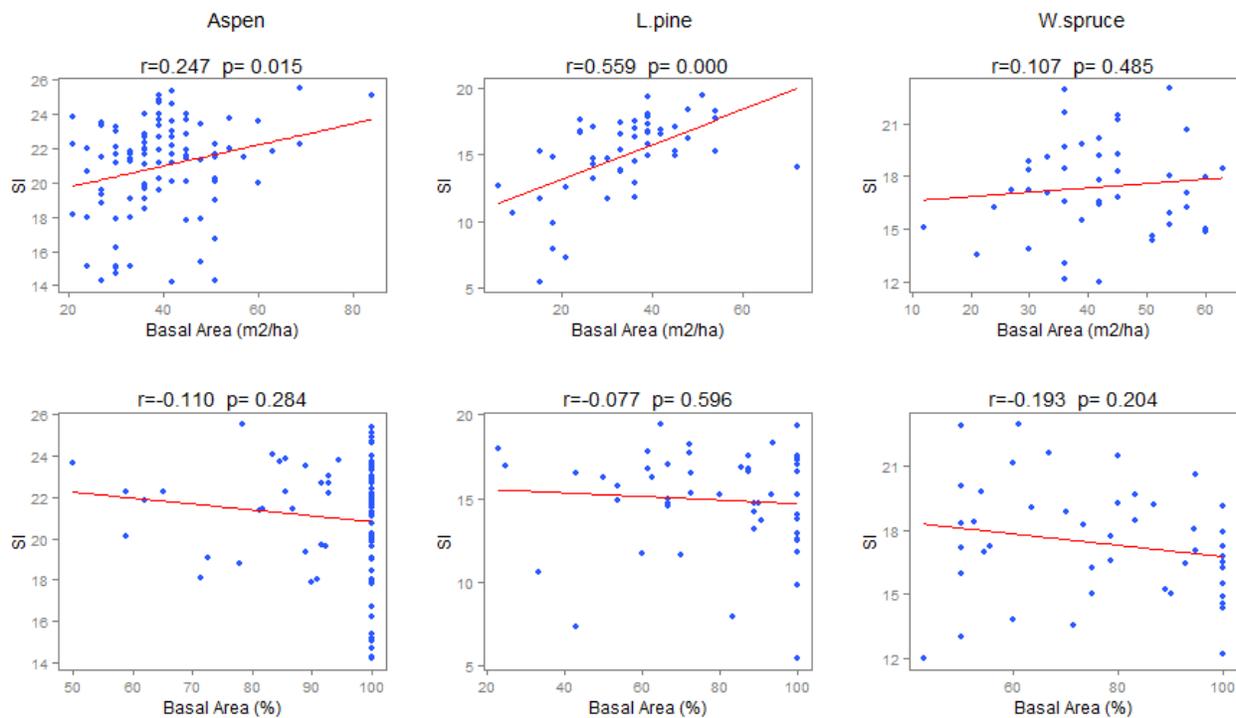


Figure 3.1. SI variation by species with total Basal Area (upper plots) and main species BA proportion (lower plots). Coefficient of correlation (r) and associated p -value are presented for each plot, red line shows linear trend in data.

As noted by Pinno et al. (2009), variable distributions usually do not follow a bell shape because the aim of sampling is to capture a range of factors with ideally even distribution. For instance sample plot distribution across aspect and altitude (Figure 3.3) – while showing continuous

distribution for aspen, has a gap in data for northerly exposed lower altitude sites for pine. For spruce sampling was unbalanced between easterly and westerly facing plots. However, due to moderately hilly relief and the fact that only a few higher hills exist in the study area sampling was limited in that extreme part of the range. As a result the majority of predictor variables show skewed distributions (Figures 3.4-3.6) with sampling being more common on lower slopes and wetter sites. Log transformation was applied for FA and TWI to better visualize distribution and relationships due to departure from a normal distribution.

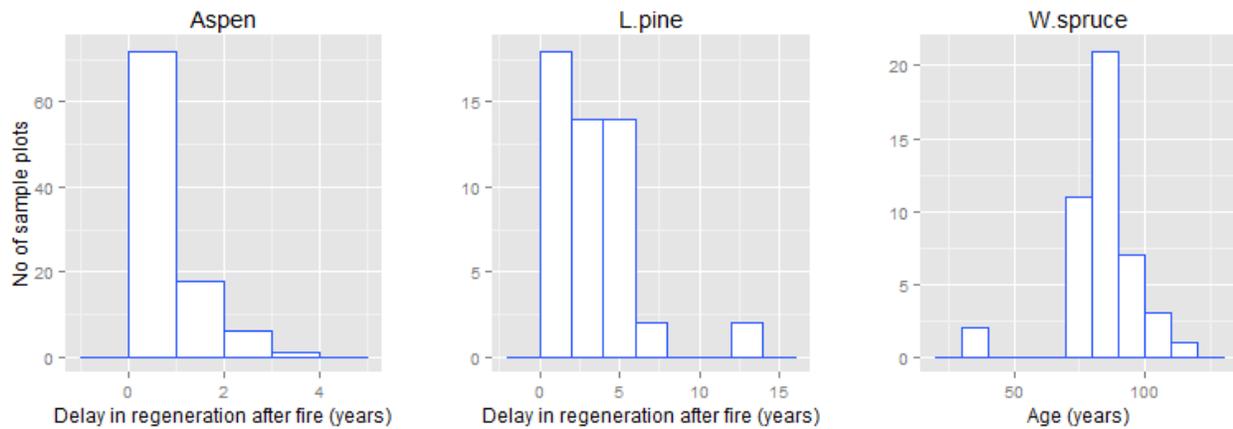


Figure 3.2. Age structure by species. Histograms for aspen and pine show time of regeneration occurrence after the most recent fire.

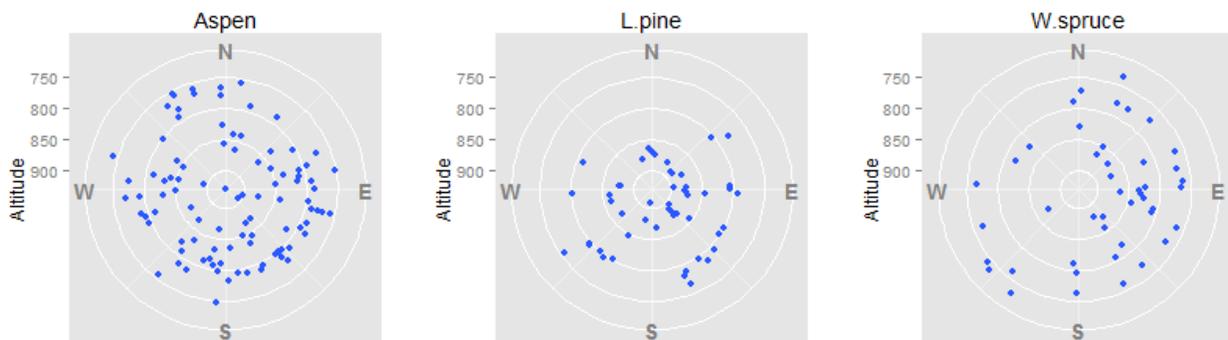


Figure 3.3. Sample plot distribution by species across aspect (polar axis) and altitude (y-axis)

The widest range of topographic variables was recorded for aspen due to its broad ecological niche and its distribution across a broad range of sites in the study area. Pine and spruce were sampled on the range of sites but were also found in subhydric conditions, that had lower DTW, higher FA, and peatymor humus, where aspen does not grow. Pine was most often found on poor

to very poor soil nutrient regimes avoiding clay based soils. Average DTW at a 4 ha flow initiation area of 10.0 m observed for aspen (ranged 0.3-57.2), 1.6 m for pine (0.0-7.9), and 2.8 m for spruce (0.0-10.3). The DTW values varied with the FIA used in the way that increasing FIA results in increasing the range of DTW values. The steepest slope was found in aspen stands (63%), then 35% in spruce, and 24% in pine stands. Also, surface roughness, both concave or convex, was highest in aspen stands, but average curvature exhibited for all species was close to straight - flat or midslope conditions. All species had a similar altitude range except that several aspen plots were established close to the highest ridges in the area.

The scatterplots with smooth splines suggest both linear and non-linear relationship between SI and explanatory variables, with the form of the relationship chosen for different variables (Figures 3.7-3.9). Several variables showed nonlinear patterns in relation to SI index, including DTW, FA, aspect, slope, TWI, SMR and slope position. Also, analogous trends for each species were noticeable for some variables indicating potential collinearity, for instance between DTW, slope and SMR.

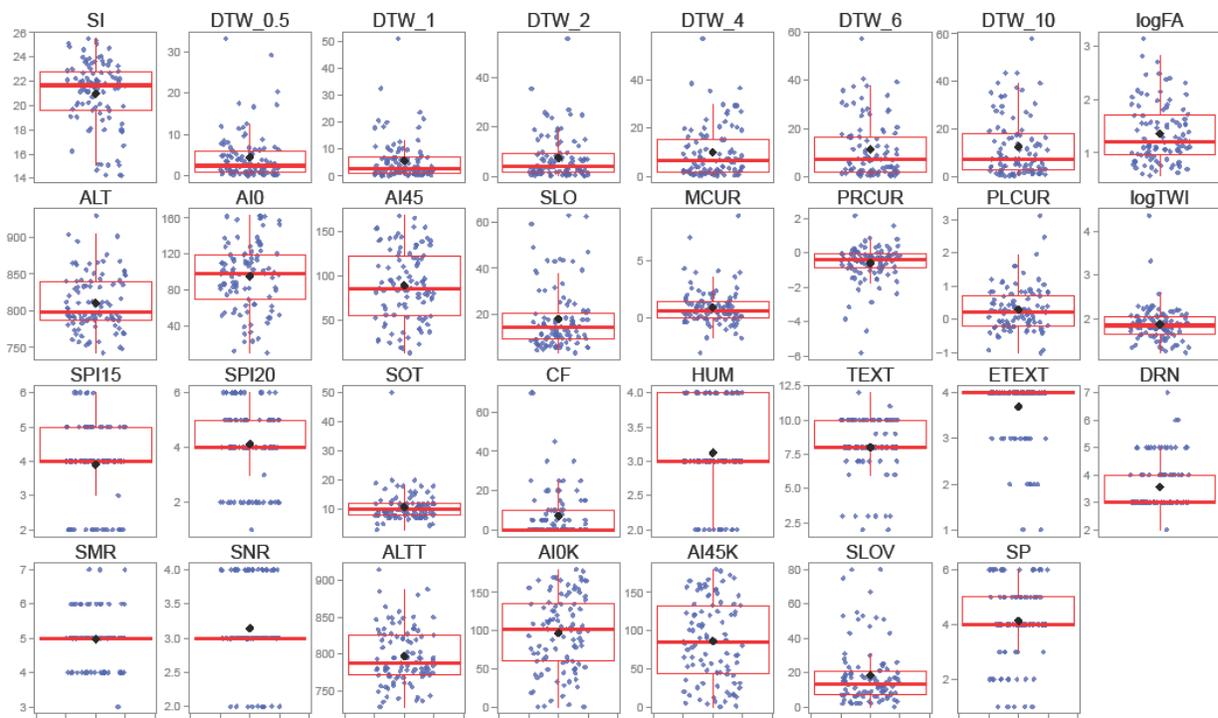


Figure 3.4. Boxplots with jitter plots of SI and environmental (remotely sensed and ground-based measured) variables for aspen. Black dot represents mean.

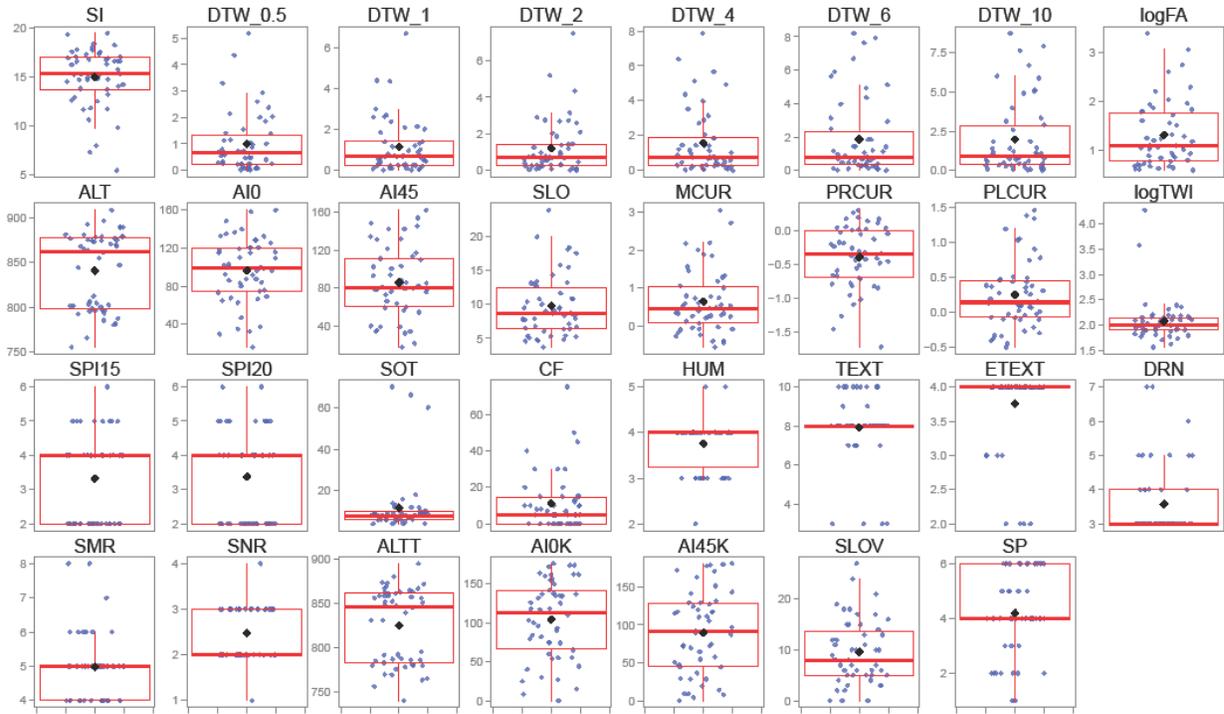


Figure 3.5. Boxplots with jitter plots of SI and environmental (remotely sensed and ground-based measured) variables for pine. Black dot represents mean.

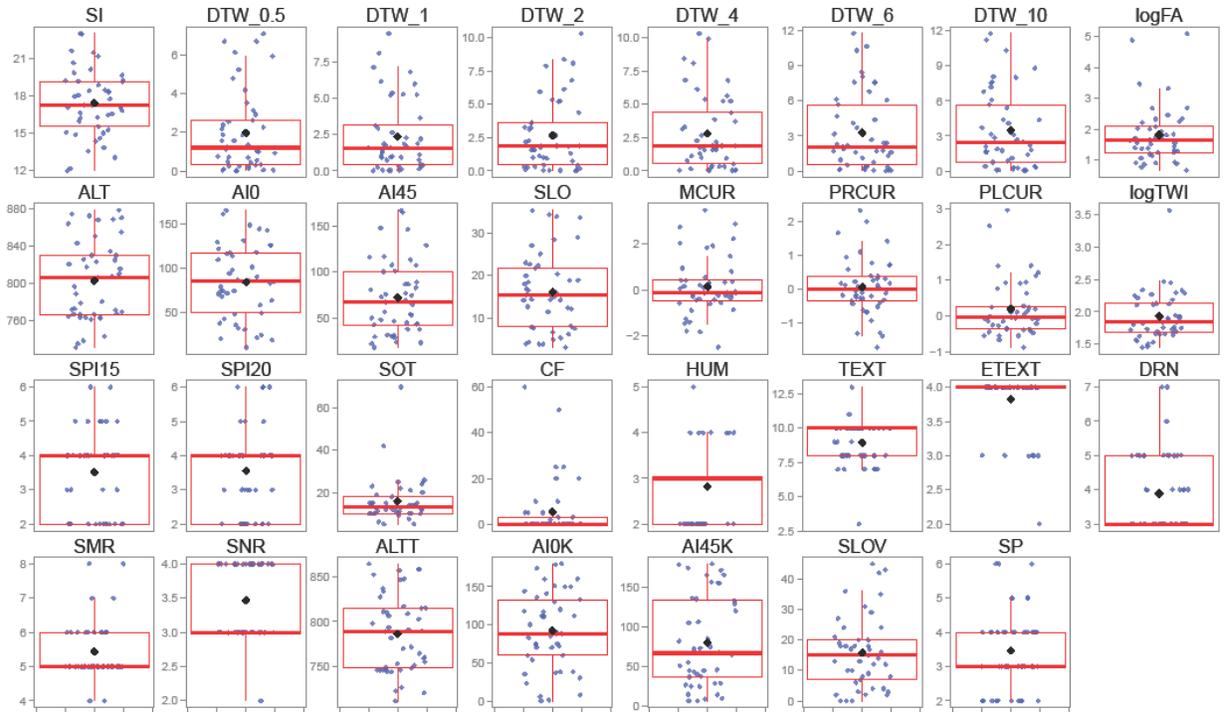


Figure 3.6. Boxplots with jitter plots of SI and environmental (remotely sensed and ground-based measured) variables for spruce. Black dot represents mean.

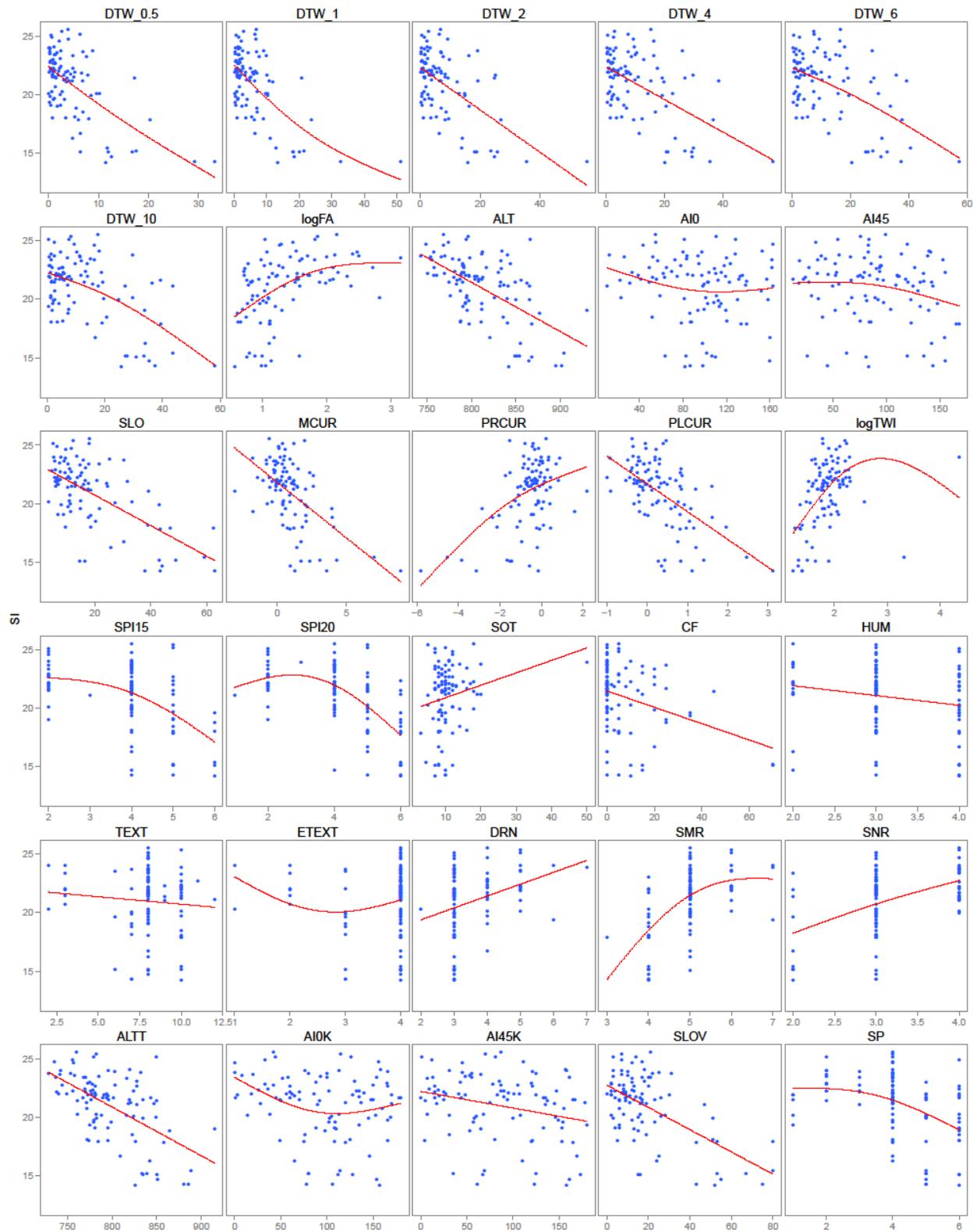


Figure 3.7. Scatter plots with penalized regression spline (3 degrees of freedom) between environmental (remotely sensed and ground-based measured) variables and Site Index for aspen.

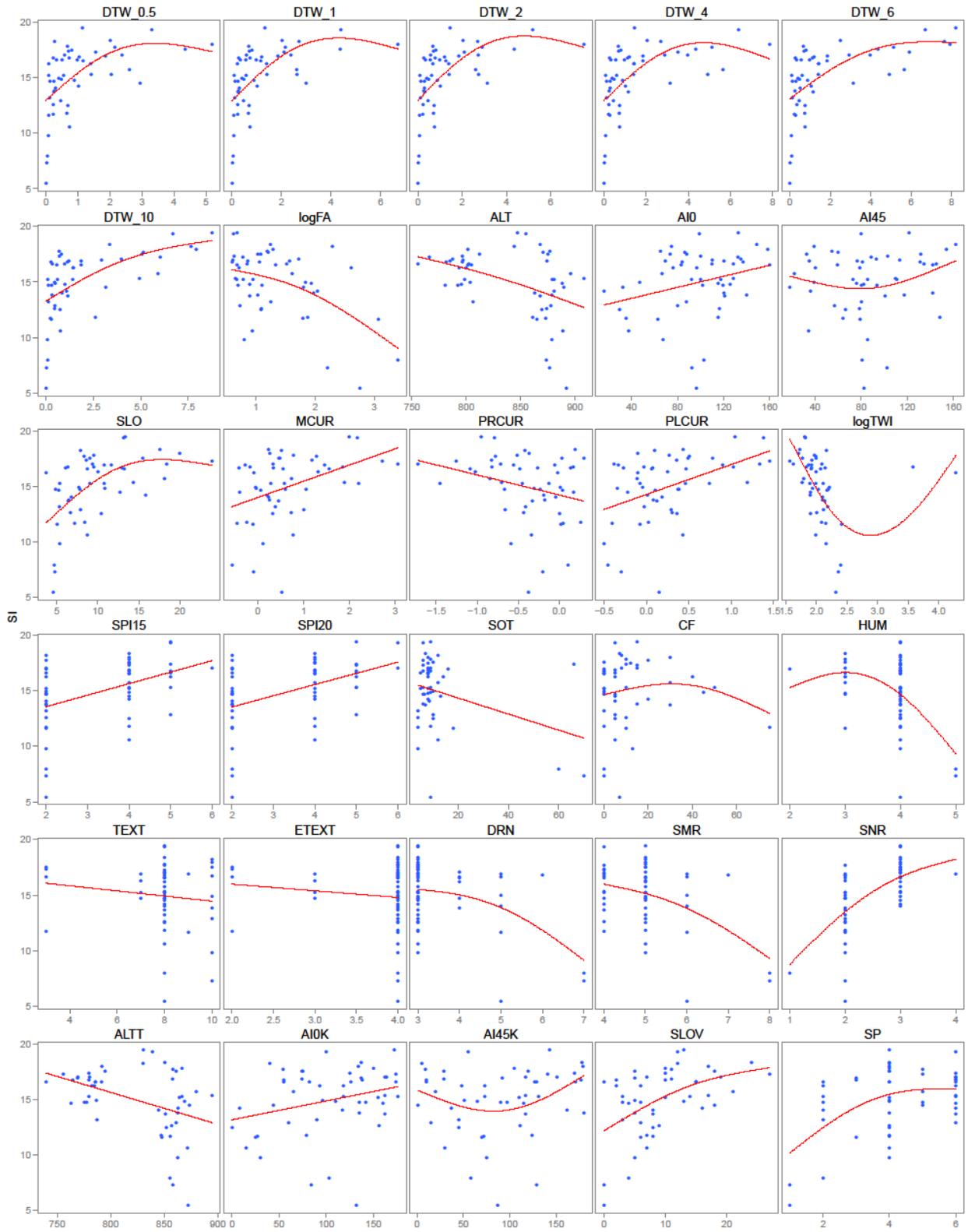


Figure 3.8. Scatter plots with penalized regression spline (3 degrees of freedom) between environmental (remotely sensed and ground-based measured) variables and Site Index for pine.

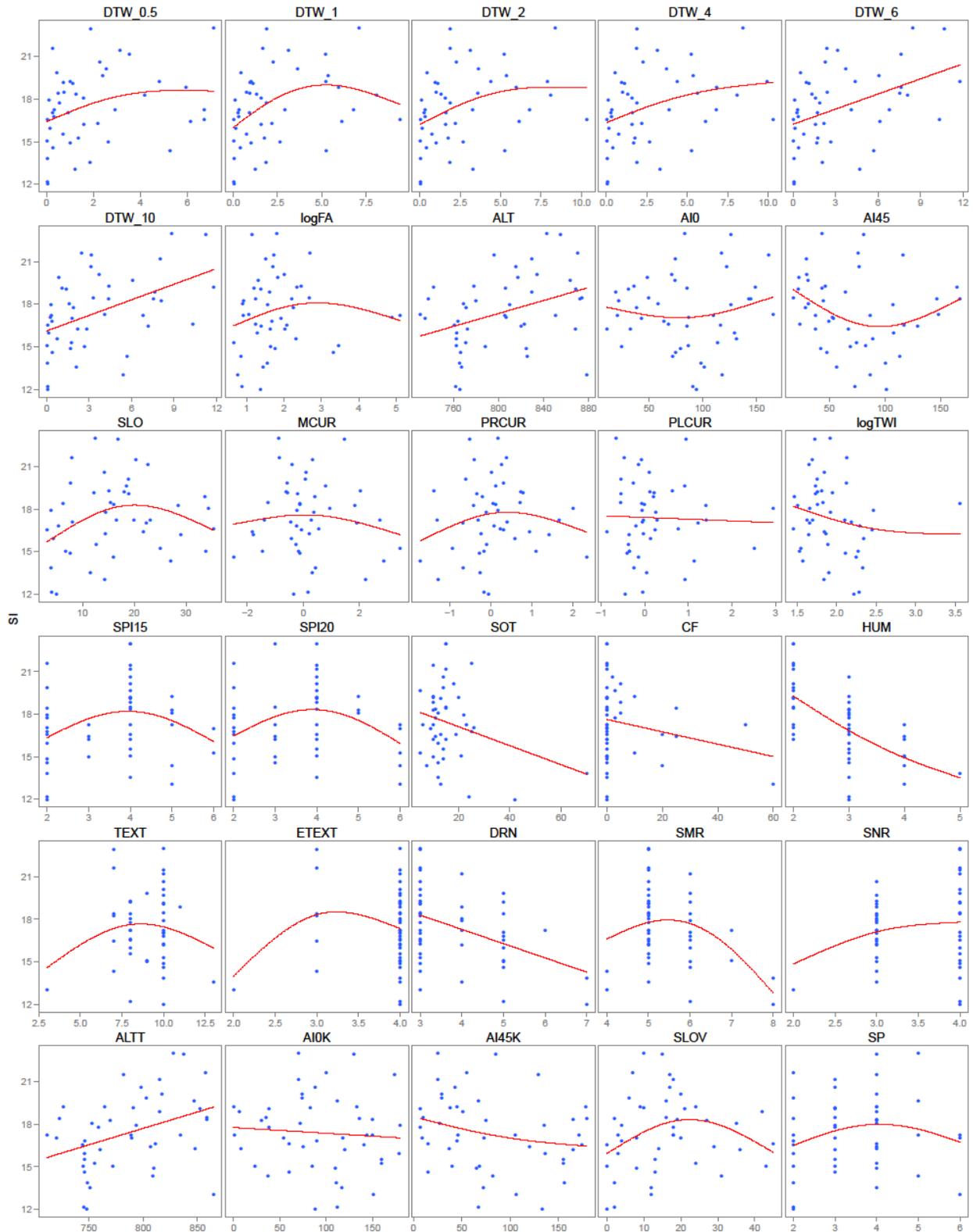


Figure 3.9. Scatter plots with penalized regression spline (3 degrees of freedom) between environmental (remotely sensed and ground-based measured) variables and Site Index for spruce.

3.2. Correlation analysis and Predictor variable selection for SI modeling

Remotely sensed topographic variables were found to be well correlated with field determined soil properties (Table 3.1) in the combined dataset for all three species (201 sample plots in total). SMR, DRN and effective texture were closely related with topography and were significantly related to all topographic indices except aspect, with the strongest positive association between FA and SMR ($r=0.46$, $p<0.001$). Humus type was also related to all topographic indices except aspect and slope position, while SNR had no significant correlation with aspect and slope (and variables closely associated to slope - TWI and DTW with smaller FIA). Soil organic thickness increased and coarse fragment content decreased significantly with increasing FA and lower topographic position measures. SMR and DRN ($r=0.74$, $p<0.001$), and HUM and SNR ($r=-0.50$, $p<0.001$) were soil variables highly correlated since DRN and HUM are used in determining SMR and SNR.

Table 3.1. Kendall’s rank correlation coefficient matrix between remotely sensed environmental and soil attributes. Bolded numbers indicate significance of relationship at $p<0.05$.

DTW_0.5	0.89	0.81	0.76	0.72	0.70	-0.17	0.17	-0.08	0.64	0.27	0.50	-0.63	-0.04	0.08	-0.13	-0.13	-0.33	-0.37	0.02
DTW_1	0.89	0.81	0.77	0.75	-0.14	0.17	-0.07	0.62	0.25	0.48	-0.60	-0.04	0.09	-0.15	-0.14	-0.33	-0.35	0.06	
DTW_2	0.89	0.84	0.80	-0.13	0.16	-0.06	0.58	0.25	0.49	-0.56	-0.03	0.10	-0.18	-0.14	-0.31	-0.35	0.08		
DTW_4	0.93	0.87	-0.13	0.14	-0.06	0.55	0.24	0.46	-0.53	-0.03	0.10	-0.17	-0.11	-0.30	-0.35	0.09			
DTW_6	0.93	-0.12	0.16	-0.03	0.54	0.24	0.46	-0.52	-0.02	0.09	-0.19	-0.13	0.28	-0.32	0.11				
DTW_10	-0.11	0.15	-0.03	0.52	0.23	0.44	-0.50	-0.02	0.06	-0.19	-0.13	0.26	-0.29	0.11					
FA	-0.11	-0.03	-0.13	-0.45	-0.41	0.22	0.31	-0.17	-0.24	0.20	0.35	0.46	0.32						
ALT	-0.04	0.16	0.12	0.19	-0.15	-0.03	0.17	0.11	-0.09	-0.28	-0.25	-0.22							
AI0	-0.05	-0.01	-0.01	0.05	-0.09	-0.01	0.00	-0.04	0.07	0.02	0.05								
SLO	0.27	0.52	-0.81	-0.02	0.08	-0.16	-0.12	-0.27	-0.30	0.05									
PLCURV	0.53	-0.31	-0.22	0.18	0.07	-0.11	-0.27	-0.40	-0.23										
SPI20	-0.53	-0.16	0.20	0.10	-0.20	-0.37	-0.47	-0.14											
TWI	0.05	-0.07	0.03	0.14	0.33	0.36	-0.01												
SOT	-0.20	-0.36	0.07	0.21	0.39	0.42													
CF	0.22	-0.01	-0.24	-0.35	-0.30														
HUM	-0.06	-0.1	-0.25	-0.50															
ETEXT	0.20	0.20	0.18																
Soil variables	DRN	0.74	0.30																
	SMR	0.39																	
	SNR																		

Topographic variables determined through ground sampling and computed from the DEM provided very similar values for Altitude ($r=1.0$, $p<0.001$), Slope ($r=0.95$, $p<0.001$), and Aspect Index ($r=0.84$, $p<0.001$), with small differences due to averaging over the plot areas. Slope position observed in the field was best represented by Slope Position Index computed using a 20 m radius ($r=0.65$, $p<0.001$).

Interestingly, FA seemed to be the most influential DEM derived variable in explaining variability of each measured soil property. DTW significantly correlates with SMR, DRN, texture classes, and humus form for all tested levels of FIA. SMR exhibited the strongest correlation with DTW, however the strength of this relationship ranged from $r=-0.37$ at 0.5ha FIA to $r=-0.29$ at 10ha FIA. In addition, all other RS topography variables explained soil properties to some extent, except that aspect did not significantly correlate with any soil variables. High intercorrelation was detected for some topographic variables, for instance between DTW, TWI and slope because slope is used in calculation of DTW and TWI. Also, curvature and slope position were correlated since it is more likely to have concave surface on lower slopes and depressions, or convex surface on ridges. Correlations between remotely sensed topographic variables which indicated potential issues with collinearity in modeling are: SLO-TWI ($r=-0.81$, $p<0.001$), SLO-DTW_0.5 ($r=0.64$, $p<0.001$), DTW_0.5-TWI ($r=-0.63$, $p<0.001$), DTW_0.5-SPI20 ($r=0.50$, $p<0.001$), MCURV-SPI20 ($r=0.57$, $p<0.001$).

3.2.1. Trembling aspen

Topographic remotely sensed variables, except aspect, had strong correlations with aspen SI with correlation coefficients ranging between 0.32 and 0.43 (Table 3.2). Declines in SI were associated with increases in DTW, altitude, slope, curvature, and slope position, and decreases in FA and TWI. To represent WAM's Depth-to-Water index, DTW_1 was selected based on analysis of optimal FIA (see chapter 3.3). Among the surface curvature indices planform curvature was selected, while SPI20 outperformed SPI15. The quadratic relationship was found to be weak but significant for AI45 explaining elevated SI from SW to NE aspects. TWI was dropped from further analysis due to its strong correlation with slope. Remotely sensed variables selected for modeling SI variation were: DTW_1, FA, ALT, AI45, SLO, PLCUR, SPI20. However, SPI20-PLCUR and SLO-DTW_1 showed strong collinearity.

Table 3.2. Kendall’s rank correlation coefficient matrix between aspen SI and environmental variables. Bolded numbers indicate significance of relationship at $p < 0.05$. Italicized numbers indicate coefficient of determination from quadratic regression.

SI	-0.32	-0.30	-0.30	-0.27	-0.25	-0.22	0.34	-0.38	<i>0.02</i>	<i>0.04</i>	-0.33	-0.34	0.21	-0.36	0.33	-0.36	-0.43
DTW_0.5	0.85	0.74	0.68	0.64	0.62	-0.10	0.37	-0.09	<i>0.06</i>	0.60	0.27	-0.18	0.29	-0.57	0.39	0.43	
DTW_1	0.82	0.73	0.68	0.67	-0.09	0.37	-0.07	<i>0.05</i>	0.58	0.26	-0.15	0.28	-0.54	0.39	0.42		
DTW_2	0.82	0.76	0.72	-0.09	0.41	-0.07	<i>0.04</i>	0.54	0.27	-0.16	0.27	-0.50	0.38	0.45			
DTW_4	0.92	0.84	-0.08	0.38	<i>0.06</i>	<i>0.04</i>	0.54	0.26	-0.15	0.26	-0.48	0.34	0.40				
DTW_6	0.89	-0.06	0.40	-0.09	<i>0.03</i>	0.53	0.25	-0.14	0.25	-0.46	0.36	0.34	0.41				
DTW_10	-0.04	0.40	<i>0.05</i>	<i>0.02</i>	0.52	0.24	-0.14	0.24	-0.46	0.33	0.38						
FA	-0.14	0.02	<i>0.00</i>	-0.13	-0.46	0.36	-0.46	0.19	-0.35	-0.42							
ALT	0.08	<i>0.00</i>	0.37	0.25	-0.18	0.21	-0.30	0.35	0.35								
AI0	0.46	0.03	0.03	-0.06	-0.04	0.01	0.05	0.02									
AI45	0.02	-0.02	0.07	-0.04	0.00	-0.02	-0.02										
SLO	0.28	-0.17	0.30	-0.79	0.46	0.49											
MCUR	-0.73	0.68	-0.27	0.54	0.64												
PRCUR	-0.41	0.15	-0.42	-0.46													
PLCUR	-0.30	0.48	0.60														
TWI	-0.42	-0.46															
SPI15	0.84																
SPI20																	
SI	0.12	-0.18	-0.16	-0.03	<i>0.02</i>	0.27	0.42	0.31	-0.38	<i>0.07</i>	-0.19	-0.26	-0.33				
SOT	-0.19	-0.36	0.01	-0.01	<i>0.30</i>	0.28	0.29	0.07	<i>0.06</i>	<i>0.00</i>	0.07	-0.11					
CF	0.21	-0.06	0.03	-0.13	-0.31	-0.22	0.03	<i>0.04</i>	0.04	0.05	0.18						
HUM	-0.17	-0.06	-0.25	-0.26	-0.46	-0.09	0.04	0.06	-0.09	0.13							
TEXT	0.72	0.13	0.04	0.17	0.09	0.03	0.00	0.01	-0.06								
ETEXT	0.12	0.06	0.20	-0.03	-0.02	-0.03	<i>0.01</i>	-0.13									
DRN	0.70	0.31	-0.16	0.06	0.04	-0.25	-0.45										
SMR	0.38	-0.21	0.00	<i>0.01</i>	-0.31	-0.61											
SNR	-0.08	0.03	<i>0.02</i>	-0.04	-0.23												
ALTT	0.10	0.08	0.32	0.24													
AI0K	0.54	<i>0.05</i>	-0.03														
AI45K	0.09	-0.13															
SLOV	0.31																
SP																	

Among the ground-based measured variables only soil organic thickness did not have a significant relationship with aspen SI (Table 3.2). AI45K was selected over AI0K, while DRN was dropped due to its similarity with SMR. SMR and SNR were the most important soil variables (0.42 and 0.31 correlation coefficients, respectively), meaning that productivity was

positively related to increases in SMR and SNR. Finally, nine variables selected as potential predictors of aspen SI were: CF, HUM, ETEXT, SMR, SNR, ALTT, AI45K, SLOV, and SP.

Table 3.3. Kendall’s rank correlation coefficient matrix between pine SI and environmental variables. Bolded numbers indicate significance of relationship at $p < 0.05$ (* $p = 0.06$). Italicized numbers indicate coefficient of determination from quadratic regression.

SI	0.46	0.52	0.52	0.52	0.54	0.52	-0.24	-0.24	0.18*	<i>0.03</i>	0.41	0.27	-0.20	0.28	-0.38	0.33	0.34
DTW_0.5		0.92	0.89	0.79	0.74	0.72	-0.24	-0.01	0.08	0.02	0.72	0.31	-0.18	0.36	-0.69	0.49	0.52
DTW_1			0.94	0.82	0.79	0.76	-0.20	-0.01	0.06	-0.01	0.68	0.30	-0.16	0.34	-0.64	0.46	0.49
DTW_2				0.87	0.83	0.80	-0.18	-0.03	0.05	-0.03	0.66	0.29	-0.15	0.35	-0.62	0.45	0.48
DTW_4					0.93	0.89	-0.14	-0.08	0.09	-0.02	0.60	0.26	-0.18	0.28	-0.54	0.37	0.39
DTW_6						0.95	-0.13	-0.08	0.10	-0.02	0.59	0.28	-0.20	0.30	-0.54	0.37	0.39
DTW_10							-0.11	-0.08	0.13	0.02	0.57	0.26	-0.17	0.28	-0.52	0.35	0.37
FA								0.12	-0.12	-0.10	-0.20	-0.55	0.44	-0.50	0.31	-0.45	-0.48
ALT									-0.25	-0.24	0.05	-0.01	0.04	-0.02	-0.07	0.08	0.06
AI0										0.57	0.05	0.06	-0.05	0.07	-0.06	-0.01	0.01
AI45											0.01	-0.04	<i>0.04</i>	0.02	-0.05	0.02	0.03
SLO												0.29	-0.20	0.33	-0.80	0.57	0.58
MCUR													-0.66	0.65	-0.39	0.42	0.47
PRCUR														-0.31	0.26	-0.28	-0.32
PLCUR															-0.43	0.42	0.46
TWI																-0.56	-0.59
SPI15																	0.97
SPI20																	
SI	-0.03	<i>0.01</i>	-0.26	-0.08	-0.09	-0.18	-0.19	0.51	-0.20	0.19	<i>0.09</i>	0.34	0.22				
SOT		-0.12	<i>0.05</i>	0.12	0.09	0.26	0.40	<i>0.13</i>	0.14	-0.14	-0.09	0.02	-0.13				
CF			0.00	0.11	0.06	-0.36	-0.38	0.11	0.29	<i>0.01</i>	0.01	0.27	0.33				
HUM				0.22	0.22	0.22	<i>0.21</i>	-0.28	0.23	-0.06	-0.02	-0.02	-0.01				
TEXT					0.70	0.07	0.17	-0.01	0.15	-0.12	-0.12	-0.11	-0.07				
ETEXT						0.05	0.08	-0.03	0.14	0.04	-0.09	0.05	-0.01				
DRN							0.70	0.23	-0.18	<i>0.05</i>	0.21	-0.31	-0.54				
SMR								<i>0.19</i>	-0.01	<i>0.08</i>	0.10	-0.29	-0.70				
SNR									-0.06	0.18	0.18	0.42	0.06				
ALTT										-0.22	-0.20	0.15	0.05				
AI0K											0.55	0.08	0.07				
AI45K												0.08	-0.11				
SLOV													0.39				
SP																	

3.2.2. Lodgepole pine

All remotely sensed variables correlated well with pine SI (Table 3.3). DTW was the strongest predictor, with $r=0.54$, with minor differences in terms of FIA. TWI was excluded due to high correlation with slope, while PLCURV and SPI20 were better than other similar variables. Warm aspects, steeper slope, convex surface, higher DTW and topographic position, while lower altitude, FA, and TWI were associated with the most productive pine sites. Selected RS variables are DTW_2, FA, ALT, AI0, SLO, PLCUR, and SPI20. However, DTW and SLO are strongly correlated.

Soil characteristics which significantly explained pine productivity were SNR and humus type (Table 3.3). Pine growth was insensitive to texture, coarse fragments, and soil organic thickness. Also, the correlation between SI and SMR was not significant due to high variability of SI within well-drained mesic and submesic SMRs. All other topographic indices were good predictors of pine growth. HUM, SNR, ALTT, AI0K, SLOV, and SP were selected indicators in the regression modeling.

3.2.3. White spruce

The strongest remotely sensed variables correlated with spruce SI were DTW and altitude (Table 3.4). While the positive trend in DTW might be plausible, increasing SI with altitude was unexpected. The correlation between SI and DTW became stronger with increasing FIA. Weaker but significant nonlinear relationships were also detected for FA, AI45, SLO, and SPI20, with SI reaching a maximum on moderate FA, slope position, slope, and aspect conditions. Surface curvature indicators and TWI were not significantly related to spruce SI. Thus, RS variables chosen for modeling were DTW_10, FA, ALT, AI45, SLO, and SPI20.

Only five variables were significantly associated with spruce SI among ground-based determined variables (Table 3.4). Humus form appeared as the strongest ($r=-0.54$, $p<0.001$), then DRN from well to very poorly drained soil exhibited correlation with $r=-0.24$, while SMR, which peaked at mesic moisture conditions, was dropped due to its being strongly related to DRN. SNR was not significant, while effect of Although was significant. Effective texture and slope were weak but significant explanatory variables and hence were selected; while SOT, CF, aspect and slope

position were uncorrelated with SI. Lastly, HUM, ETEXT, DRN, ALTT, and SLOV were selected for further analysis.

Table 3.4. Kendall’s rank correlation coefficient matrix between w. spruce SI and environmental variables. Bolded numbers indicate significance of relationship at $p < 0.05$. Italicized numbers indicate coefficient of determination from quadratic regression (*exponential transformation).

SI	0.23	0.26	0.26	0.25	0.31	0.34	<i>0.09*</i>	0.25	<i>0.01</i>	<i>0.09</i>	<i>0.10</i>	<i>0.05</i>	<i>0.05</i>	-0.01	-0.12	<i>0.09</i>	<i>0.12</i>
DTW_0.5	0.89	0.80	0.78	0.72	0.69	0.69	-0.24	0.26	-0.17	-0.13	0.59	0.16	-0.11	0.13	-0.62	0.45	0.44
DTW_1	0.87	0.84	0.77	0.72	0.72	-0.22	0.28	-0.17	-0.14	0.54	0.17	-0.10	0.13	-0.57	0.40	0.39	
DTW_2	0.96	0.88	0.82	0.82	0.82	-0.16	0.25	-0.14	-0.10	0.47	0.20	-0.15	0.14	-0.52	0.42	0.41	
DTW_4	0.90	0.83	0.83	0.83	0.83	-0.15	0.23	-0.13	-0.09	0.46	0.19	-0.14	0.14	-0.51	0.40	0.39	
DTW_6	0.93	0.83	0.82	0.82	0.82	-0.14	0.28	-0.09	-0.10	0.44	0.18	-0.13	0.13	-0.50	0.39	0.38	
DTW_10	0.93	0.83	0.82	0.82	0.82	-0.14	0.32	-0.07	-0.06	0.43	0.15	-0.10	0.13	-0.48	0.40	0.38	
FA	-0.12	0.00	-0.03	-0.17	-0.40	0.25	-0.31	0.28	-0.33	-0.34							
ALT	-0.21	-0.30	<i>0.06</i>	0.02	-0.06	-0.07	-0.19	<i>0.12</i>	0.15								
AIO	0.56	-0.26	-0.05	0.03	-0.02	0.25	-0.16	-0.14									
AI45	-0.17	-0.02	-0.01	0.01	0.17	-0.11	-0.07										
SLO	<i>0.09</i>	<i>0.08</i>	0.22	-0.82	<i>0.51</i>	0.51											
MCUR	-0.65	0.55	-0.25	0.48	0.48												
PRCUR	<i>0.23</i>	0.17	-0.37	-0.37													
PLCUR	-0.23	0.40	0.44														
TWI	-0.59	-0.59															
SPI15	0.95																
SPI20																	
SI	-0.03	-0.01	-0.54	<i>0.04</i>	0.05	-0.28	<i>0.18</i>	0.12	0.26	-0.08	-0.16	<i>0.10</i>	<i>0.05</i>				
SOT	-0.19	-0.21	0.07	0.09	0.32	0.45	0.49	0.02	0.14	-0.03	-0.03	-0.51	-0.55				
CF	0.06	-0.26	-0.18	-0.19	-0.29	-0.45	0.11	-0.05	0.00	0.22	0.22						
HUM	-0.11	-0.18	0.13	<i>0.18</i>	-0.34	-0.22	-0.03	0.21	0.07	0.06							
TEXT	0.63	0.08	0.12	0.11	0.03	-0.13	-0.22	-0.02	0.01								
ETEXT	0.35	<i>0.33</i>	<i>0.38</i>	-0.35	-0.04	-0.18	-0.15	-0.25									
DRN	0.84	0.34	-0.46	0.19	0.25	-0.36	-0.58										
SMR	0.49	-0.39	0.17	0.15	-0.48	-0.68											
SNR	-0.20	0.17	0.02	-0.46	-0.49												
ALTT	-0.19	-0.25	0.19	0.28													
AI0K	0.56	-0.25	-0.17														
AI45K	-0.15	-0.07															
SLOV	<i>0.40</i>																
SP																	

3.3. Depth-To-Water (DTW) index selection

In order to select optimal DTW for modeling SI, six different Flow Initiation Areas (0.5, 1, 2, 4, 6, and 10 ha) were tested using two statistical methods for the three investigated species. GAM models were fit with an upper limit of 3 degrees of freedom for aspen and spruce, and 4 for pine. Absolute and relative goodness-of-fit measures were applied in selection of the best model for each species and statistical technique (Table 3.5). All selected models passed validation graphically based on residual and actual vs. fitted plots (Figure 3.12) and using the Shapiro-Wilk test on residuals ($p > 0.05$).

Table 3.5. Fit statistics for SI models based on DTW calculated at six different WAM's Flow Initiation Area (FIA) by species. Models were fitted using Generalized Additive Models (GAM) and Linear Regression (LR) approaches. Abbreviations: edf - effective degrees of freedom; RMSE - root mean squared error; R^2 - coefficient of determination (deviance explained); AIC - Akaike information criteria; Significance codes - *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. $p > 0.05$. The best models are shown in bold.

FIA	edf	RMSE (m)	R^2	AIC	Differences	RMSE (m)	R^2	AIC	Model significance
Aspen									
	Fit: GAM (3 edf max)					Fit: LR (DTW no transf.)			
0.5 ha	1.56***	2.17	0.38	433.0		2.18	0.38	432.9	***
1 ha	2.75***	2.08	0.44	427.0	>*	2.16	0.39	430.8	***
2 ha	1.53***	2.25	0.34	439.4		2.25	0.34	438.5	***
4 ha	1.90***	2.33	0.29	447.4		2.33	0.29	445.7	***
6 ha	1.98***	2.32	0.30	446.6		2.34	0.28	446.4	***
10 ha	2.58***	2.28	0.32	443.9		2.34	0.28	446.6	***
L. pine									
	Fit: GAM (4 edf max)					Fit: LR (DTW log transf.)			
0.5 ha	3.87***	2.12	0.50	228.6		2.00	0.55	216.9	***
1 ha	3.90***	1.94	0.58	219.8		1.84	0.62	209.1	***
2 ha	3.90***	1.90	0.60	218.0	n.s.	1.81	0.63	207.2	***
4 ha	3.86***	1.98	0.56	222.2		1.86	0.61	209.8	***
6 ha	3.87***	1.95	0.57	220.6		1.83	0.63	208.2	***
10 ha	3.89***	2.00	0.55	222.8		1.90	0.60	212.2	***
W. spruce									
	Fit: GAM (3 edf max)					Fit: LR (DTW log transf.)			
0.5 ha	1.32 ^{n.s.}	2.50	0.10	216.8		2.40	0.17	212.5	**
1 ha	1.60*	2.39	0.18	213.5		2.33	0.22	209.8	**
2 ha	1.24 ^{n.s.}	2.45	0.14	214.7		2.33	0.22	210.0	**
4 ha	1.22 ^{n.s.}	2.47	0.13	215.4		2.35	0.20	210.7	**
6 ha	1.10*	2.37	0.19	211.7		2.30	0.24	208.6	***
10 ha	2.58**	2.26	0.26	210.3	n.s.	2.25	0.27	206.5	***

Aspen SI models had the highest R^2 and lowest AIC when DTW was calculated using 1 ha FIA threshold (Table 3.5). Comparing fit statistics for different values of FIA a similar pattern can be noticed with both GAM and LR. Model performance peaked at 1 ha FIA, declined slightly at 0.5 ha, and declined rapidly to 4 ha, while no significant changes occurred with FIA above 4 ha. The GAM 10 ha model had slightly improved performance due to inflated degree of smoothness. A chi-square test confirmed that GAM outperformed the LM model with 1 ha flow initiation area. In addition, GAM provided a fit that was more ecologically plausible than the linearly declining fit (Figures 3.10-3.11). The non-linear trend in GAM model peaks at around 1.95 m DTW_1 with SI of 22.2 m. Log transformation of SI yielded R-squared values for both methods to 0.48 of GAM and 0.43 for LR, however all comparisons and trends remained fairly similar.

Pine SI prediction models using DTW calculated at a 2 ha FIA worked best for both GAM ($R^2=0.60$) and LR ($R^2=0.63$) models, with no significant difference between these two models (Table 3.5). Similarly, the strength of the relationship for other FIAs was almost identical for both methods. Despite the weaker models based on 0.5 ha FIA threshold area, within the range

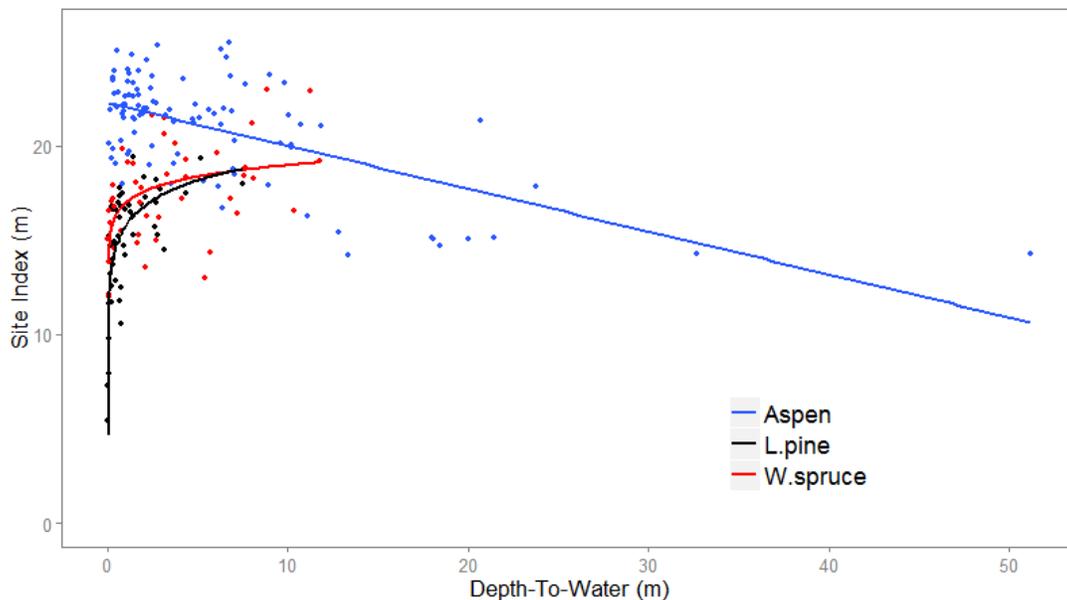


Figure 3.10. Relationship between SI and selected DTW index fitted linear function for aspen and logarithmic function for pine and spruce. DTW is calculated at FIA of 1 ha for aspen, 2 ha for pine, and 10 ha for spruce.

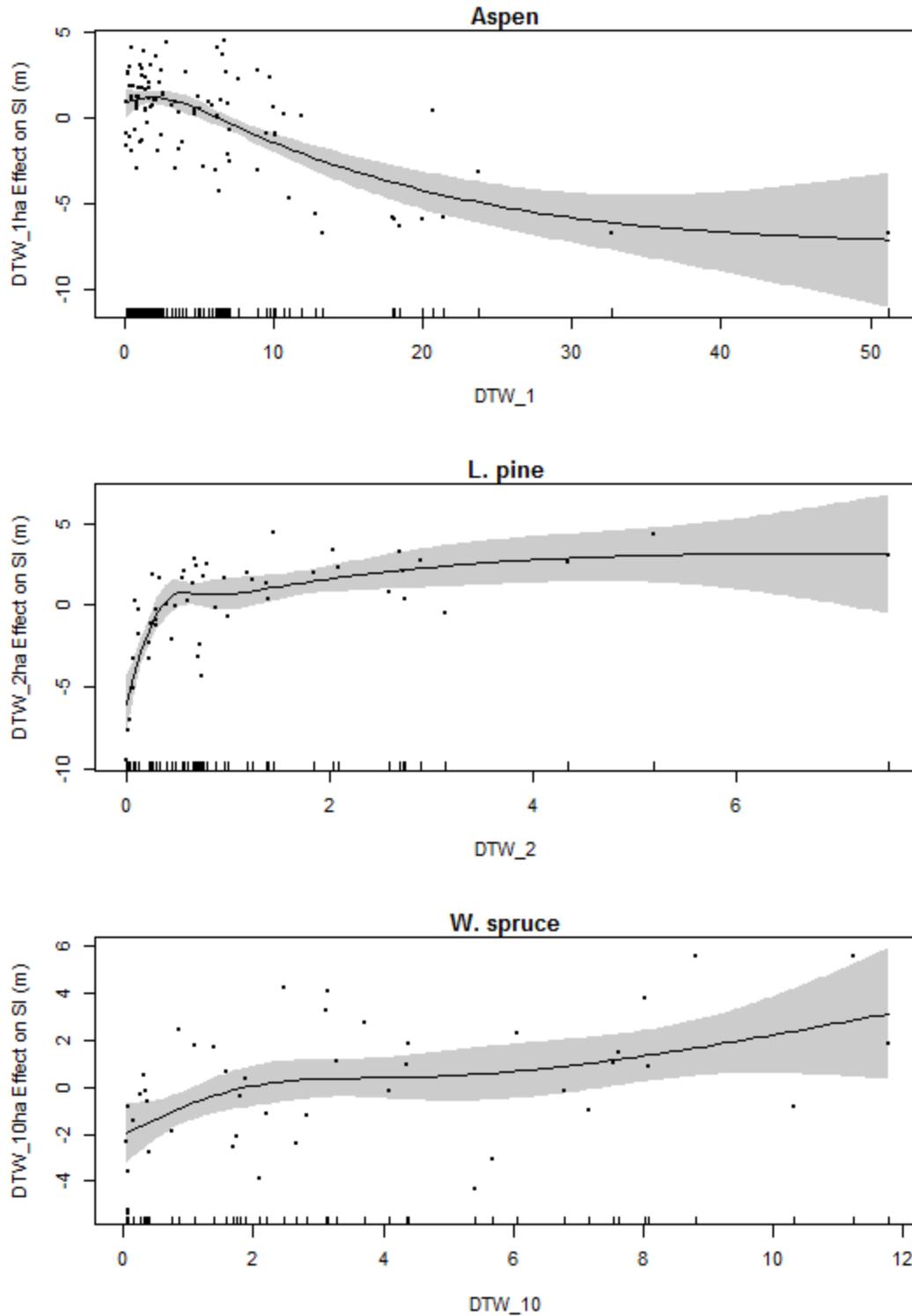


Figure 3.11. Response curves (GAM method) to selected DTW index for each species. The y-axis shows effect on SI with values centered on zero. "Rug" lines along x-axis show covariate values, dots represent residuals. Shading shows confidence bands (± 2 se) of the estimates.

from 1 ha to 10 ha fluctuations in model performance were non-significant. A non-linear relationship between SI and DTW was evident. Therefore, logarithmic transformation applied on DTW prior to LR analysis, and with a maximum possible edf of 4 resulted in a better fit for the cubic regression spline in GAM analysis. Hence both approaches exhibited, to some extent, similar results (Figures 3.10-3.11). In the GAM model a steep positive slope occurred up to 0.6 m DTW_2 reaching a plateau at 15.6 m SI, then gradual increases between 0.9 to 4 m DTW_2 and then levelling-off at around 18.0 m. The logarithmic fit demonstrated a progressive decrease in slope reaching SI 12.4 m at 0.1 m DTW_2, with SI 14.8 m at 0.5 m DTW_2, SI 15.8 m at 1 m DTW_2, SI 16.8 m at 2 m DTW_2, and a SI of 18.8 m at the end of the range. Log transforming SI resulted in an R-squared to 0.62 for the GAM model and 0.69 for the LR model based on DTW_2.

Spruce SI was found to be best estimated using DTW at the largest tested catchment area of 10 ha (Table 3.5). DTW_10 accounted for 27% and 26% of variation in SI based on a LR, with log transformation of DTW, and GAM respectively, and were not statistically different. Outputs from GAM SI models based on DTW_0.5, DTW_2, and DTW_4 were unreliable due to non-significant parameters. Comparison of fit statistics for different FIAs suggest model improvement with increasing FIA. However, both fitted curves (LR and GAM) showed a positive relationship but with different patterns (Figures 3.10-3.11). The GAM SI model sharply increased up to 2 m DTW_10, then remained constant at about 17.8 m SI, and then gradually increased after 5 m DTW_10. The positive logarithmic shape of the SI curve indicates a gradual change in slope, with the rise in SI being rapid from 14 to 17 m for DTW_10 up to 1 m, then a slower rise of 1 m SI at DTW_10 between 1-3 m, reaching 19 m SI at 11 m DTW_10.

Evaluation of both modeling approaches suggested that resulting models showed very similar results and confirmed the strength and nature of relationships between SI and DTW. Also, both methods provided similar inference for each species in terms of predictive capacity of DTW with different FIA. Firstly, a decline could be noticed from 1 ha to 0.5 ha FIA for all species suggesting that 0.5 ha FIA is unnecessarily small. Secondly, R² from 1 ha to 10 ha FIA exhibited different trends for each species. For aspen SI the smaller FIA worked better, for spruce the largest FIA worked best, while pine seems insensitive to changes in size of the FIA.

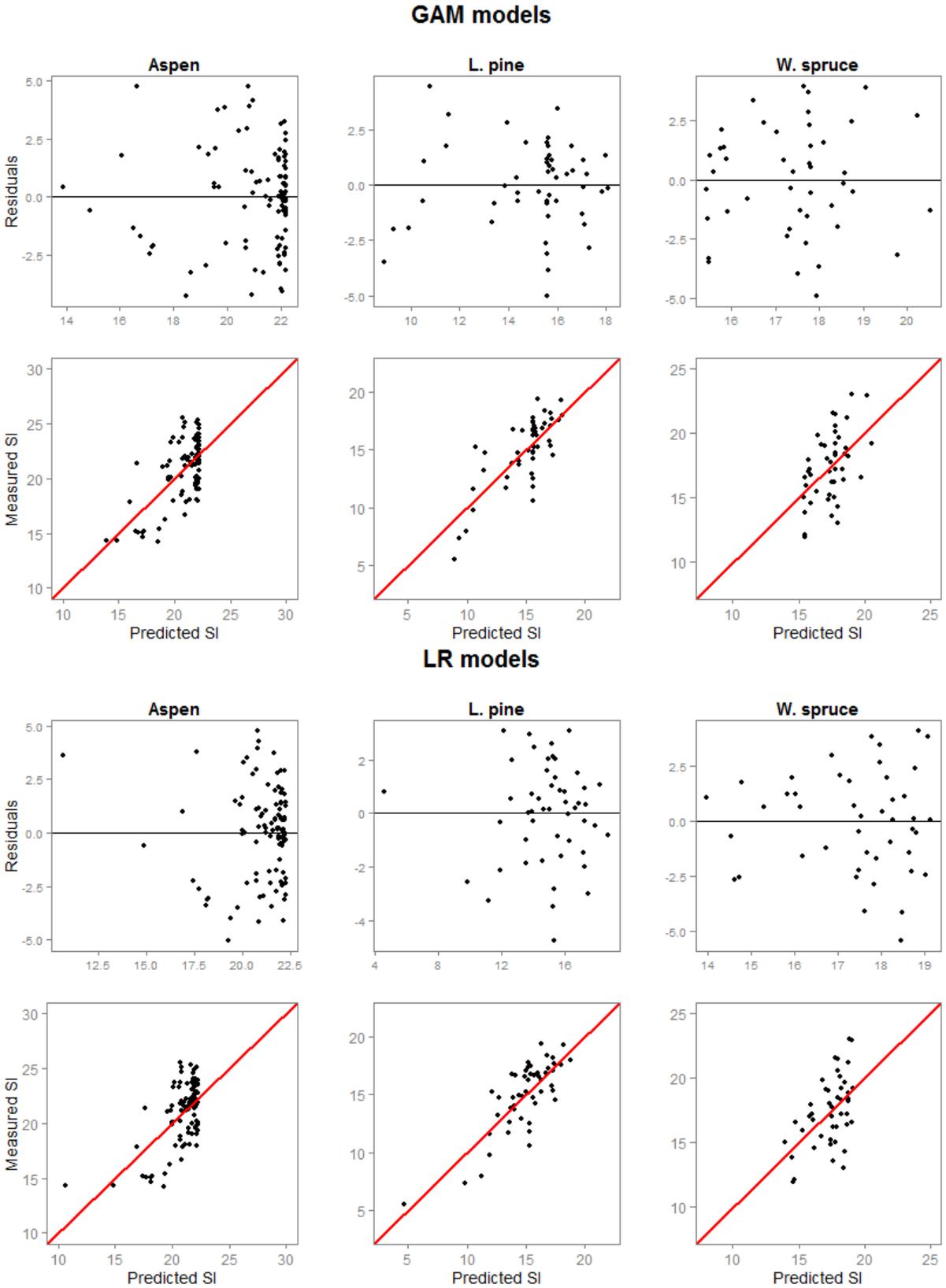


Figure 3.12. Scatterplots of residuals vs. predicted SI values and measured vs. predicted SI values with 1:1 line (red color) by species and statistical method of SI models based on selected DTW index.

3.4. Modeling Site Index variation

Species-specific Site Index models were developed by applying four different statistical methods. In addition, models were built from three different data sources: (1) remotely sensed variables (RS), (2) only WAM variables, (3) ground-based measured variables (GB). Backward variable selection, maximizing variance explained and/or minimizing BIC and RMSE, was employed to select optimal models for each of statistical approaches and data sources. For covariates selected to be considered in modeling refer to section 3.2. and 3.3. Assumptions of homoscedasticity and normality of residuals were evaluated visually on graphs and using the Shapiro-Wilk test ($p > 0.05$) for each selected model. No collinearity issues between variables in selected MLR and GAM models were observed with $VIF < 5$ for each covariate.

3.4.1. Trembling aspen

Multiple Linear Regression

Selected optimal models predicting SI from different data sources are presented in Table 3.6. The FA was linearized using a log function. Among remotely sensed variables only SPI20 could not be fitted in the model with significant ($p < 0.05$) slope parameter and was dropped, subsequently models with the best possible combination of independent variables for particular numbers of variables were examined (Figure 3.13). Ecologically plausible interactions were tested but they were either not significant or did not improve model performance. The final SI prediction model ($p < 0.001$; $RMSE = 1.63$ m; $adj.R^2 = 0.64$) developed from remotely sensed data included five variables which were all highly significant and with biologically plausible behaviour. Aspen productivity was highest at low DTW values, high FA, lower altitude, northeast aspect, and low surface curvature. According to partial R^2 DTW was the most important variable in explaining SI variation, then aspect and curvature, while FA and altitude were weaker predictors. The model that contained just WAM variables (Figure 3.14) also had significant fit with $RMSE = 1.91$ m and $adj.R^2 = 0.51$. Log transformation of SI improved on the amount of variance explained for both RS ($adj.R^2 = 0.66$) and WAM ($adj.R^2 = 0.54$) models.

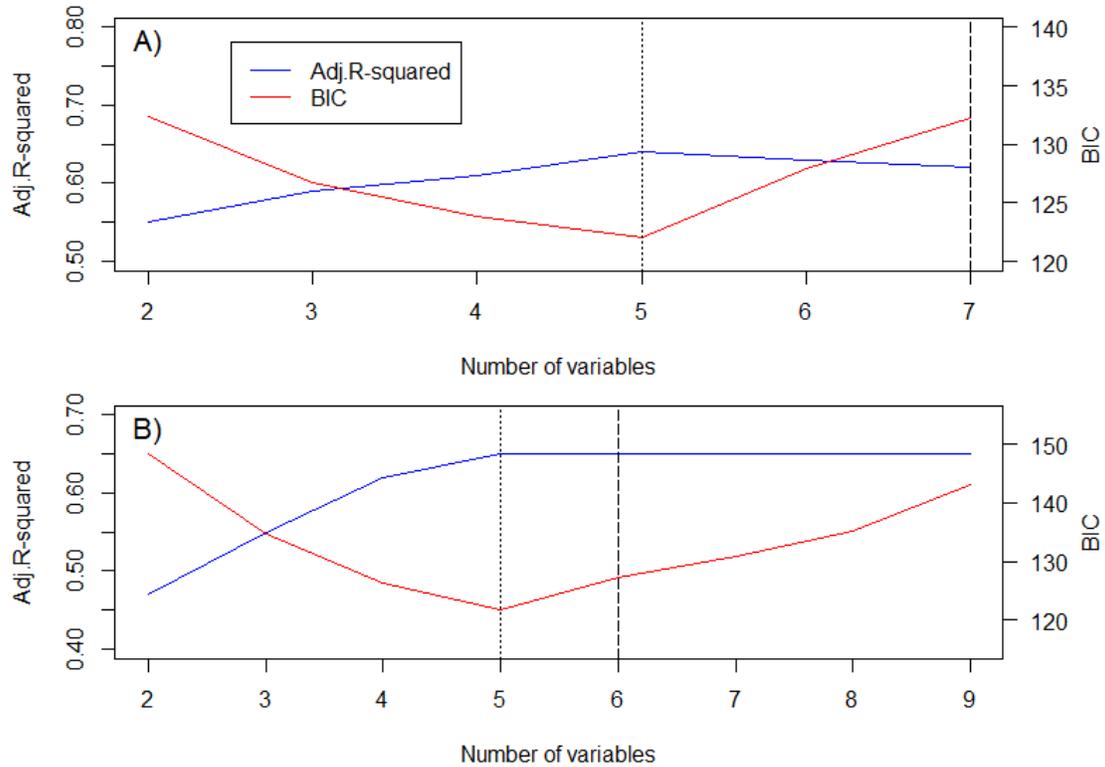


Figure 3.13. Decision support plots for aspen MLR model selection based on adj.R² and BIC statistics; A) remotely sensed variables, B) ground-based measured variables. Adj.R² and BIC are given for the best fitted model for particular number of variables included. Dotted line indicates selected model. Dashed line means that model (and all models to the right) includes at least one non significant variable.

Soil properties (CF, HUM, ETEXT, SMR) provided unreliable ($p > 0.05$) parameters in the SI model, only SNR was retained in the final model but accounted relatively little for SI variation compared to topographic indices. Although SMR had the strongest bivariate correlation with SI, it was dropped in modeling due to high correlation with SP ($r = -0.61$). SP was transformed using a second order polynomial function. The best fitted model, according to adj.R² and BIC statistics (Figure 3.13), contained SNR, ALTT, AI45K, SLOV, and SP ($p < 0.001$; RMSE=1.59 m; adj.R²=0.65), with the fit statistics being almost identical to RS model. Rich soil nutrient regime, lower altitude, northeast aspect, no inclination, and lower slope to level topographic position were associated with very productive aspen stands. Variable importance measured by partial R² showed slope and slope position as stronger predictors. The GB model yielded slightly better results (adj.R²=0.66) when SI was log transformed.

Table 3.6. Selected MLR SI prediction models for aspen using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model).

	Model	RMSE (m)	Adj. R ²	p-value
<u>RS</u>	$SI=34.544-0.136*DTW_1+0.407*\log(FA)-0.015*ALT-0.014*AI45-1.243*PLCUR$	1.63	0.64	<0.001
<u>WAM</u>	$SI=19.682-0.207*DTW_1+0.794*\log(FA)$	1.91	0.51	<0.001
<u>GB</u>	$SI=33.612+1.145*SNR-0.021*ALTT-0.010*AI45K-0.062*SLOV+1.932*SP-0.302*SP^2$	1.59	0.65	<0.001

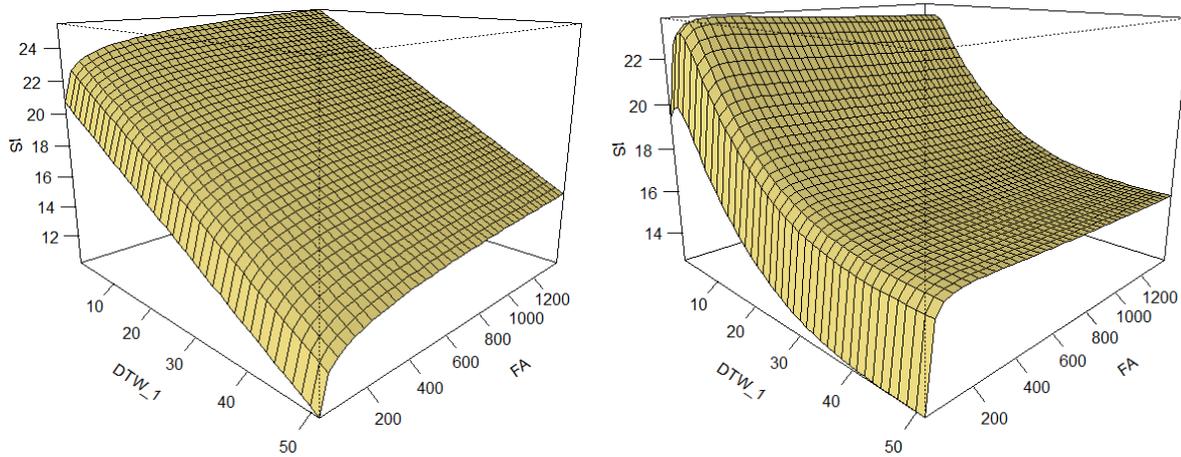


Figure 3.14. Selected aspen SI prediction models based on WAM variables using MLR method (left plot) and GAM method (right plot). For model description refer to Table 3.6 and Table 3.7.

Generalized Additive Models

The SI models selected as optimal for each data source are given in Table 3.7. ALT, PLCUR and SLO were fitted as linear terms, and DTW_1, AI45 and SPI20 applied as smooth parameters with max 3 edf (the number of knots in the splines), while FA was used untransformed and allowed a max 4 edf to capture the nonlinearity. However, after excluding non-significant parameters, the best selected model (RMSE=1.54 m; adj.R²=0.66) had four included variables (Figure 3.15). Partial response curves for each variable were as expected in terms of aspen ecology (Figure 3.16). Aspect had a negative linear relationship with SI from NE to SW, DTW_1 peaked at around 3 m and decreased with excess moisture or moisture deficit and levelled-off at higher than 30 m values, FA had positive and logarithmic curve shape with a slight decrease at high values, and SI declined on midslope and upper sites. DTW_1 was the dominant covariate in

explaining SI variation in terms of relative variable importance. The WAM model (Figure 3.14) also provided strong and significant parameters for explaining aspen SI (RMSE=1.70 m; adj.R²=0.60). When the log function was applied to SI in selected models adj.R² shifted to 0.69 and 0.63 for RS and GAM models, respectively.

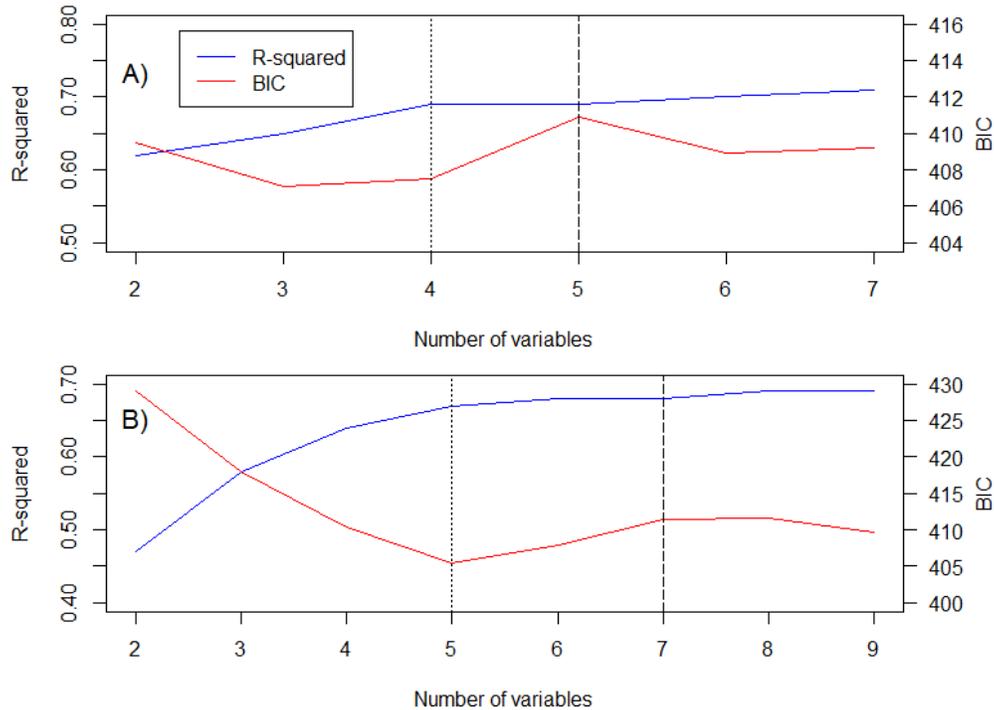


Figure 3.15. Decision support plots for aspen GAM model selection based on R² and BIC statistics; A) remotely sensed variables, B) ground-based measured variables. R² and BIC are given for the best fitted model for particular number of variables included. Dotted line indicates selected model. Dashed line means that model (and all models to the right) includes at least one non significant variable.

Similarly to the MLR model, all soil variables, except SNR, appeared as nonsignificant in the GAM model. The SI GAM model developed from ground-based measured variables (RMSE=1.59 m; adj.R²=0.65) included the same variables and accounted the same amount of variation as the MLR model. The form of relationship for all fitted variables was also fairly similar to that in the MLR model. Non-linear behaviour of slope and slope position are shown in the Figure 3.16 SI log transformation resulted in only a small change in variance explained (adj.R²=0.66).

Table 3.7. Selected GAM SI prediction models for aspen using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model).

	Model	RMSE (m)	Adj. R ²	
	$SI = \beta_0 + f_1(DTW_1) + f_2(FA) + \beta_1 AI45 + f_3(SPI20) + \epsilon$	1.54	0.66	
<u>RS</u>	Parametric coefficients (β):	Smooth terms (f):		
	Intercept	Estimate (\pm SE)	edf	
	AI45	21.995 (± 0.408)***	DTW_1	2.884***
		-0.012 (± 0.004)**	FA	2.822*
			SPI20	1.881***
	$SI = \beta_0 + f_1(DTW_1) + f_2(\log(FA)) + \epsilon$	1.70	0.60	
<u>WAM</u>	Parametric coefficients (β):	Smooth terms (f):		
	Intercept	20.972 (± 0.179)***	DTW_1	2.838***
			log(FA)	2.048***
	$SI = \beta_0 + \beta_1 SNR + \beta_2 ALTT + \beta_3 AI45K + f_1(SLOV) + f_2(SP) + \epsilon$	1.59	0.65	
<u>GB</u>	Parametric coefficients (β):	Smooth terms (f):		
	Intercept	35.857 (± 4.257)***	SLOV	1.730***
	SNR	1.171 (± 0.317)***	SP	2.120***
	ALTT	-0.022 (± 0.005)***		
	AI45K	-0.010 (± 0.003)**		

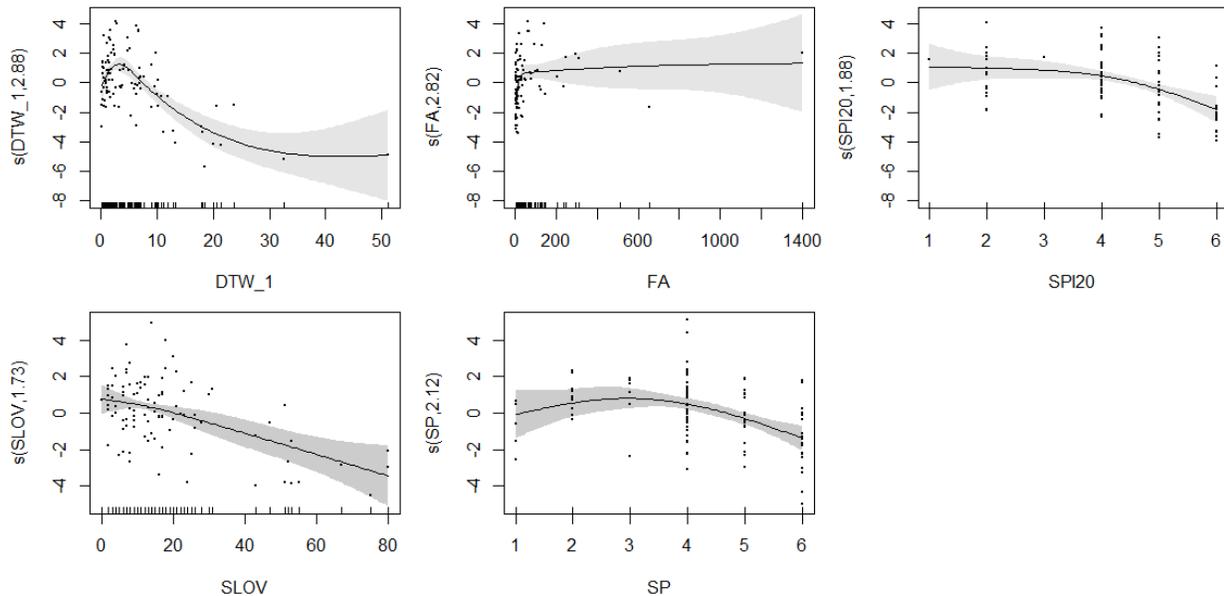


Figure 3.16. Partial response curves of aspen SI to selected smooth terms in RS (upper plots) and GB (bottom plots) models from Table 3.6. The y-axis shows effect on SI with values centered on zero, with label indicating smooth term and resulting edf. "Rug" lines along x-axis indicating covariate values, dots represent residuals. Shading shows confidence bands (± 2 se) of the estimates.

Regression Trees

The RT models were optimized by pruning the trees to number of splits determined as maximizing the variance explained after applying cross-validation (Figure 3.17), and are presented in Table 3.8 and Figure 3.18. The final RS model (RMSE=1.47 m; adj.R²=0.70) used five variables in modeling SI variation. The primary split (R²=0.47), introducing the most important variable in the model, was based on DTW_1 with the threshold of 10.9 m indicating lower productive sites at higher DTW values, and vice versa. The model predicts the lowest productivity of SI of 15 m at plots with high DTW and low FA (FA<13), while the most productive sites (23.7 m SI) were found at plots with low DTW, PLCUR<0.5, and lower altitude. Remaining terminal nodes predict intermediate SI classes using DTW_1 and one or two of other variables (FA, PLCUR, SLO, ALT), but with uneven class distribution across SI range. The WAM model had the same primary split and left part of the tree as the RS model and explained just slightly less variation (RMSE=1.55 m; adj.R²=0.68) than the RS model. This is probably due to WAM variables being strongly correlated with other RS variables.

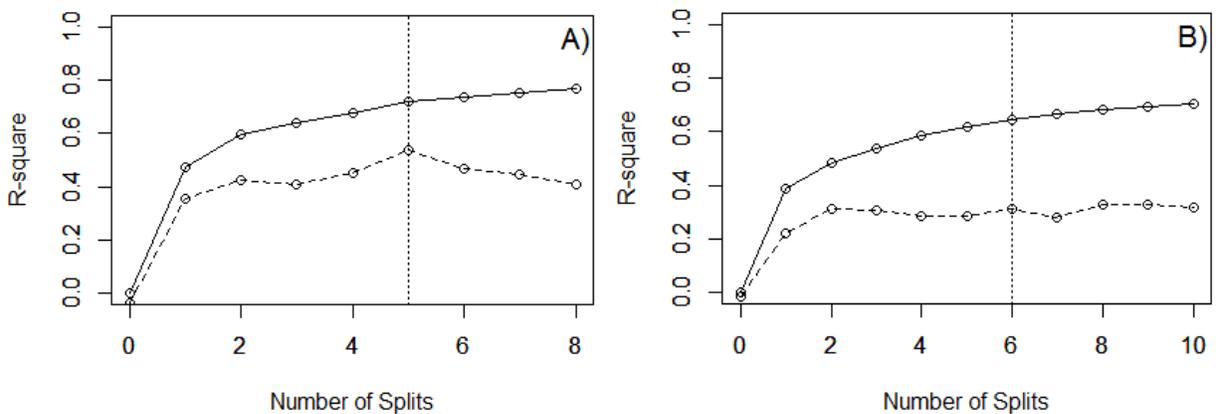


Figure 3.17. Decision support plots for determining the pruning RT for aspen: A) remotely sensed variables, B) ground-based measured variables. Solid line plots the variation explained (R-squared) by the split in the model. Dashed line plots variation explained (R-squared) by the split from cross-validation (CV). Dotted line indicates the point of pruning (No of splits left) to optimal RT based on maximizing CV R-squared.

Aspen SI for the GB model (RMSE=1.65 m; adj.R²=0.63) includes four topographic variables. Splits based on soil variables were pruned out because they resulted in overfitting the model. Slope value of 37% defined the primary split explaining the largest amount of variation in SI (38%). In addition, alteration of slope position, altitude and aspect explained seven SI classes

predicted in terminal nodes. SI was highest (23.4 m) in cases when slope was lower than 37%, slope position was not crest, and altitude was lower than 770 m, while SI was lowest (15.6 m) on slopes steeper than 51.5%.

Table 3.8. Fit statistics and models rules in explaining aspen SI for selected RT models using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model). The Terminal node numbering refers to corresponding models in Figure 3.18.

Model	RMSE (m)	Adj.R ²	Terminal node	Number of obs.	Node rules	Predict SI (m)
<u>RS</u>	1.47	0.70		97		
			Node 3	8	(DTW_1 \geq 10.9) and (FA $<$ 12.9)	14.97
			Node 4	5	(DTW_1 \geq 10.9) and (FA \geq 12.9)	18.02
			Node 7	11	(DTW_1 $<$ 10.9) and (PLCUR \geq 0.5) and (SLO \geq 17.7)	18.82
			Node 8	13	(DTW_1 $<$ 10.9) and (PLCUR \geq 0.5) and (SLO $<$ 17.7)	21.09
			Node 10	47	(DTW_1 $<$ 10.9) and (PLCUR $<$ 0.5) and (ALT \geq 774.1)	22.02
			Node 11	13	(DTW_1 $<$ 10.9) and (PLCUR $<$ 0.5) and (ALT $<$ 774.1)	23.73
<u>WAM</u>	1.55	0.68		97		
			Node 3	8	(DTW_1 \geq 10.9) and (FA $<$ 12.9)	14.97
			Node 4	5	(DTW_1 \geq 10.9) and (FA \geq 12.9)	18.02
			Node 7	11	DTW_1 $<$ 10.9) and (FA $<$ 16) and (DTW_1 \geq 5.2)	18.82
			Node 8	13	DTW_1 $<$ 10.9) and (FA $<$ 16) and (DTW_1 $<$ 5.2)	21.23
			Node 9	60	DTW_1 $<$ 10.9) and (FA \geq 16)	22.68
<u>GB</u>	1.65	0.63		97		
			Node 3	8	(SLOV \geq 51.5)	15.59
			Node 4	5	(37 \leq SLOV $<$ 51.5)	18.24
			Node 7	12	(SLOV $<$ 37) and (SP \geq 5.5) and (ALTT \geq 771.6)	18.79
			Node 11	20	(SLOV $<$ 37) and (SP $<$ 5.5) and (ALTT \geq 770.1) and (AI45K \geq 107.5)	20.87
			Node 8	5	(SLOV $<$ 37) and (SP \geq 5.5) and (ALTT $<$ 771.6)	22.22
			Node 12	31	(SLOV $<$ 37) and (SP $<$ 5.5) and (ALTT \geq 770.1) and (AI45K $<$ 107.5)	22.27
			Node 13	16	(SLOV $<$ 37) and (SP $<$ 5.5) and (ALTT $<$ 770.1)	23.38

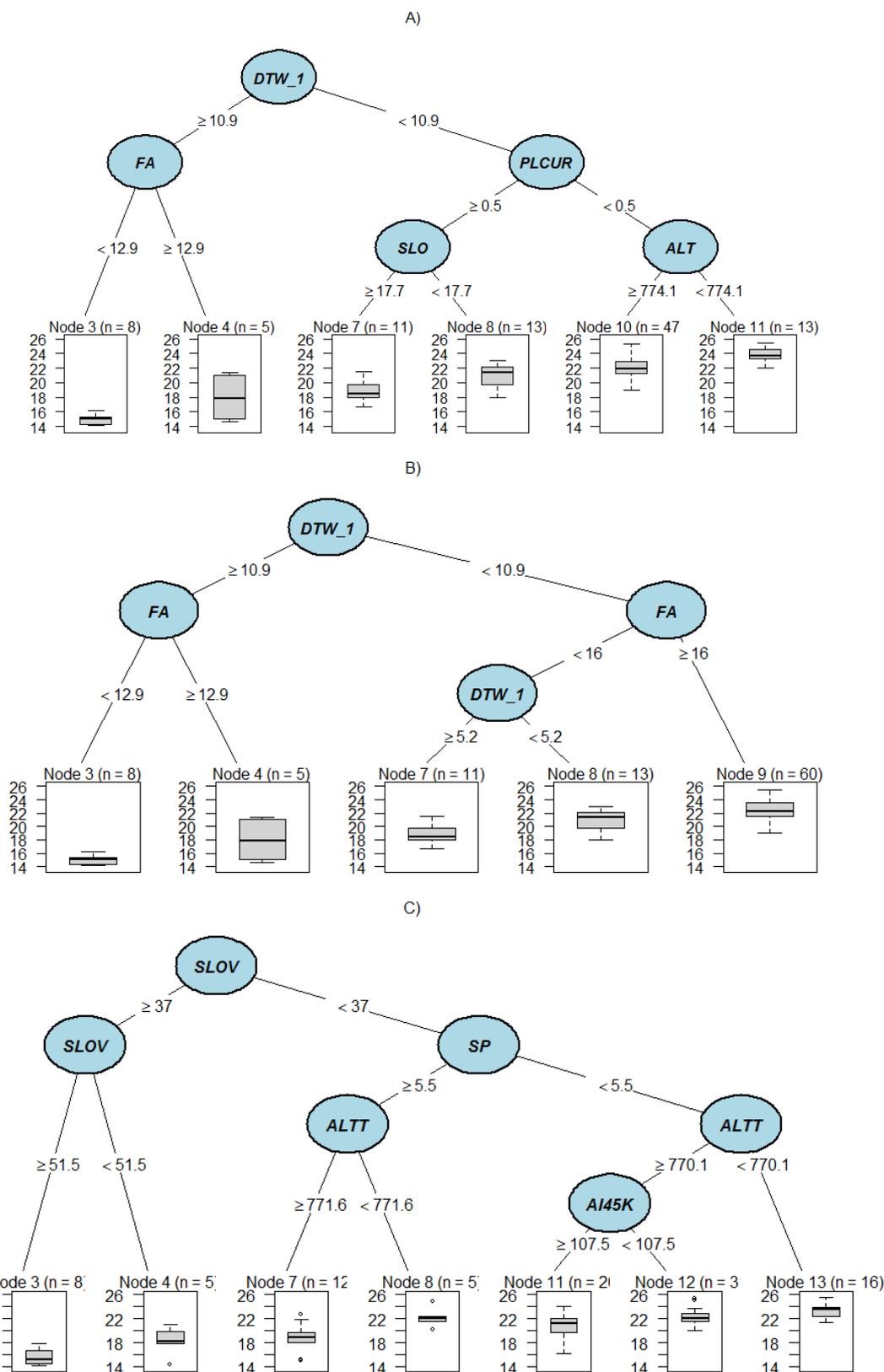


Figure 3.18. Aspen SI Regression Tree models; A) remotely sensed variables (RS model), B) only WAM variables (WAM model), C) ground-based measured variables (GB model).

Random Forest

Aspen SI models were fitted for both datasets, remotely sensed and ground-based, including all potential predictors with measures of importance computed for each variable in the "full" model (Figure 3.19). Optimal models were chosen by backward selection maximizing R^2 and minimizing RMSE (Figure 3.20).

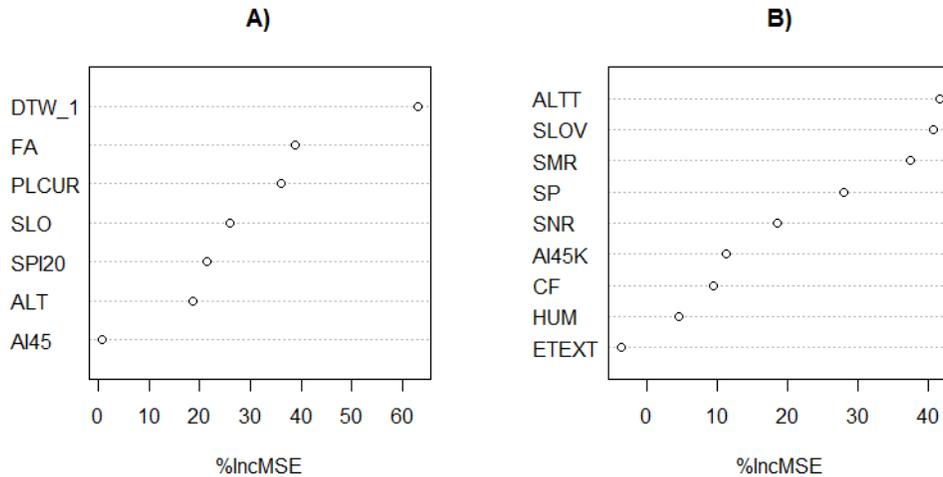


Figure 3.19. Variable importance plots for aspen SI Random Forest models built with all potential explanatory variables; A) remotely sensed variables, B) ground-based measured variables. %IncMSE - variable importance measure indicating average percent change in the mean square error when the particular variable is permuted while all others are retained unchanged. Important variables should cause relatively large %IncMSE.

Basic fit statistics ($\text{adj.}R^2$ and RMSE) and variable importance for resulting optimal SI models for each of the data sources are shown in Table 3.9. The RS model (RMSE=1.74 m; $\text{adj.}R^2=0.59$) consisted of DTW_1, FA and SPI20 with the general trends of partial effects on SI being similar to the fit of the same variables in the aspen GAM model (Figure 3.21). DTW appeared as the most influential variable in the model. The WAM model accounted for slightly less variation in SI (RMSE=1.82 m; $\text{adj.}R^2=0.56$) with regard to RS model. Finally, the GB model (RMSE=1.83 m; $\text{adj.}R^2=0.55$) included three topographic indices, and SMR as the best variables for explaining SI variation. Trends in the partial effects of topographic variables on SI were plausible and consistent with the fits obtained using other methods. SMR was positively related with SI reaching maximum under a subhygric soil moisture regime.

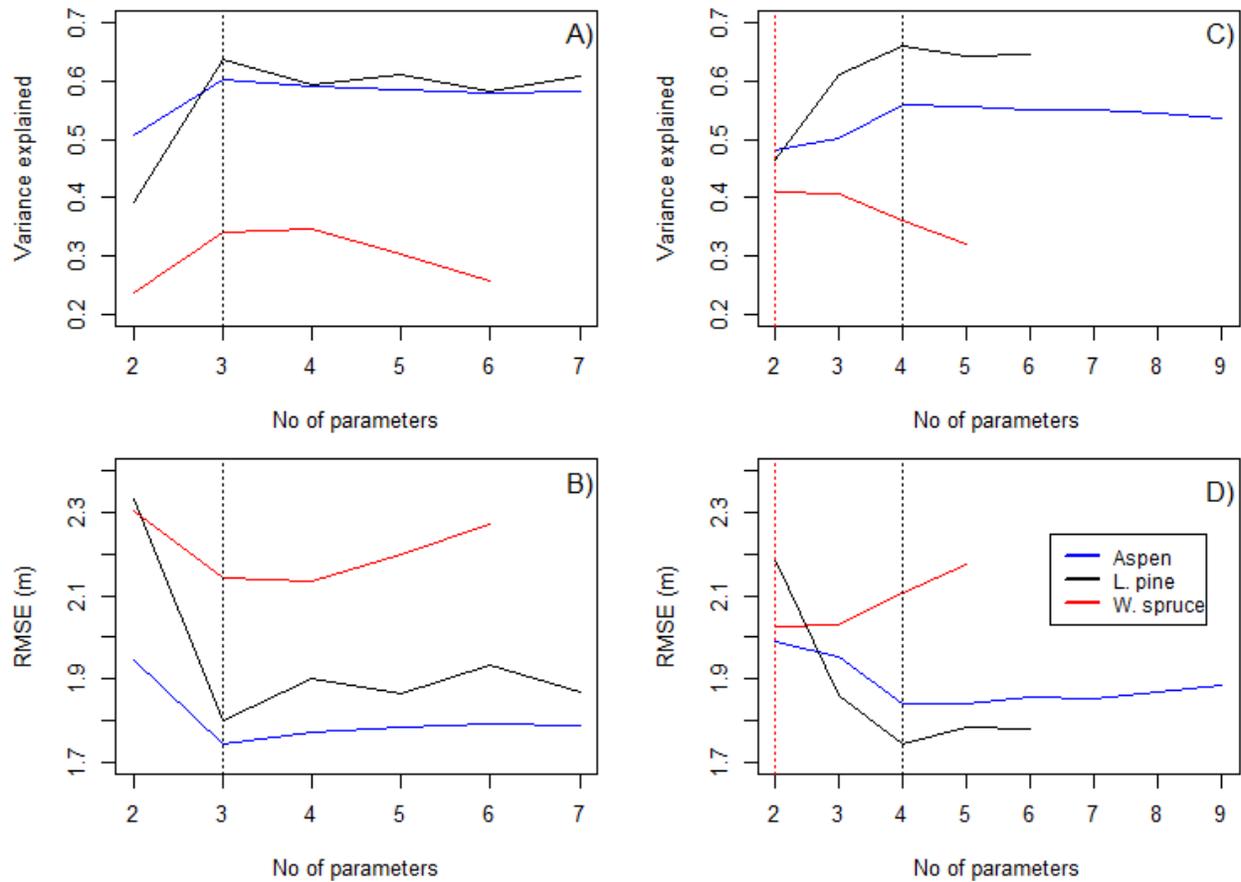


Figure 3.20. Decision support plots based on R^2 and RMSE for selected SI models from remotely sensed variables (plots A and B) and ground-based measured variables (plots C and D). R^2 and RMSE are given for the best fitted model for particular number of variables included. Dotted line indicates selected model (black color - for aspen, pine and RS spruce; red color for GB spruce).

Table 3.9. Fit statistics and variable importance for selected RF models explaining aspen SI based on remotely sensed variables (RS model), WAM variables (WAM model), and ground-based measured variables (GB model). Adj. R^2 - calculated using "pseudo R^2 " (validation samples); %IncMSE - variable importance measure indicating average percent change in the mean square error when the particular variable is permuted while all others are retained unchanged.

	RS		WAM		GB	
	Variable	%IncMSE	Variable	%IncMSE	Variable	%IncMSE
	DTW_1	65.9	DTW_1	95.1	SMR	51.1
	FA	51.9	FA	70.8	SLOV	46.9
	SPI20	56.1			ALTT	45.2
					SP	42.0
Adj. R^2	0.59		0.56		0.55	
RMSE (m)	1.74		1.82		1.83	

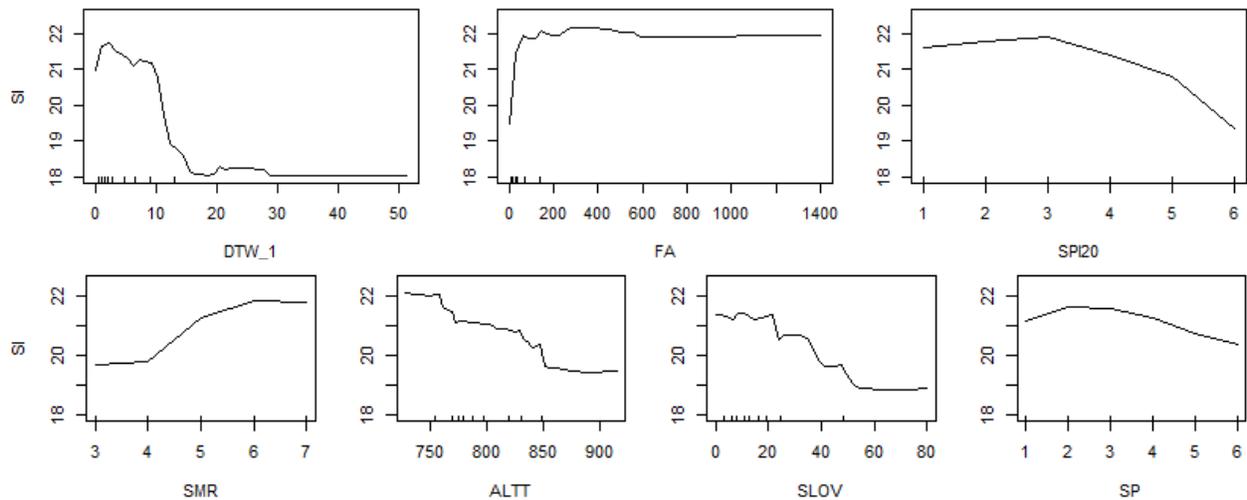


Figure 3.21. Partial dependence plots showing effect of selected variables on aspen SI in RS (upper plots) and GB (bottom plots) models. Hush marks at the bottom of the plot show the deciles of the variable.

3.4.2. Lodgepole pine

Multiple Linear Regression

The best MLR SI models for pine were selected based on significance of variables included and model performance measures (Figure 3.22 and Table 3.10). Necessary transformations on DTW_2, FA and SLO were applied prior to model fitting. The RS model with only two explanatory variables, including DTW_2 and one of FA, AI0 or ALT, were significant ($p < 0.05$). No significant interactions were found. The selected RS model ($p < 0.001$; RMSE=1.54 m; $\text{adj.}R^2=0.72$) showed a sharp decline in SI as DTW_2 decreases from 2 to 0 m, and relatively lower influence of altitude with negative linear response in SI (Figure 3.23). The WAM model (Figure 3.23) explained slightly less variance ($p < 0.001$; RMSE=1.73 m; $\text{adj.}R^2=0.65$) than the RS model, with a similar strong effect of DTW_2 and marginal effect of FA. Altitude, slope position and SNR were selected predictors in the optimal GB model ($p < 0.001$; RMSE=1.69 m; $\text{adj.}R^2=0.66$), while aspect, slope and humus form provided unreliable ($p > 0.05$) parameter estimates. Pine productivity declined on poor nutrient sites, lower altitude, and lower (lower slope, level and depression) topographic positions. Applying logarithmic transformation on SI resulted in notable improvements for both the RS ($\text{adj.}R^2=0.77$) and WAM ($\text{adj.}R^2=0.71$) models.

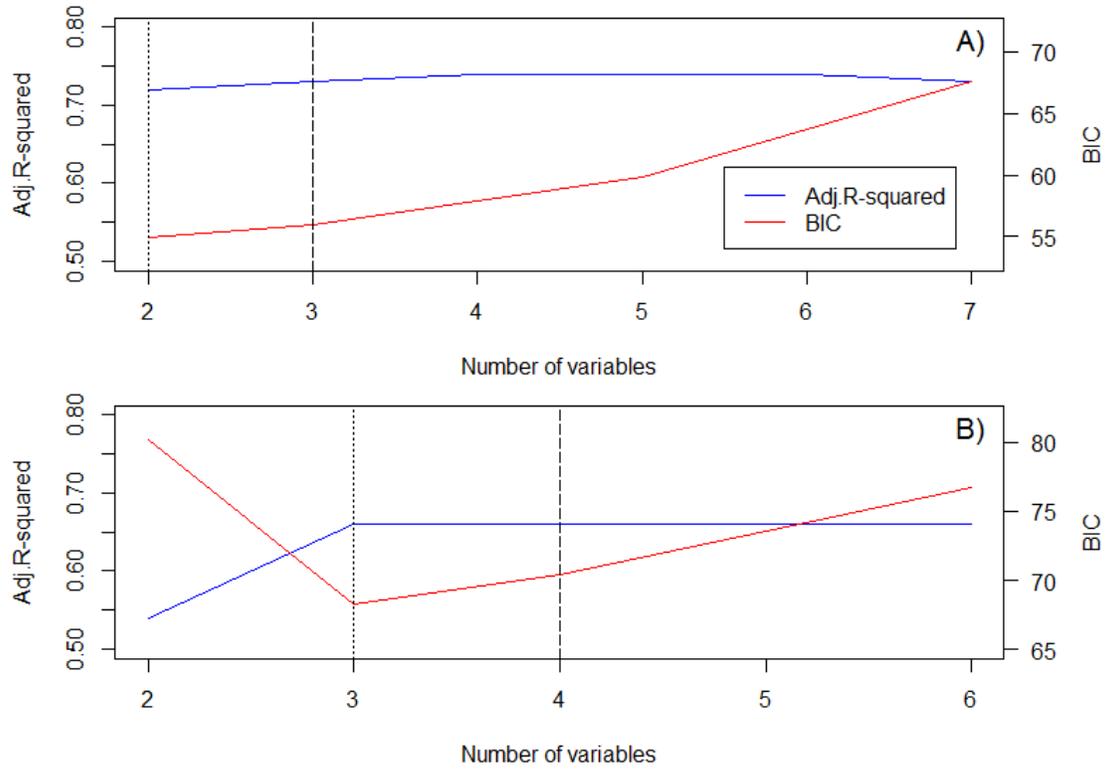


Figure 3.22. Decision support plots for pine MLR model selection based on $\text{adj.}R^2$ and BIC statistics; A) remotely sensed variables, B) ground-based measured variables. $\text{Adj.}R^2$ and BIC are given for the best fitted model for particular number of variables included. Dotted line indicates selected model. Dashed line means that model (and all models to the right) includes at least one non significant variable.

Table 3.10. Selected MLR SI prediction models for pine using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model).

	Model	RMSE (m)	Adj. R^2	p-value
<u>RS</u>	$\text{SI}=34.375+1.398*\log(\text{DTW_2})-0.022*\text{ALT}$	1.54	0.72	<0.001
<u>WAM</u>	$\text{SI}=16.817+1.336*\log(\text{DTW_2})-0.366*\log(\text{FA})$	1.73	0.65	<0.001
<u>GB</u>	$\text{SI}=24.860+2.525*\text{SNR}-0.025*\text{ALTT} +3.096*\log(\text{SP})$	1.69	0.66	<0.001

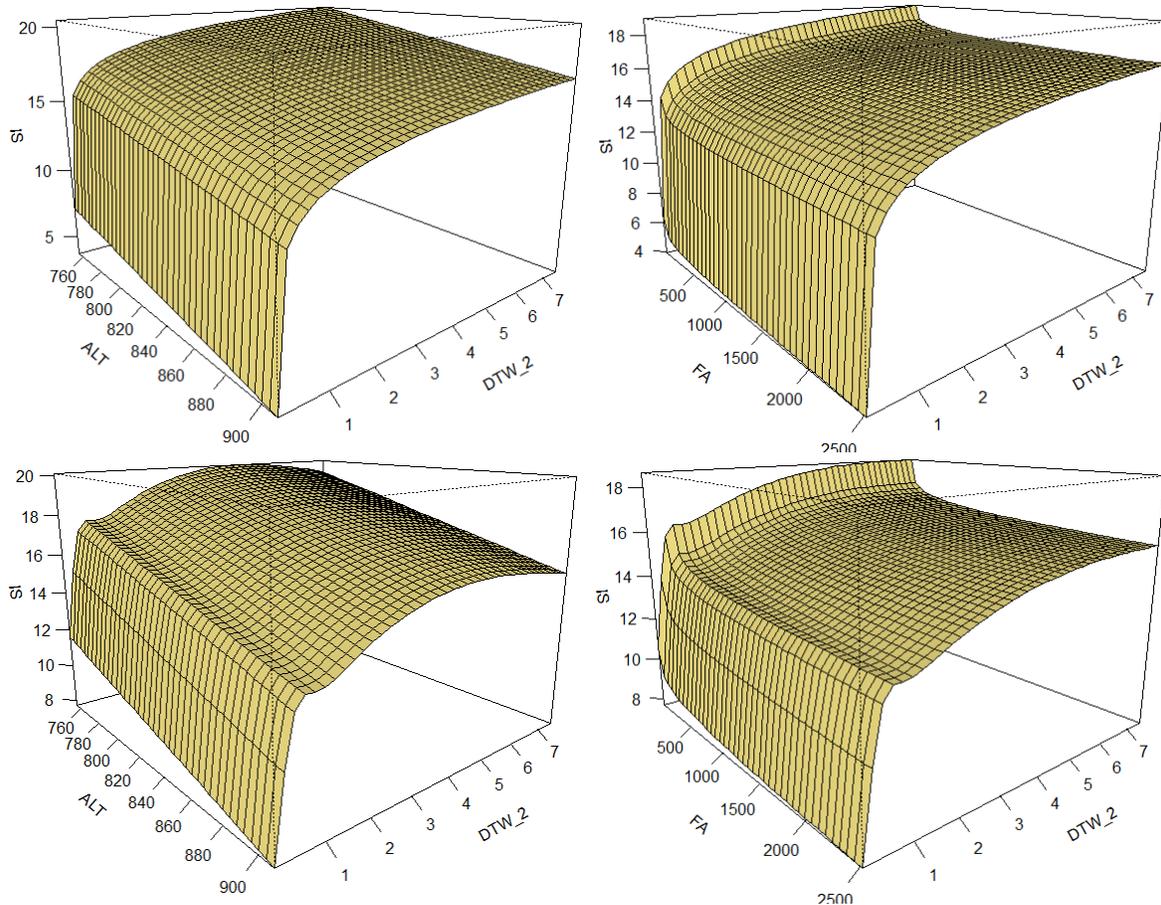


Figure 3.23. Selected pine SI prediction models based on RS variables using MLR method (upper left plot), WAM variables using MLR method (upper right plot), RS variables using GAM method (bottom left plot), and WAM variables using GAM method (bottom right plot). For model description refer to Table 3.10 and Table 3.11.

Generalized Additive Models

All selected models were developed using the same variables as in selected MLR models for all data sources (Table 3.11). The edf was limited to 4 for DTW_2 to properly detect nonlinearity, while for FA a cubic smooth function could not capture the strong nonlinear trend and therefore FA was used in a log transformed parametric term. The optimal RS model (RMSE=1.59 m; $\text{adj.}R^2=0.69$) was built from a smooth function based on DTW_2 and ALT as linear term. Although fit statistics suggested the best RS model included DTW_2, FA and ALT (Figure 3.24), it was not accepted since ALT could be fitted only as nonlinear and was implausible. The WAM variables also provided strong and significant relationships with SI (RMSE=1.80 m; $\text{adj.}R^2=0.60$). However, in both GAM models (RS and WAM; Figure 3.24.) DTW_2 smooth fit

resulted with around 3.9 edf and provided a pattern with SI levelling off at 0.5-1 m DTW. Decreasing smoothness to max 3 edf provided straighter but weaker fit for RS ($\text{adj.R}^2=0.64$) and WAM ($\text{adj.R}^2=0.56$) models. The selected GB model included the same variables as the MLR. Smoothness applied to SNR and slope position did not improve overall model performance ($\text{RMSE}=1.68$ m; $\text{adj.R}^2=0.66$) compared to MLR, and did not differ in terms of covariate behaviour in the model (Figure 3.25). Relative variable importance was higher for SNR than for the other two variables. Log transformation of SI had a negligible effect on results for RS ($\text{adj.R}^2=0.70$), WAM ($\text{adj.R}^2=0.62$) and GB ($\text{adj.R}^2=0.68$) models.

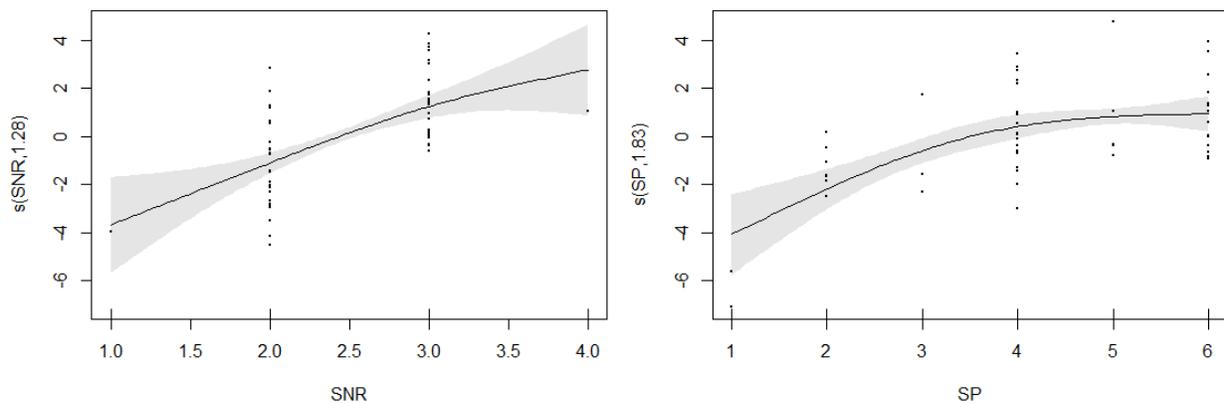


Figure 3.25. Partial response curves of pine SI to selected smooth terms in GB model from Table 3.10. The y-axis shows effect on SI with values centered on zero, with label indicating smooth term and resulting edf. "Rug" lines along x-axis indicating covariate values, dots represent residuals. Shading shows confidence bands (± 2 se) of the estimates.

Regression Trees

Pine SI models developed using the RT method also included the same variables and showed similar strength and form of relationships (Table 3.12) as the previous methods. The RS model resulting from the pruning process (Figure 3.26) accounted for 73% of variation in SI with an RMSE of 1.5 m and was based on DTW_2 and ALT in three splits (Figure 3.27). DTW_2 value of 0.1 m divided the dataset contributing to majority of explained variation and predicting low productive pine sites of 8.4 m SI at $\text{DTW}_2 < 0.1$. With two additional splits a total of four SI classes were predicted. The WAM model ($\text{RMSE}=1.77$ m; $\text{adj.R}^2=0.65$) was a single variable model which used DTW_2 with thresholds of 0.1 and 1.1 m to estimate low productivity (8.4 m SI), intermediate (14.8 m SI) and high productivity (17.1 m SI) for pine. For ground-based variables, the optimal model ($\text{RMSE}=1.71$ m; $\text{adj.R}^2=0.66$) indicated that the most productive

sites were on medium and rich soils (SNR>2.5), and the least productive sites were on poor soils, lower slopes, levels and depressions (SP<4.5), and at lower altitude (Figure 3.27). The primary split in the GB model was created by SNR, with SNR explaining about half of the variance explained by the model and emphasizing the importance of SNR as the only soil variable selected.

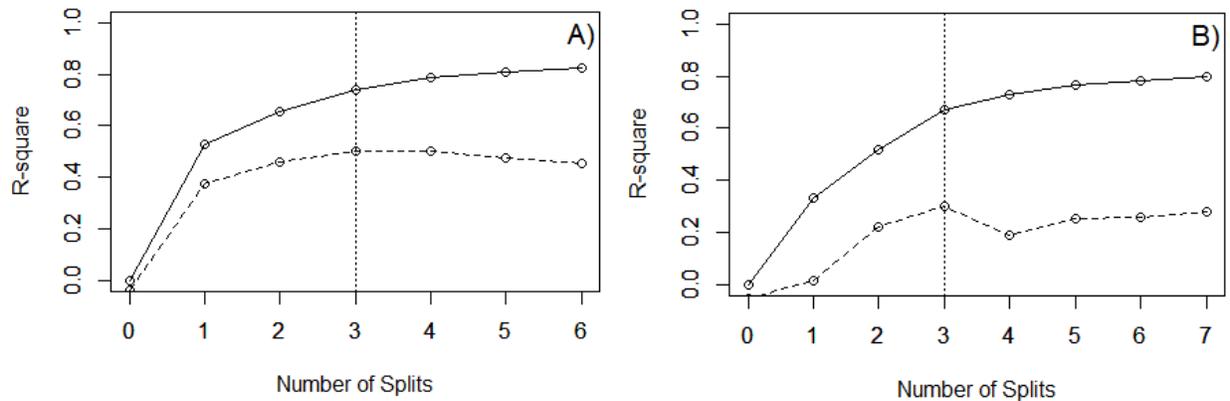


Figure 3.26. Decision support plots for determining the pruning RT for lodgepole pine; A) remotely sensed variables, B) ground-based measured variables. Solid line plots the variation explained (R-squared) by the split in the model. Dashed line plots variation explained (R-squared) by the split from cross-validation (CV). Dotted line indicates the point of pruning (No of splits left) to optimal RT based on maximizing CV R-squared.

Table 3.12. Fit statistics and models rules in explaining pine SI for selected RT models using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model). The Terminal node numbering refers to corresponding models in Figure 3.27.

Model	RMSE (m)	Adj.R ²	Terminal node	Number of obs.	Node rules	Predict SI (m)
<u>RS</u>	1.52	0.73		50		
			Node 2	5	(DTW_2<0.1)	8.42
			Node 5	13	(0.1≤DTW_2<0.8)and (ALT≥805.4)	13.43
			Node 6	10	(0.1≤DTW_2<0.8)and (ALT<805.4)	16.01
<u>WAM</u>	1.77	0.65		50		
			Node 2	5	(DTW_2<0.1)	8.42
			Node 4	23	(0.1≤DTW_2<1.1)	14.80
<u>GB</u>	1.71	0.66		50		
			Node 4	11	(SNR<2.5) and (SP<4.5) and (ALTT≥817.2)	10.46
			Node 5	5	(SNR<2.5) and (SP<4.5) and (ALTT<817.2)	14.96
			Node 6	10	(SNR<2.5) and (SP≥4.5)	15.54
			Node 7	24	(SNR≥2.5)	16.73

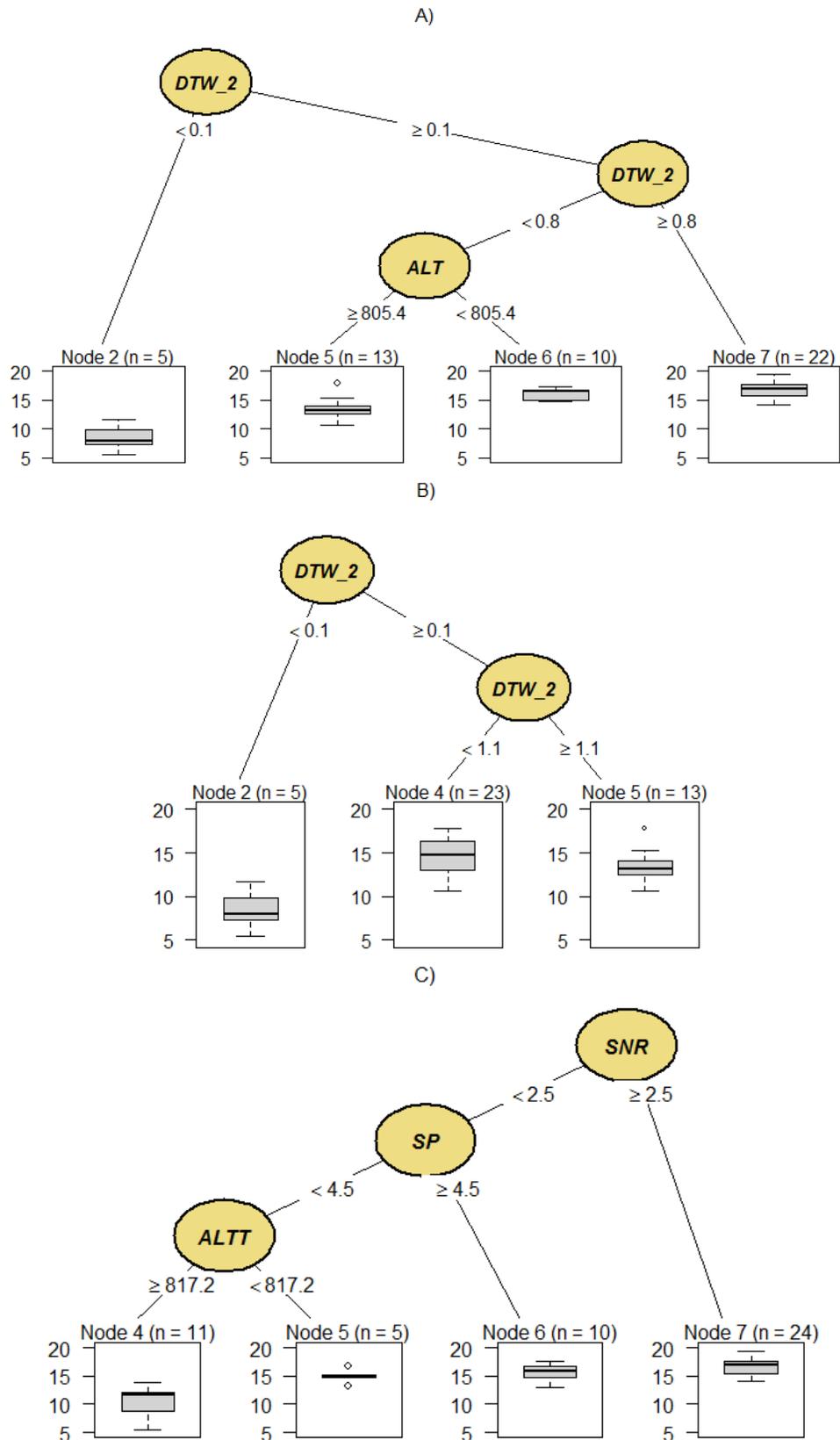


Figure 3.27. Pine SI Regression Tree models; A) remotely sensed variables (RS model), B) only WAM variables (WAM model), C) ground-based measured variables (GB model).

RandomForest

Variable importance for models that included all possible predictors are presented for each data source in Figure 3.28. Selected models optimized in terms of number of variables and R^2 are shown in the Table 3.13. In addition to the variables selected by other methods, slope appeared in both RS and GB final models. The final RS model, which included three variables, and the GB model, which included four variables, had similar amounts of variance explained ($\text{adj.}R^2=0.63\text{-}0.64$). However, the WAM model provided weaker results ($\text{RMSE}=2.26\text{ m}$; $\text{adj.}R^2=0.42$). This may be because FA is included in the model but it showed as the least important variable (Figure 3.28). While the RF method works in the way that at each split it randomly selects variables to evaluate to make a split, in this case it randomly selects one variable ($\text{mtry}=1$) because there are only two variables in the model, hence FA is often chosen for splitting even though it is not a good predictor. All partial response curves confirmed plausibility and consistency with other methods (Figure 3.29). Moreover, DTW_2 among remotely sensed variables and SNR among ground-based variables were found to be the best predictors of pine SI as shown earlier for other statistical methods.

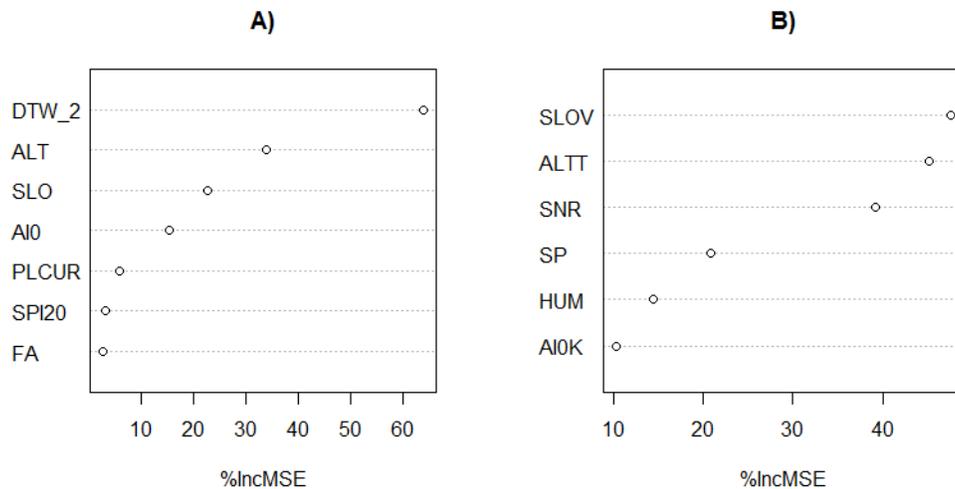


Figure 3.28. Variable importance plots for pine SI Random Forest models built with all potential explanatory variables; A) remotely sensed variables, B) ground-based measured variables. %IncMSE - variable importance measure indicating average percent change in the mean square error when the particular variable is permuted while all others are retained unchanged. Important variables should cause relatively large %IncMSE.

Table 3.13. Fit statistics and variable importance for selected RF models explaining pine SI based on remotely sensed variables (RS model), WAM variables (WAM model), and ground-based measured variables (GB model). Adj.R² - calculated using "pseudo R²" (validation samples); %IncMSE - variable importance measure indicating average percent change in the mean square error when the particular variable is permuted while all others are retained unchanged.

	RS		WAM		GB	
	Variable	%IncMSE	Variable	%IncMSE	Variable	%IncMSE
	DTW_2	49.8	DTW_2	58.6	SNR	47.3
	ALT	42.1	FA	13.4	ALTT	41.2
	SLO	37.8			SLOV	40.0
					SP	29.2
Adj.R ²	0.63		0.42		0.64	
RMSE (m)	1.77		2.26		1.73	

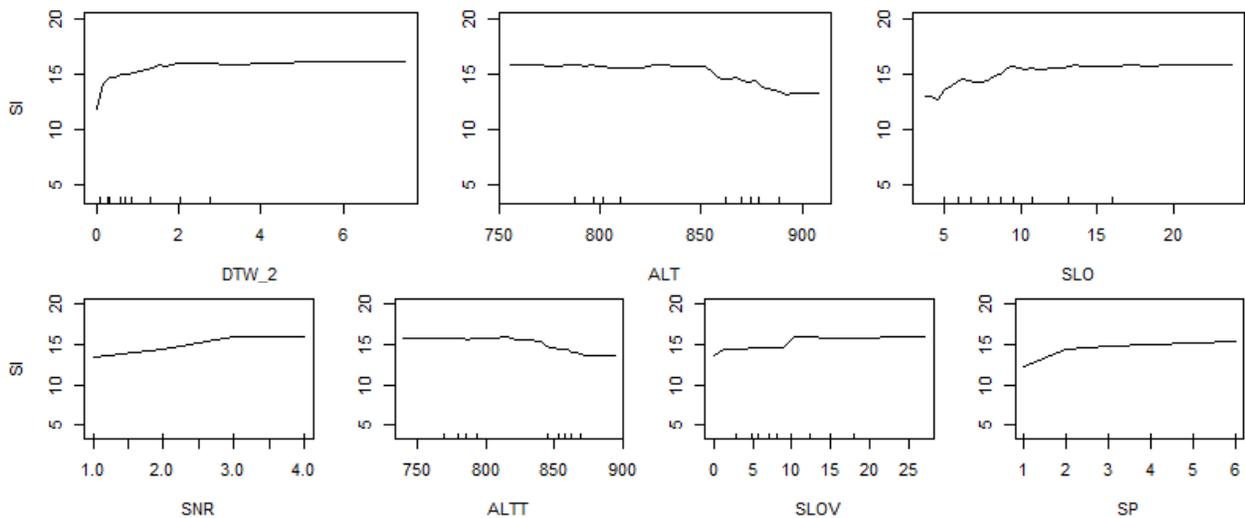


Figure 3.29. Partial dependence plots showing effect of selected variables on pine SI in RS (upper plots) and GB (bottom plots) models. Hush marks at the bottom of the plot show the deciles of the variable.

3.4.3. White spruce

Multiple Linear Regression

Spruce SI models by data sources optimized in terms of model performance measures (Figure 3.30) are presented in Table 3.14. Variables not fitting a linear relationship with SI were transformed. Including terms for interactions among variables did not improve model performance. Only three variables were found to be statistically important in both cases, RS

model (DTW_10, FA, ALT) and GB model (HUM, DRN, ETEXT). The selected RS model for WAM variables was highly significant ($p < 0.001$; $RMSE = 2.01$ m; $adj.R^2 = 0.39$) with DTW_10 as the dominant covariate according to partial R-squared. DTW_10 had a positive logarithmic relationship with sharp decline in SI with excess moisture within 1 m DTW_10, while FA was fitted as an inverse exponential function and indicated an increase in SI with increasing FA and asymptote at high values of FA (Figure 3.31). The final GB model ($p < 0.001$; $RMSE = 1.84$ m; $adj.R^2 = 0.49$) included only soil variables. Topographic variables did not explain additional variation in spruce SI. The best site in terms of spruce height growth is associated with well-drained soil and moder humus type, and the least productive site is very poorly drained with very thick (>40 cm) organic matter (peatmor humus form) (Figure 3.31).

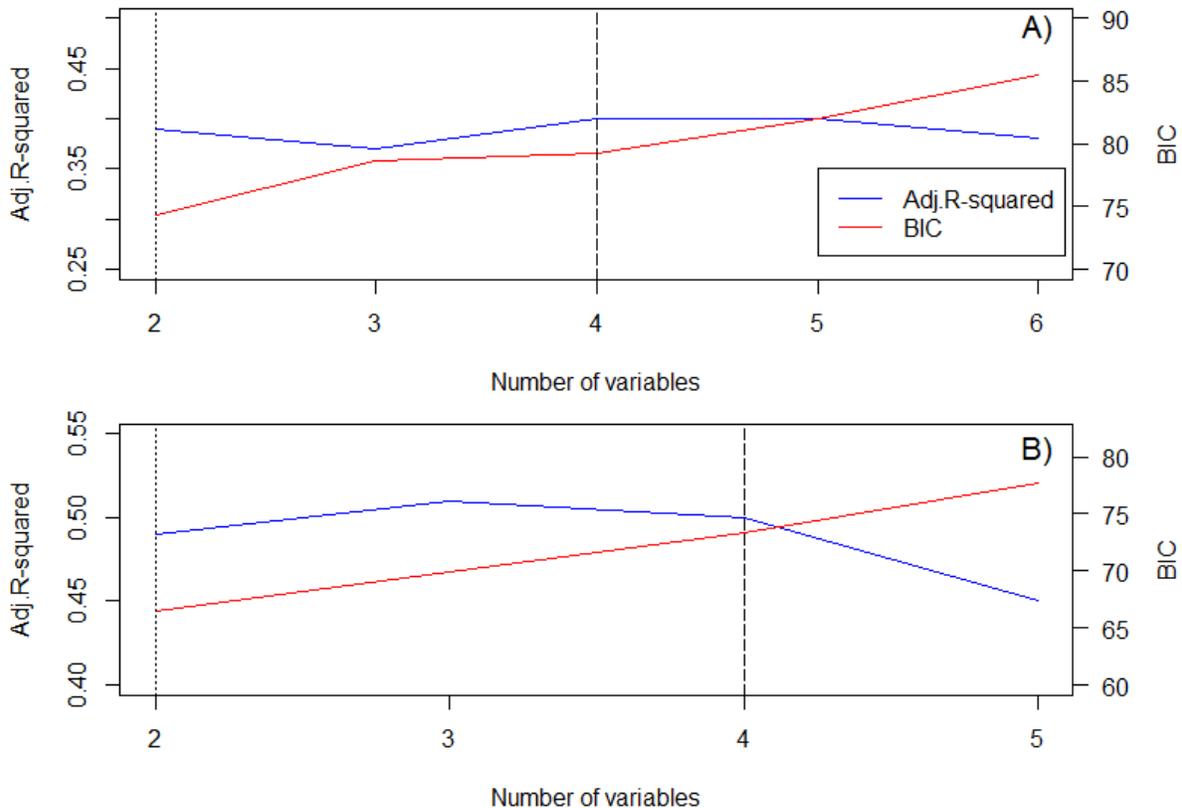


Figure 3.30. Decision support plots for spruce MLR model selection based on $adj.R^2$ and BIC statistics; A) remotely sensed variables, B) ground-based measured variables. $adj.R^2$ and BIC are given for the best fitted model for particular number of variables included. Dotted line indicates selected model. Dashed line means that model (and all models to the right) includes at least one non significant variable.

Table 3.14. Selected MLR SI prediction models for spruce using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model).

	Model	RMSE (m)	Adj. R ²	p-value
<u>RS</u> <u>(WAM)</u>	$SI=35.960+0.967*\log(DTW_10)-18.185*\exp(1/FA)$	2.01	0.39	<0.001
<u>GB</u>	$SI=26.088-5.763*\log(HUM)-0.750*DRN$	1.84	0.49	<0.001

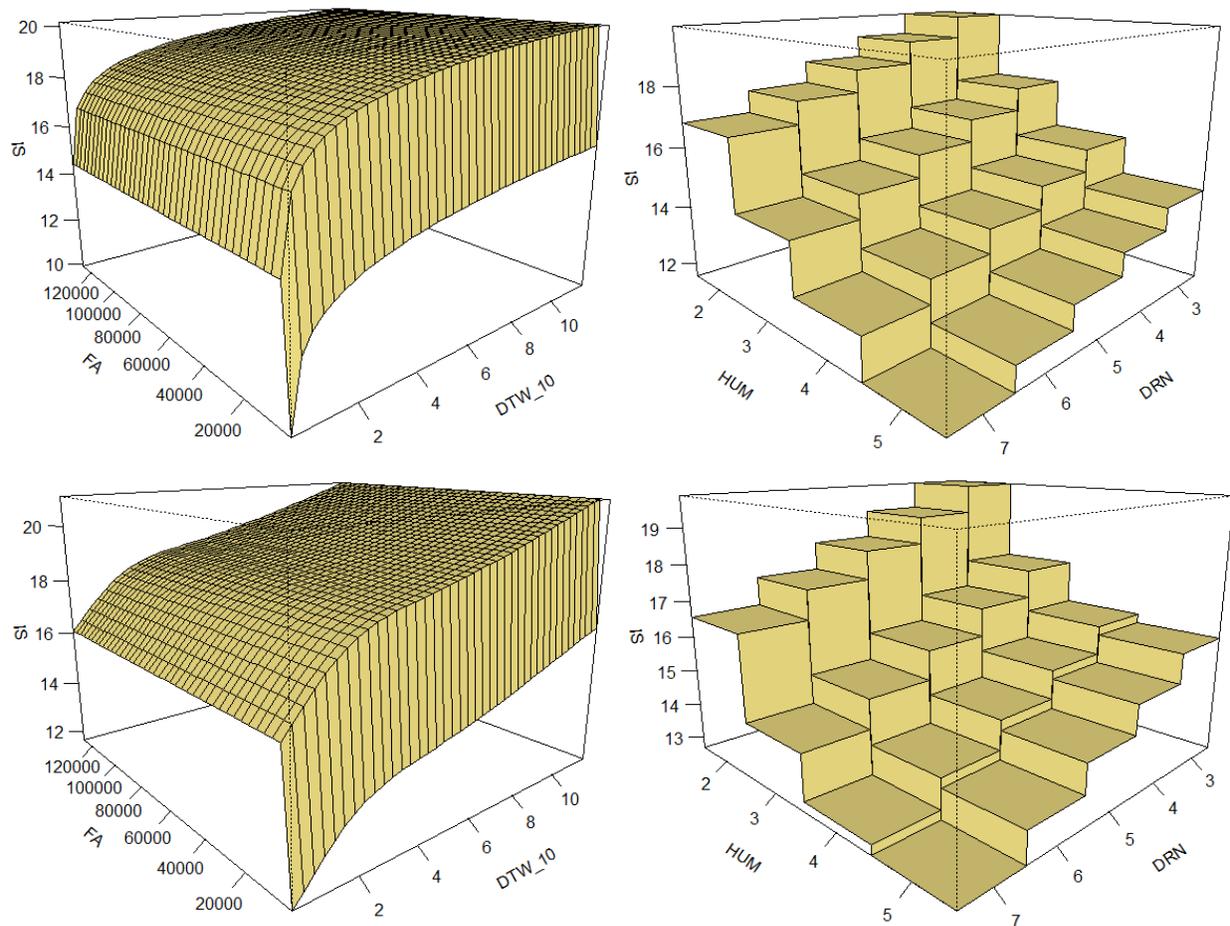


Figure 3.31. Selected spruce SI prediction models using MLR method (upper plots) and GAM method (bottom plots) based on RS (WAM) variables (left plots) and GB variables (right plots). For model description refer to Table 3.14 and Table 3.15.

Generalized Additive Models

The best spruce biophysical SI GAM models are given in Table 3.15. Similar to the MLR method, the WAM variables and soil variables are the most important in spruce SI estimation. The decision support plot (Figure 3.32) suggested that the best model was the one including DTW, FA and ALT (RMSE=2.02 m; adj.R²=0.37) but since a positive altitude trend is not plausible, the model with DTW and FA (RMSE=2.06 m; adj.R²=0.34) was selected. In this model SI increases with increasing DTW_10, but due to smoothness (edf=2.46) the curve slope changes. The selected GB model included HUM and DRN (RMSE=1.82 m; adj.R²=0.50), with a positive linear effect of DRN and a nonlinear relationship between HUM and SI.

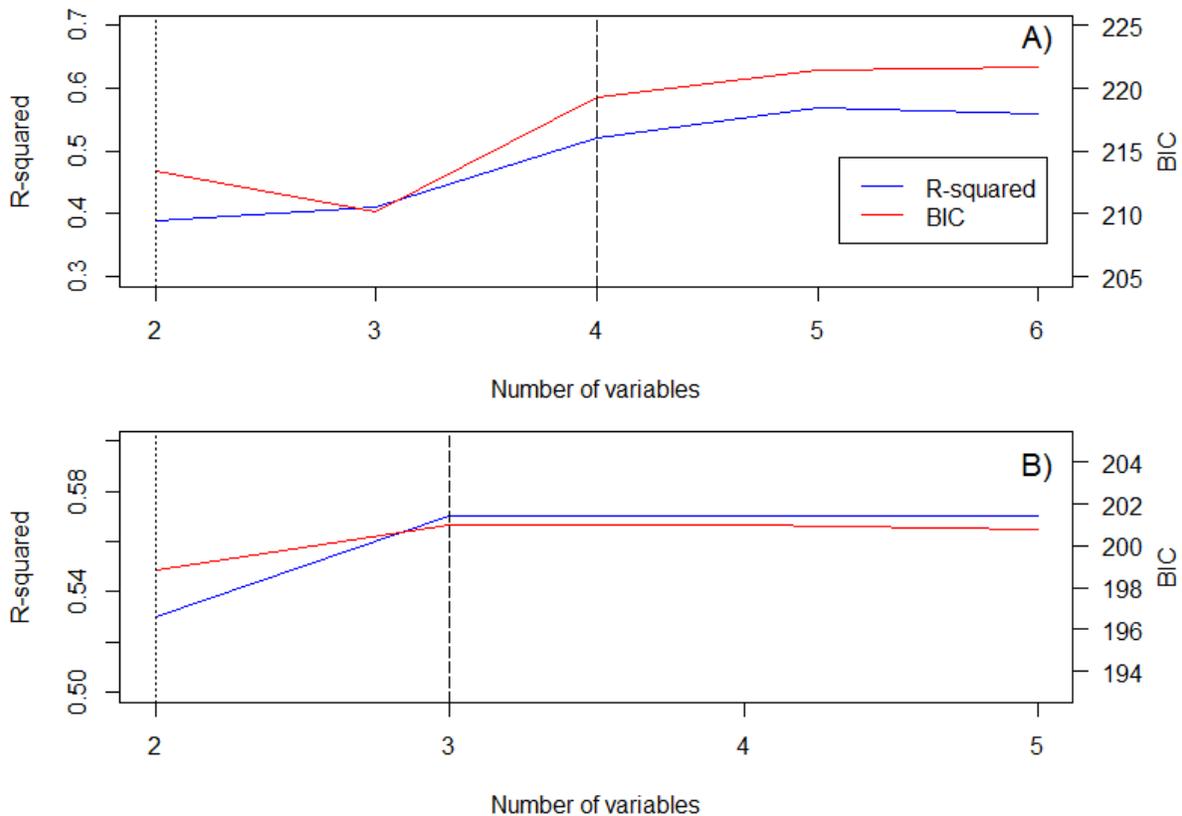


Figure 3.32. Decision support plots for spruce GAM model selection based on R² and BIC statistics; A) remotely sensed variables, B) ground-based measured variables. R² and BIC are given for the best fitted model for particular number of variables included. Dotted line indicates selected model. Dashed line means that model (and all models to the right) includes at least one non significant variable.

Table 3.15. Selected GAM SI prediction models for spruce using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model).

	Model	RMSE (m)	Adj. R ²	
	$SI = \beta_0 + f_1(\text{DTW}_{10}) + \beta_1(\exp(1/\text{FA})) + \varepsilon$	2.06	0.34	
<u>RS</u> (WAM)	Parametric coefficients (β):	Smooth terms (f):		
	Estimate (\pm SE)			
	Intercept	36.015 (± 6.228)***	DTW_10	2.464**
	exp(1/FA)	-17.794 (± 5.945)**		
	$SI = \beta_0 + f_1(\text{HUM}) + \beta_1 \text{DRN} + \varepsilon$	1.82	0.50	
<u>GB</u>	Parametric coefficients (β):	Smooth terms (f):		
	Estimate (\pm SE)			
	Intercept	20.579 (± 1.051)***	HUM	1.557***
	DRN	-0.818 (± 0.260)**		

Regression Trees

Selected RT models for each data source are shown in Table 3.16 and Figure 3.34. The final RS model (RMSE=1.91 m; adj.R²=0.44) used ALT, DTW_10, and SPI20 with ALT as the primary split in the tree. The RS model was chosen based only on fitted R-squared because cross-validation did not provide reasonable results (Figure 3.33). Failure of cross-validation suggested instability of the model and uncertainty of SI estimates. In addition, the RS model included ALT as an important explanatory variable which has an unexplained positive relationship with SI. Testing only WAM variables in the RT model yielded slightly better results (RMSE=1.90 m; adj.R²=0.46), with cross-validation confirming 4 splits as optimal but with the low cross-validation R-squared value of 0.10 suggesting potential issue of over-fitting. The GB RT model was grown from HUM and DRN (RMSE=1.91 m; adj.R²=0.45) with an optimum of two splits proven by CV results (Figure 3.33). Good productivity (average 19.5 m SI) by spruce was achieved at sites with moder humus form, moderate productivity (average 16.8 m SI) with raw moder to peatymor humus and well to moderately-well drained soil, and low productivity (average 14.7 m SI) with raw moder to peatymor humus and imperfectly to very poorly drained soil.

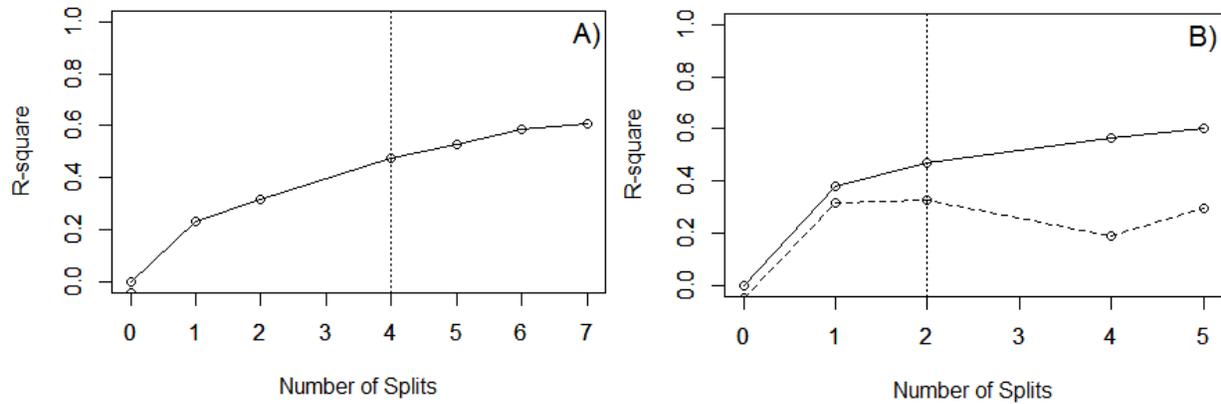


Figure 3.33. Decision support plots for determining the pruning RT; A) remotely sensed variables, B) ground-based measured variables. Solid line plots the variation explained (R-squared) by the split in the model. Dashed line plots variation explained (R-squared) by the split from cross-validation (CV). Dotted line indicates the point of pruning (No of splits left) to optimal RT based on maximizing CV R-squared.

Table 3.16. Fit statistics and models rules in explaining spruce SI for selected RT models using remotely sensed variables (RS model), using WAM variables (WAM model), and using ground-based measured variables (GB model). The Terminal node numbering refers to corresponding models in Figure 3.34.

Model	RMSE (m)	Adj.R ²	Terminal node	Number of obs.	Node rules	Predict SI (m)
<u>RS</u>	1.91	0.44		45		
			Node 3	5	(ALT<775.3) and (DTW_10<0.1)	13.92
			Node 4	13	(ALT<775.3) and (DTW_10≥0.1)	16.58
			Node 6	5	(ALT≥775.3) and (SPI20≥4.5)	16.42
			Node 8	12	(ALT≥775.3) and (SPI20<4.5) and (ALT<829.6)	17.92
			Node 9	10	(ALT≥775.3) and (SPI20<4.5) and (ALT≥829.6)	20.06
<u>WAM</u>	1.90	0.47		45		
			Node 2	6	(DTW_10<0.2)	14.25
			Node 5	5	(DTW_10≥0.2) and (DTW_10<7.6)and (FA<9.8)	15.58
			Node 7	17	(DTW_10≥0.2) and (DTW_10<7.6)and (FA≥9.8)and (DTW_10<3)	17.03
			Node 8	10	(DTW_10≥0.2) and (DTW_10<7.6)and (FA≥9.8)and (DTW_10≥3)	19.00
			Node 9	7	(DTW_10≥0.2) and (DTW_10≥7.6)	20.00
<u>GB</u>	1.91	0.46		45		
			Node 3	9	(HUM≥2.5) and (DRN≥4.5)	14.66
			Node 4	19	(HUM≥2.5) and (DRN<4.5)	16.83
			Node 5	17	(HUM<2.5)	19.49

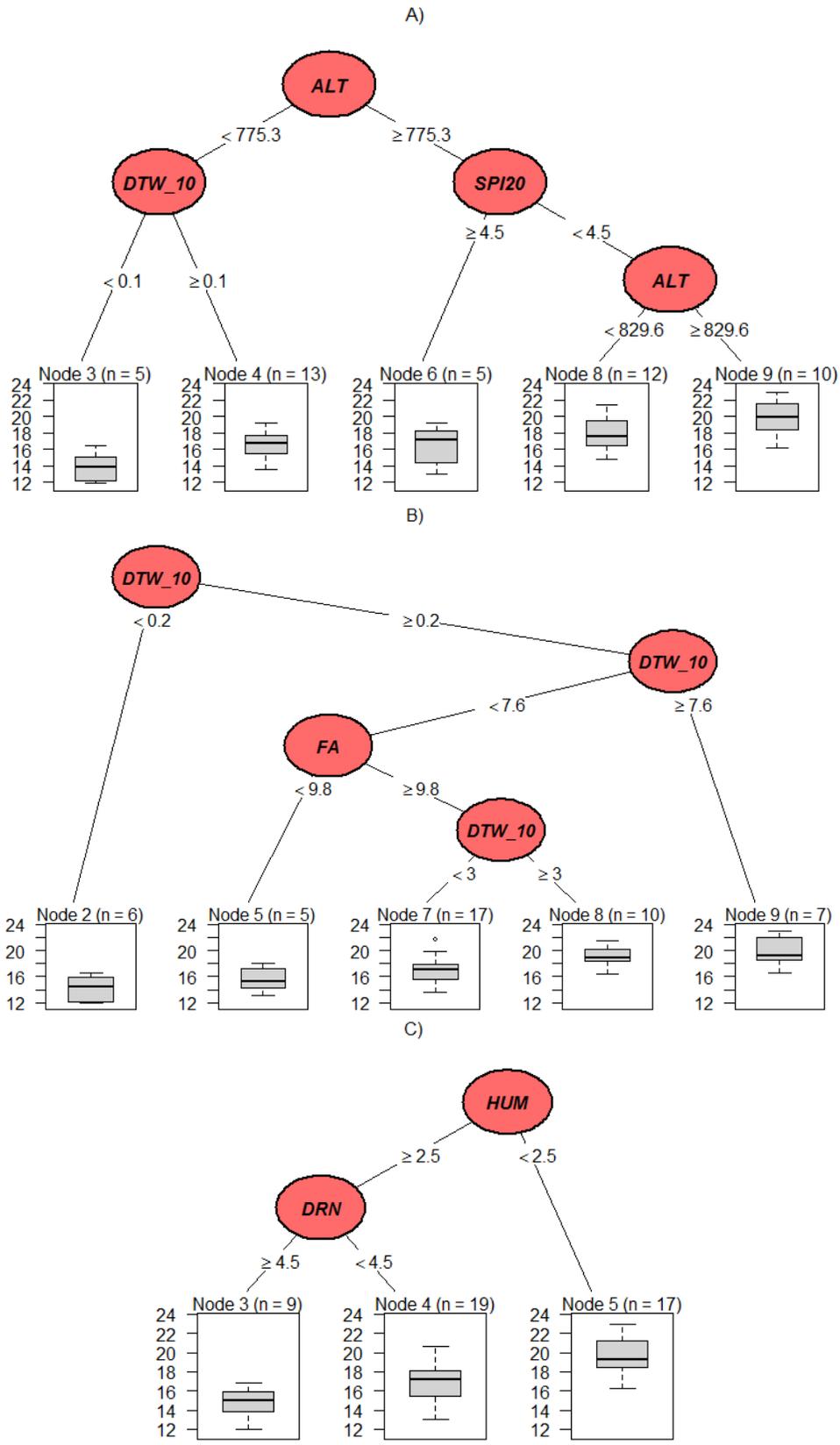


Figure 3.34. Spruce SI Regression Tree models; A) remotely sensed variables (RS model), B) only WAM variables (WAM model), C) ground-based measured variables (GB model).

RandomForest

Variable importance computed for all potential explanatory variables for both datasets, remotely sensed and ground-based, are presented in Figure 3.35. After excluding unimportant variables optimal models were developed (Table 3.17). The RS model (RMSE=2.14 m; adj.R²=0.31) was based on DTW_10, FA and ALT with effects of each variable (Figure 3.36) on SI equivalent with behaviour of the same variables in other methods, however with the same issue relating to the impact of ALT. The WAM model provided apparently weaker results (RMSE=2.33 m; adj.R²=0.20) than all other selected spruce SI models. This may result from the RF cross-validation method applied on the two-variable model with a weak predictor. Finally, the selected GB model (RMSE=2.03 m; adj.R²=0.39) contained the same predictors (HUM, DRN) with fairly analogous PRC (Figure 3.36) as obtained from the other three modeling methods.

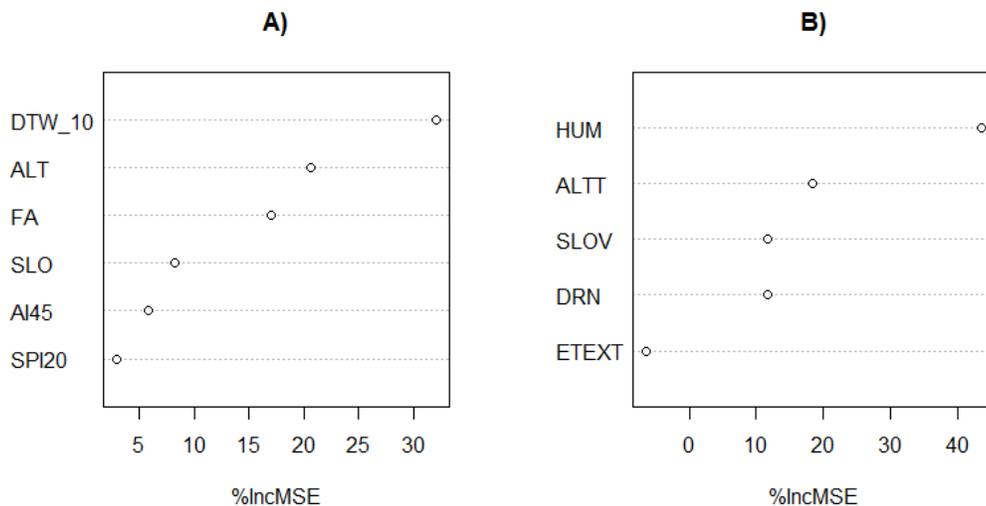


Figure 3.35. Variable importance plots for spruce SI Random Forest models built with all potential explanatory variables; A) remotely sensed variables, B) ground-based measured variables. %IncMSE - variable importance measure indicating average percent change in the mean square error when the particular variable is permuted while all others are retained unchanged. Important variables should cause relatively large %IncMSE.

Table 3.17. Fit statistics and variable importance for selected RF models explaining spruce SI based on remotely sensed variables (RS model), WAM variables (WAM model), and ground-based measured variables (GB model). Adj.R² - calculated using "pseudo R²" (validation samples); %IncMSE - variable importance measure indicating average percent change in the mean square error when the particular variable is permuted while all others are retained unchanged.

	RS		WAM		GB	
	Variable	%IncMSE	Variable	%IncMSE	Variable	%IncMSE
	DTW_10	28.8	DTW_10	46.3	HUM	86.2
	FA	19.2	FA	20.3	DRN	34.0
	ALT	31.1				
Adj.R ²		0.31		0.20		0.39
RMSE (m)		2.14		2.33		2.03

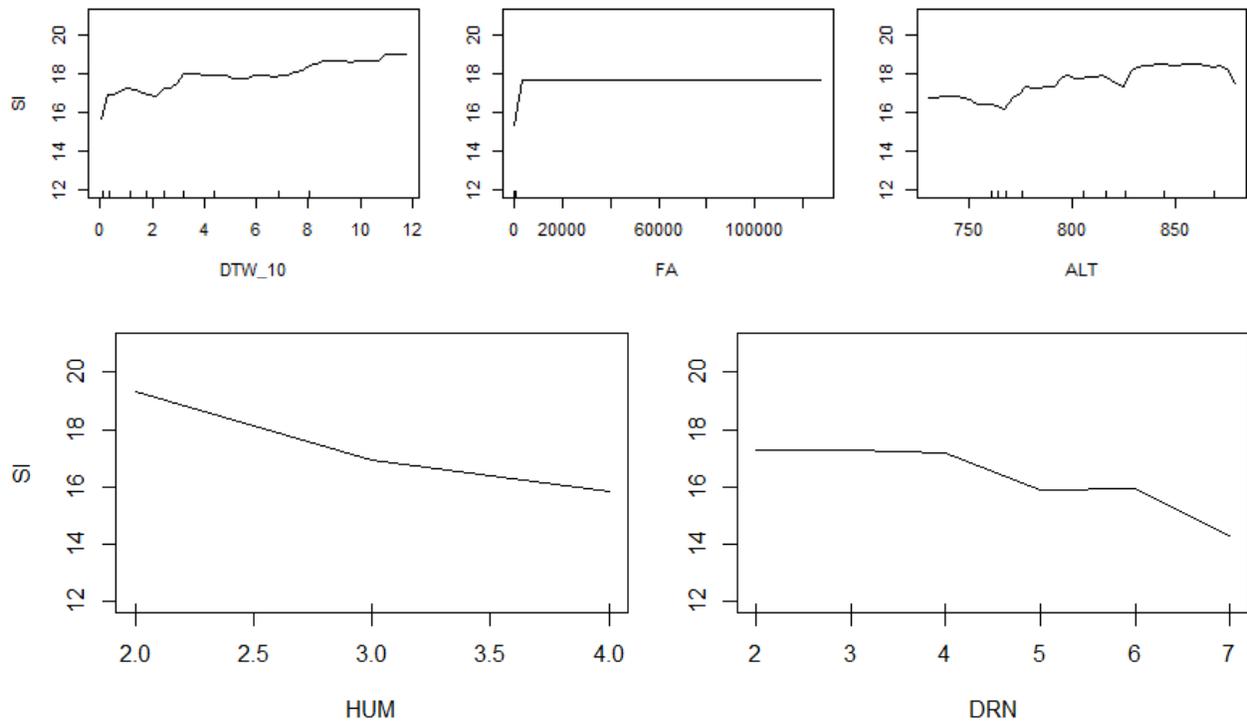


Figure 3.36. Partial dependence plots showing effect of selected variables on spruce SI in RS (upper plots) and GB (bottom plots) models. Hush marks at the bottom of the plot showing the deciles of the variable.

4. DISCUSSION

4.1. Site Index variation and environmental variables

Relationships between environmental factors and Site Index for three major tree species found in boreal forests of central Alberta were examined in this study. Significant relationships between SI and stand basal area were observed for aspen and pine, however, there were no significant relationship between SI and primary species BA proportion (Figure 3.1). An age trend for spruce SI indicates possible violation of assumptions for Site Index determination. This could be due to competition effects in young stands which are common for shade tolerant white spruce, or to a poor fit of the height age curves used for these particular stands. This has been noted as a potential issue in several other similar studies (Aertsen et al., 2012b; Albert and Schmidt, 2010; Anyomi et al., 2015, 2013) with changes in climate and fertility over time often being given as an explanation for this violation of the assumption of time independence for SI estimates. Due to apparent changes in forest productivity, including stand age in SI models has been suggested to account for environmental changes and focus examination on impacts of more constant site factors on productivity (Aertsen et al., 2012b). Additional problems arise as a result of all ecosites not being represented in the research area and with limited sampling at low and high values for most of variables. However, sampling for spruce was unbalanced and resulted in variable distributions that were skewed toward lower slopes and wetter sites producing unstable estimates in parts of ranges where the number of observations was small. In addition, restricted availability of appropriate spruce stands for sampling resulted in a lack of sample independence in some cases. While most results are biologically plausible, some relationships obtained in this study may result from productivity indicator chosen, quality of dataset, and ranges over which sampling was possible.

Topographic variables, although considered as indirect biophysical factors, have often been utilized in forest growth and yield modeling since they are easy to measure. I evaluated relationships between soil properties and topography in order to confirm suitability of topographic attributes for explaining variation in Site Index. Strong correlations were found between soil properties and topography in this study area. SMR had the strongest association with topographic indices, but that may be a consequence of the nature of SMR with it being a

complex soil attribute determined based on soil texture, slope and slope position. Other moisture-related soil variables (drainage, texture, content of coarse fragments) were also closely and plausibly related to topography. Weaker relationships, that were not always significant, were observed with nutrient availability related soil attributes (SNR, humus form, organic thickness), which is in line with findings of Oltean et al. (2016) in young aspen stands. This is likely associated with stand composition being correlated with soil nutrient status and humus layer characteristics. Humus and other biologically related soil properties are affected by stand history and consequently will vary spatially and temporally (Pinno et al., 2009). In summary, my findings are in line with findings reported by Little (2001) that forest floor thickness, clay-size particle content, and available water for plants increased from crest to toe in mixed aspen-spruce forests in the lower foothills of west-central Alberta. Similarly, results from a hillslope transect study in the Foothills Natural Region of Alberta indicate that similar soil attributes - moisture, drainage, texture, forest floor depth, as well as other physical and chemical soil properties, were controlled by topography (Murphy et al., 2011).

The DTW index, although topographically derived and not representing the actual water table, was strongly related to moisture-based soil properties as has been demonstrated by other studies (Ågren et al., 2014; Hiltz et al., 2012; Murphy et al., 2011, 2009; Oltean et al., 2016). However, in my study flow accumulation appeared as the most important topographic index in relation to soil properties. FA represents upslope contributing area which drains to any particular point, and is associated with hydrological connectivity and water movement and associated pathways. Therefore, FA is expected to be related to moisture-based variables, soil texture and other physical properties, as well as to increased accumulation of organic matter on moister sites. In addition to FA, curvature and topographic position were variables important in explaining SNR and soil organic thickness. DTW, slope and TWI were significantly related to SMR but highly intercorrelated, and were therefore interchangeable. TWI was excluded from modeling SI in the first steps of variable selection because it was not as strongly correlated with SI as DTW and slope. DTW outperformed slope in all selected final models, consistent with results from other studies that found DTW to be a better predictor of soil and vegetation characteristics than TWI (Ågren et al., 2014; Murphy et al., 2011, 2009). On the other hand, aspect was not related to field determined soil attributes even though it is associated with temperature and solar radiation which influence stand productivity.

Results from correlation analysis support the hypothesis that topography controls physical soil properties and ecological factors in this lower foothills landscape. In addition, correlation of SMR with almost all remotely sensed topographic indices implies that good models can be developed for predicting SMR (see Appendix), as well as some other soil properties.

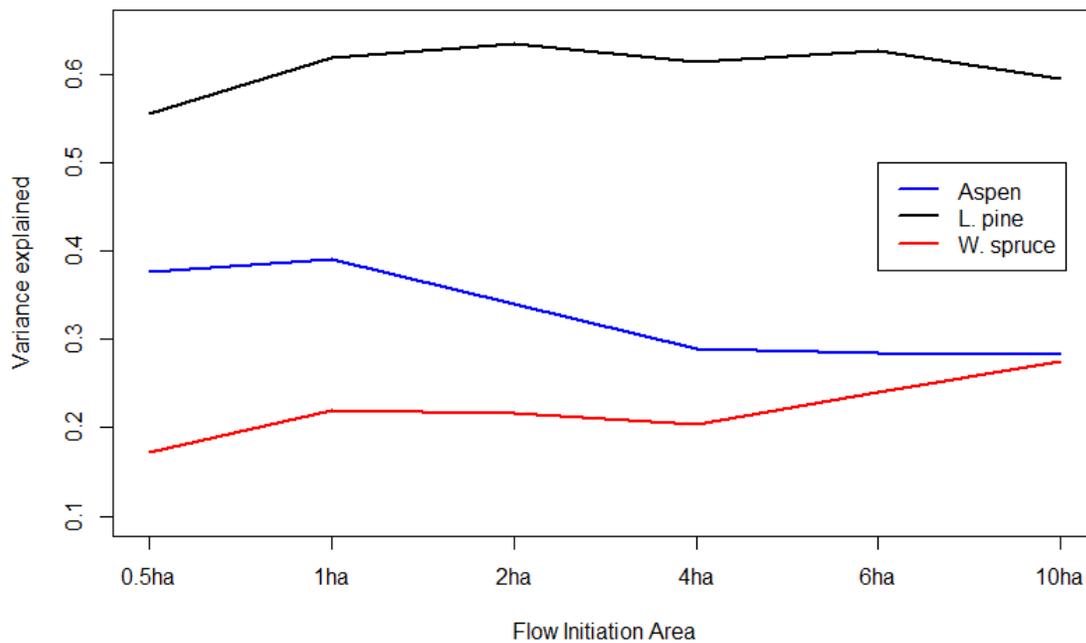


Figure 4.1. Flow Initiation Area in relation to SI variation explain (linear regression) by DTW which is calculated based on associated FIA.

Six different Flow Initiation Areas, from 0.5 ha to 10 ha, were tested in this study to determine the optimal FIA to use in predicting SI from DTW for each species. Divergence in the optimal FIA was found in literature in modeling different soil and vegetation characteristics using DTW (Ågren et al., 2014; Campbell et al., 2013; Murphy et al., 2011; Nijland et al., 2015; Oltean et al., 2016). Relying on the assumption that water movement depends only on gravity and topography, FIA defines the size of contributing area which initiates potential stream flow (Murphy et al., 2009). Therefore, optimal FIA will vary seasonally with a 4 ha FIA being chosen to represent the late summer stream network (White et al., 2012). Different FIA values may be optimal for different purposes (White et al., 2012). Parent material and effective soil texture will also influence FIA with an optimal FIA of 1 to 2 ha for slowly permeable materials and 8-16 ha for highly permeable substrates for soil wetness prediction using DTW (Ågren et al., 2014). Optimal

FIA may also be influenced by topography with suggested FIA values of 2-4 ha for plain landscapes and 12-16 ha for foothills landscapes (Oltean, 2015).

In this foothills landscape dominated by glacial till parent material, the strength of the relationship between SI and DTW differs with different FIA for all three species (Figure 4.1). DTW based on smaller FIA was better in estimating aspen SI, the largest size (10 ha) of FIA was best for spruce, while the size of FIA does not appear to have an influence for pine SI estimation. Therefore, DTW computed at FIAs of 1 ha, 2 ha, and 10 ha was selected as optimal in modeling SI variation of aspen, pine and spruce, respectively, and used in subsequent analysis. This is similar to findings of Oltean (2015) for aspen and spruce height in relation to DTW. Results from correlation analysis (Table 3.1) show that decreasing FIA resulted in increasing correlations between DTW and topography (SLO, CURV, SPI, FA, TWI) and between DTW and soil moisture (SMR, DRN), while correlations with humus form had the opposite trend (decrease with FIA decrease), and altitude and soil texture were insensitive to changes in FIA. Aspen SI appeared moisture sensitive with SMR ($r=0.42$) being the strongest SI predictor, implying optimal FIA of 1 ha. Pine SI is not significantly correlated to SMR and DRN but is strongly associated to SNR ($r=0.51$), which is in turn not related to DTW, consequently FIA size was irrelevant for pine productivity estimation. For spruce SI, soil parameters, with HUM having the strongest correlation ($r=-0.54$), were more important than topography, and due to HUM being better related to a larger FIA, 10 ha showed as the best in SI prediction. Generally, results indicate that if response is moisture sensitive then lower FIA is optimal, but when some other factor drives the response optimal FIA will depend on its relationship with DTW calculated at different FIA. Thus, optimal size of FIA, in addition to parent material and topography, is also dependant on the response variable (application) (Nijland et al., 2015). This is in reference to results reported by Murphy et al. (2011) that the strongest relationship between soil properties and DTW was found at the lowest tested FIA, and by Campbell et al. (2013) that soil resistance to compaction positively correlates with DTW but more strongly at a 1 ha than at 4 ha FIA.

4.2. The main productivity drivers

The main objective of this study was to explore potential to estimate forest productivity using fine resolution remotely sensed environmental data which enable spatial prediction of

productivity, therefore a case study in a relatively small area was chosen to control for variation in climate. Spatial variability in productivity drivers can cause issues of scaling in modeling forest site productivity since model accuracy, selected predictors and their relative contribution to total variance explained will vary with the size of the study area (Aertsens et al., 2012b; Anyomi et al., 2014; Chen et al., 2002). Several studies at regional (provincial) scales report that coarse scale climatic variables describe general patterns of SI for aspen (Ung et al., 2001), lodgepole pine (Monserud et al., 2006), and white spruce (Nigh et al., 2004). In aspen forests of western Quebec, Anyomi et al. (2014) successfully used a combination of microscale (plot level) and macroscale (bioclimatic domains) site factors for aspen SI predictions. In contrast, for aspen SI prediction locally, Pinno et al. (2009) found it only possible to use plot level soil and site factors, and were not able to explain SI using coarse resolution mappable climate and soil data, suggesting that the resolution of predictor variables should correspond to scale of SI measurements. In addition, relationships between ecosite and aspen productivity in the case study in Saskatchewan were not significant due to variability in site conditions included within each ecosite class and due to overlap in site conditions between ecosites (Pinno and Bélanger, 2011). In this study, although an ecosite map and other ecological information are available from forest inventory (AVI), it was not useful in SI estimation because the scale of mapping and the variability of the terrain resulted in mapped polygons not corresponding to plot specific characteristics.

Temporal variability in site factors makes productivity predictions even more complex. Anyomi et al. (2014) found that stand dynamics (stand structure, composition, and age) can influence aspen productivity in mixed stands more than direct climatic factors, but in pure aspen stands Ung et al. (2001) found no improvement over site factors. However, they also report that stand successional stage (Shannon index) is important in SI prediction for two shade tolerant boreal species (Ung et al., 2001). Age was the most important variable in explaining common beech SI in Flanders (Aertsens et al., 2012b). Consequently, sampling in even-aged stands in a landscape where stands are similar in age provides control over effects of stand development stage and reduces errors associated with SI estimation, as was the case for aspen and pine in my study, but not for white spruce.

Table 4.1. Variables included and model performance measures of selected models by species, data sources and modeling techniques. Abbreviations: RMSE - root mean squared error; rRMSE - normalized RMSE to the mean of measured data; Variance explained - R^2 adjusted for the sample size and the number of variables included in the model; BIC - Bayesian information criteria. Note: performance measures of RF models were computed from cross-validation (OOB) samples.

Species	Data source	Mod. method	No. par.	Variables	RMSE (m)	rRMSE (%)	Variance explained	BIC	
Aspen	RS	MLR	5	DTW_1, FA, ALT, AI45,PLCUR	1.63	7.77	0.64	117.7	
		GAM	4	DTW_1, FA, AI45, SPI20	1.54	7.34	0.66	102.1	
		RT	5	DTW_1, FA, ALT, SLO, PLCUR	1.47	7.01	0.70	97.6	
		RF	3	DTW_1, FA, SPI20	1.74	8.30	0.59	121.2	
	WAM	MLR	2	DTW_1, FA	1.91	9.11	0.51	134.7	
		GAM	2	DTW_1, FA	1.70	8.11	0.60	112.1	
		RT	2	DTW_1, FA	1.55	7.39	0.68	94.2	
		RF	2	DTW_1, FA	1.82	8.68	0.56	125.3	
	GB	MLR	5	SNR, ALTT, AI45K, SLOV, SP	1.59	7.58	0.65	117.4	
		GAM	5	SNR, ALTT, AI45K, SLOV, SP	1.59	7.58	0.65	112.8	
		RT	4	ALTT, AI45K, SLOV, SP	1.65	7.87	0.63	115.4	
		RF	4	SMR, ALTT, SLOV, SP	1.83	8.73	0.55	135.5	
	L. pine	RS	MLR	2	DTW_2, ALT	1.54	10.31	0.72	51.0
			GAM	2	DTW_2, ALT	1.59	10.65	0.69	54.2
			RT	2	DTW_2, ALT	1.52	10.18	0.73	49.7
			RF	3	DTW_2, ALT, SLO	1.77	11.86	0.63	68.8
WAM		MLR	2	DTW_2, FA	1.73	11.59	0.65	62.6	
		GAM	2	DTW_2, FA	1.80	12.06	0.60	66.6	
		RT	2	DTW_2, FA	1.77	11.86	0.65	64.9	
		RF	2	DTW_2, FA	2.26	15.14	0.42	89.4	
GB		MLR	3	SNR, ALTT, SP	1.69	11.32	0.66	64.2	
		GAM	3	SNR, ALTT, SP	1.68	11.25	0.66	63.6	
		RT	3	SNR, ALTT, SP	1.71	11.45	0.66	65.4	
		RF	4	SNR, ALTT, SLOV, SP	1.73	11.59	0.64	70.5	
W. spruce		RS	MLR	2	DTW_10, FA	2.01	11.55	0.39	70.4
			GAM	2	DTW_10, FA	2.06	11.84	0.34	72.7
			RT	3	DTW_10, ALT, SPI20	1.91	10.98	0.44	69.7
			RF	3	DTW_10, FA, ALT	2.14	12.30	0.31	79.9
	WAM	MLR	2	DTW_10, FA	2.01	11.55	0.39	70.4	
		GAM	2	DTW_10, FA	2.06	11.84	0.34	72.6	
		RT	2	DTW_10, FA	1.90	10.92	0.47	65.4	
		RF	2	DTW_10, FA	2.33	13.39	0.20	83.7	
	GB	MLR	2	HUM, DRN	1.84	10.57	0.49	62.5	
		GAM	2	HUM, DRN	1.82	10.46	0.50	61.5	
		RT	2	HUM, DRN	1.91	10.98	0.46	65.9	
		RF	2	HUM, DRN	2.03	11.67	0.39	71.3	

A total of 36 species-specific Site Index models were developed for each data subset and modeling method. Selected models (Table 4.1) differ in explanatory variables included, as well

as in the number of variables which ranged from two to five, for different species, data sources, and modeling methods. Aertsen et al. (2011, 2010) also found differences in variables selected by five different modeling techniques applied to three species in two contrasting ecoregions, one where topographic factors control productivity and one where soil properties control productivity. However, in my study DTW was the only variable selected by all statistical methods for each of the studied species. In addition, predominance of topographic parameters over soil properties is noticeable. Only the spruce GB SI model included exclusively soil variables (humus form and drainage) but topographic DTW and FA were also able to account for comparable amount of spruce SI variation. Therefore, topography appeared as the main driver of forest productivity in this study area based on examined predictor variables. In a similar ecological area in aspen-spruce mixed forests Little (2001) reported productivity models with high predictive power ($R^2=0.70$) based on laboratory determined soil physical and chemical attributes. Strong relationship between logdepole pine SI and understory vegetation, soil physical and chemical properties were found in a study in south-western Alberta (Szwaluk and Strong, 2003). Concentrations of some foliar nutrients were also found to be significant in relation to SI of aspen (Chen et al., 1998) and spruce (Wang, 1995) in B.C. Results from these studies suggest that adding explanatory variables obtained from more detailed ecological surveys and laboratory analysis of soil and foliage might improve productivity predictions by accounting additional source of variation. However, that will lead to trade-offs between model effectiveness, cost and potential utility.

Biologically plausible behaviour of each selected variable was evaluated based on relationships between SI and each variable (Figures 3.7-3.9) and based on partial response curves in selected models. The general influence of each variable selected in the final SI models is summarized in Table 4.2 (similar to approach of Caouette et al. (2016)), confirming that the three species, which have different autoecology, respond differently to environmental factors. For moisture-related variables (DTW, FA, slope, slope position, curvature, SMR, drainage), aspen SI generally increases with increasing moisture but declines under conditions of excess water, pine SI declines with increasing moisture, while spruce SI reaches an optimum under moderate conditions (mesic SMR, midslope, etc.). These findings are also consistent with the close association between topographic features and soil moisture. Elevation was negatively related to aspen and pine SI, while it was positively related to spruce SI. Finally, higher nutrient

availability (SNR and humus form) was associated with increases in productivity of all three species.

Table 4.2. Summary by species of responses in SI on site factors included in final models; **↑** - positive, **↓** - negative, **→** - no response, **()** - trend detected only by some statistical methods, * no significant relationship between SI and single particular predictor.

Selected site factors	Aspen SI	L. pine SI	W. spruce SI
↑ DTW	(↑)↓	↑	↑
↑ FA	↑→	↓→	↑→
↑ Altitude	↓	↓	↑
↑ Aspect (cold-warm)	↓	↑*	→*
↑ Slope	↓	↑	↑↓
↑ Slope position (depression-crest)	(↑)↓	↑	↑↓
↑ Curvature (convex-concave)	↓	↑	↑↓*
↑ SMR (dry-wet)	↑(↓)	↓*	↑↓
↑ Drainage (well-poor)	↑	↓*	↑↓
↑ SNR (poor-rich)	↑	↑	↑*
↑ Humus form (mor-mull)	↑	↑	↑

Aspen productivity depended strongly on environmental factors as indicated by strong correlations between SI and almost all tested soil and topographic factors (Table 3.2). However, several statistical techniques (Table 4.1) indicate that topographic variables are more important in explaining variation in aspen SI, since aspen SI is sensitive to changes in soil moisture. Other studies report similar relationships between aspen SI and soil moisture (Anyomi et al., 2015; Chen et al., 2002, 1998; Oltean, 2015; Pinno et al., 2009; Ung et al., 2001). Excess water (shallow water table or stagnant water) and conditions with poor soil aeration limit aspen growth (Burns and Honkala, 1990; Chen et al., 2002; Pinno and Bélanger, 2011) as also shown in my results. As a result, maximum SI of aspen is found on sites with a subhygric soil moisture regime, flat to lower slope topographic position and around 3 m DTW at 1 ha FIA. Since it is a nutrient demanding species, aspen growth was positively related to increasing SNR, as shown by other studies (Anyomi et al., 2015; Chen et al., 2002, 1998). Higher aspen SI was found at lower

altitudes and north aspects, which is also consistent with findings from other studies (Chen et al., 2002, 1998; Pinno et al., 2009).

Lodgepole pine SI was significantly correlated with all examined topographic variables and only with SNR and humus form among soil variables (Table 3.3). Although a dramatic decline in SI was noticed as soil moisture increased going from subhygric to subhydric soil moisture conditions, SMR and other soil physical properties were not significant in the models. This might be due to pine being able to tolerate both very low and very high soil moisture levels as well as the range of other site conditions, leading to substantial variability and a lack of a consistent general relationship between SI and SMR. A lack of samples at the extremes of SMR and a preponderance of a plots in the submesic and mesic range may also be causing a weak relationships between SI and SMR. However, significance of slope position and DTW in the models using RS and WAM data may indicate issues relating to the subjective determination of SMR in the field. Results are consistent with findings by Wang et al. (2004) that pine SI was strongly associated with spatial location (i.g. negatively responded to elevation), weakly associated with SNR, and not related to SMR. Additionally, Szwaluk and Strong (2003) reported that SNR and humus form were good SI predictors, and that soil physical variables were weaker in terms of explaining variance than soil chemical, understory plants and humus form variables. In addition to climate factors, increasing fertility (Fries et al., 2000) and decreasing altitude (Wang et al., 2005) significantly contributed to explaining variation in pine SI. Five different variables were selected in my final models: altitude, slope position and SNR were selected in the best GB models, while slope position and SNR could be interchanged with DTW accounting for similar amounts of SI variation in each of the modeling methods. In general for pine, topographic indexes were better growth predictors than soil properties. Slope was selected only by the Random Forest method, however with no improvement in model goodness-of-fit over other methods. Results indicate that pine productivity was lowest on sites with poor nutrient regimes, higher altitude, lower topographic position (depression, level and lower slope), slight slope, low DTW and high FA values. Those findings acknowledged general pine silvics that good productivity can be achieved on rocky soil, steep slopes and ridges (Burns and Honkala, 1990).

For white spruce major factors influencing productivity were soil moisture (drainage and SMR), soil nutrient availability (humus form and SNR), DTW and altitude. Slope position and FA were

selected in final models but they were weak predictors of SI. Humus form, in contrast to aspen and pine, was found to be the variable most closely related to SI. However, temporal variability in soil organic layer characteristics was not accounted for. In terms of drainage and DTW at 10 ha FIA, decreases in moisture availability positively influenced spruce productivity. Interestingly, spruce grew best on sites with mesic SMR, and growth was lower on submesic and hygric sites, but only two observations were sampled for submesic and hygric sites. SI was also highest on midslope sites and declined with increasing slope. Spruce SI also increased with increasing FA but flattened off at high values. A non-monotonic relationship between SI and DTW calculated at lower FIAs (Figure 3.9) was also found. According to Burns and Honkala (1990) white spruce has a wide ecological amplitude, with limited growth on sites with stagnant water or under very dry (xeric and subxeric) conditions, and good productivity on sites with well balanced water-aeration conditions. Spruce is a nutrient demanding species, but although relatively strong, SNR was not significant due to the limited range of nutrient regimes (medium and rich) sampled and only one observation in poor nutrient regime. Similar results were reported for sub-boreal white spruce where SMR was the strongest factor with moderate moisture conditions having the highest SI, while SAR and SNR were less important (Wang and Klinka, 1996). Wang (1995) found that soil physical properties were much better predictors of spruce SI than soil nutrient properties. Altitude was significant and was positively related to spruce SI. While other studies indicate that white spruce SI is insensitive to climate (Nigh et al., 2004; Wang and Klinka, 1996), interactions between elevation and soil moisture regime may reflect the influence of climate on water availability and drought stress. Spatial autocorrelation issues or a lack of sample independence may also be influencing the results from my study. Stand attributes, including composition and age effects, have been found to be important for shade tolerant species (Ung et al. 2001), such as white spruce, leading to increases in variability in SI under different environmental conditions.

4.3. Site Index prediction models - performance and application

Performance of selected models was examined based on model accuracy and bias. Absolute and relative measures of fit, used for evaluation and comparison of selected models, are summarized in Table 4.1 by species, data source and modeling method. Minimizing RMSE, relative RMSE

and BIC, and maximizing adjusted R^2 indicate models with better prediction power. It is important to note that fit statistics and predictions for the Random Forest method were computed based on cross-validation data ("out of bag" observations not used to develop particular regression tree), with $\text{adj.}R^2$ being calculated based on RF's "pseudo R^2 ", and with RMSE being root mean prediction error.

Basic assumptions were investigated to ensure unbiased model predictions. Collinearity (overfitting) between predictor variables in final models was avoided by using a two step method for variable selection which involved first applying correlation analysis to exclude all predictors uncorrelated to SI and to identify collinear variables, and second optimizing models in terms of number of variables included based on fit statistics. Normal distribution of residuals with no trends and biases was confirmed using the Shapiro-Wilk test on residuals ($p > 0.05$), and graphically based on plots of residual and measured vs. predicted (Figures 4.3-4.5). Some biases in the models are apparent and are reflected by the fact that SI is overestimated on poorer sites and underestimated on better sites. What's more, models built using non-parametric methods (RT and RF) appeared to provide much narrower ranges in predicted values due the nature of their algorithms, which are robust to outliers. Similar differences in predicting SI ranges were reported for RT and BRT methods compared to MLR and GAM (Aertsen et al., 2010). While variable weighting was applied by Ung et al. (2001) to minimize biases resulting from unbalanced variable distribution, I did not weight variables in my study.

For easier model comparison $\text{adj.}R^2$ results are visualized in bar graphs grouped by species, data sources and modeling approaches (Figure 4.2). In addition to different responses to site factors, species show differences in the amount of variation explained by the models. Models developed in this study accounted for up to 73%, 70% and 50% of variation in SI for pine, aspen and spruce, respectively. This is consistent with other similar SI-environment studies. Aertsen et al. (2012a) developed SI models based on edaphic characteristics in Flanders and obtained R^2 values of 67%, 77% and 68% for oak, beech and pine, respectively. Soil physical and chemical attributes explained up to 73% of the variation in SI for Douglas-fir plantations in central Europe (Eckhart et al., 2014). Species-independent RF climate sensitive SI models accounted for 78% of overall variation of western US tree species (Weiskittel et al. 2011) and 36-63% of variation of eastern US tree species (Jiang et al., 2015). Sharma et al. (2012)

developed a SI model using site and climate as independent variables with R^2 of 0.86 for Norway spruce and 0.72 for Scots pine in Norway. Soil and climatic data accounted for up to 57% of the variation in SI of Mediterranean pine in Spain (Bravo-Oviedo et al., 2011). In the boreal forest of Quebec Hamel et al. (2004) described models based on climate and biophysical factors which explained 51% of black spruce and 49% for Jack pine SI variation. Finally, using stand age and remotely sensed environmental data, including full feature LiDAR, Watt et al. (2015) presented a SI model with an R^2 value of 0.88 for *Pinus radiata* plantations in New Zealand.

Generally, SI prediction models are scale dependent, with trade-offs between scale and model accuracy, explaining between 50 and 80% of SI variation in models built from local site factors (Bontemps and Bouriaud, 2014). Results from this study fall within this range, but with pine and aspen reaching the upper limit, and spruce approaching the lower end of this range. In terms of relative RMSE, aspen models show better predictive power than pine and spruce. My best models for spruce explained 34-50% of SI variation (31-39% after cross-validation in RF method). Nigh et al. (2004) obtained a weak relationship ($R^2=0.08$) when climate factors were fitted to white spruce SI in interior B.C., which was also much lower than other examined species. Wang and Klinka (1996) explained up to 90% of variance in spruce SI using soil and physiographic variables as predictors, but the model did not perform well when validated against an independent dataset (Kayahara et al., 1998). Therefore, site factors important for spruce growth not included in the study, or issues relating to stand ages and sampling design as explained above, might cause poorer results for spruce than for aspen and pine. While Ung et al. (2001) suggest including stand attributes as predictors for shade tolerant species to improve biophysical SI prediction models. This may only be needed where sampling includes trees that may have experienced suppression during portions of their lives, or where there is uncertainty regarding the age of top height trees.

Several studies report strong relationships between aspen growth and environmental factors in boreal forests. Chen et al. (1998) reported strong relationships between SI and topographic indices (adj. $R^2=0.60$), SMR and SNR (adj. $R^2=0.63$), and detailed soil physical and chemical properties (adj. $R^2=0.69$) for northern B.C. Even stronger models were described in interior B.C. based on geographic position, elevation, SMR, and SNR (adj. $R^2=0.61$), while variance explained was 82% when biogeoclimatic zones were included as a variable in the model (Chen et al.,

2002). In my study the best models explained between 63 and 70% of the variation aspen in aspen SI (55-59% after cross-validation in RF procedure), similar to results from these other studies. Stand characteristics and soil physical and chemical variables were found to improve parent material-specific aspen SI models with R^2 of 0.63 and 0.88 in Quebec (Pinno et al., 2009) and R^2 from 0.45 to 0.54 in Saskatchewan (Pinno and Bélanger, 2011). These results suggest the possibility of improving aspen productivity prediction if field assessment is followed up with soil laboratory analysis. However, our results indicated that field survey or remote sensing was sufficient to capture the majority of variability of lower foothills aspen productivity caused by topography/soil factors. At the broad scale, soil and climate variables explained 69% of aspen SI variation across Quebec (Ung et al., 2001) and 53% of aspen SI variation across Canada (Anyomi et al., 2015).

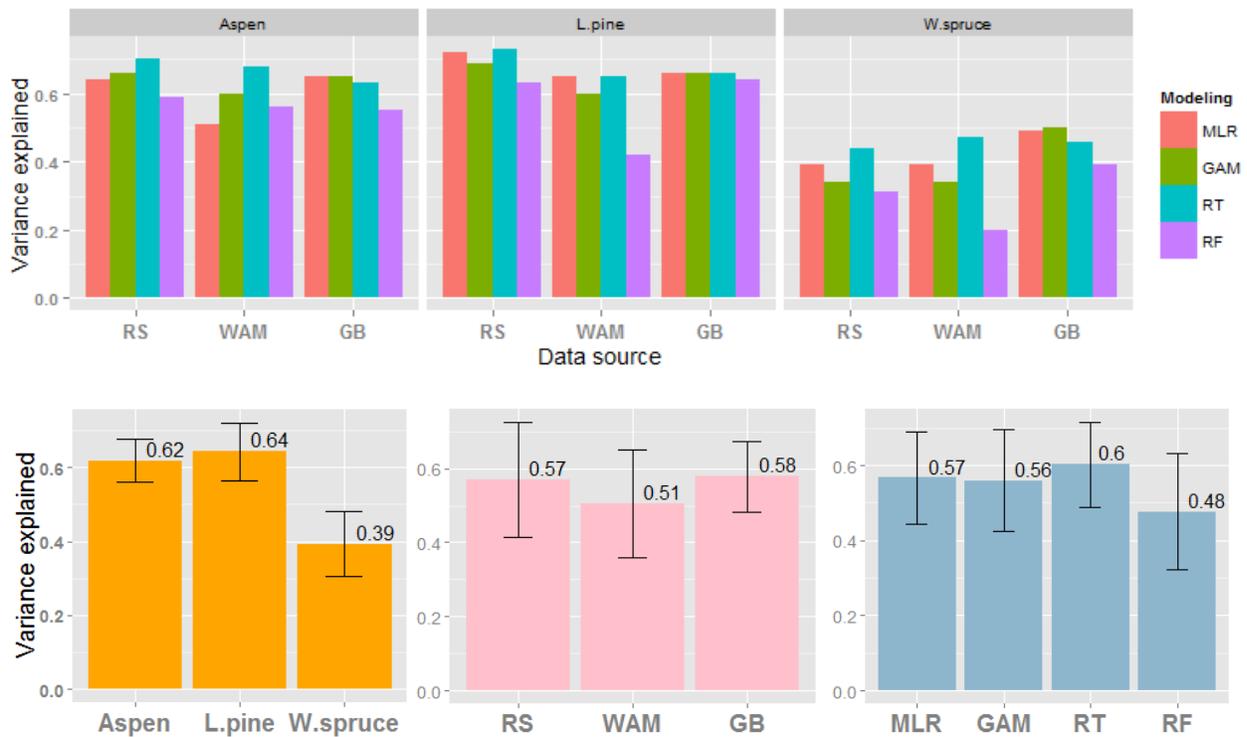


Figure 4.2. Model comparison by species, data sources and modeling techniques based on variance explained. Numbers represent the mean, error bars represent standard error of the mean.

The best models for pine accounted for 66-73% of the variation in SI (63-64% after bootstrapping in RF method). In a similar study in south-western Alberta physical and chemical properties and understory vegetation accounted for up to 67% of the variation in SI variance

(Szwaluk and Strong, 2003). While in an area with contrasting topography (across four NSR) in west-central Alberta Wang et al. (2005) calibrated SI models using climate, topography and soil sand content which explained up to 75% of the variation in SI. Hence variables that were tested for SI prediction in this study, both field and aerial, provided comparable fit to models reported by other similar regional (local) studies. At a provincial level a weaker relationships were reported between SI and spatial location, soil moisture and nutrients ($R^2=0.41$) (Wang et al., 2004), and climatic factors ($R^2=0.27$) (Monserud et al., 2006) in natural pine stands. It is interesting to note that DTW was the strongest predictor of pine SI in my study, with potential to explain up to 63% of the variation in pine SI by itself.

In order to evaluate the use of environmental data in spatial prediction of forest site productivity, SI models were built from different data sources, remotely sensed (RS) and ground-based measured variables (GB). For aspen, RS and GB models provided almost identical results using parametric methods while RS models had slightly better fit when nonparametric methods were used. Pine RS models explained up to 73% of the variance in SI, compared to GB models which explained up to 65%, as a result of including DTW as the strongest predictor. While spruce GB models explained on average 10% of variance they were better than RS models since some soil properties were more closely related to SI than was topography. Generally, comparing results from different modeling approaches across all species (Figure 4.2, bottom middle plot), a factorial ANOVA revealed no significant differences in $\text{adj.}R^2$ values between RS and GB data used as explanatory variables. In addition, although it provided significantly lower $\text{adj.}R^2$, WAM data by itself was able to explain much of the variation in SI.

For DTW, GAM provided more plausible fit for aspen and MLR was able to better detect logarithmic trends for pine. In terms of best predictors, among RS variables DTW was selected by each statistical method for each species and in most cases it is the strongest predictor in the model, while among GB variables different variables appeared as the most important for each species according to species silvics. Integrated moisture index, which uses variables computed from DEM, was employed in white oak SI prediction explaining 64% of variation (Iverson et al., 1997). In *Cryptomeria japonica* plantations in Japan Topographic Wetness Index and Solar Radiation Index were the best predictors of SI ($\text{adj.}R^2=0.65$) among many tested DEM-derived variables (Mitsuda et al., 2007). Laamrani et al. (2014) described a regression tree with 31% of

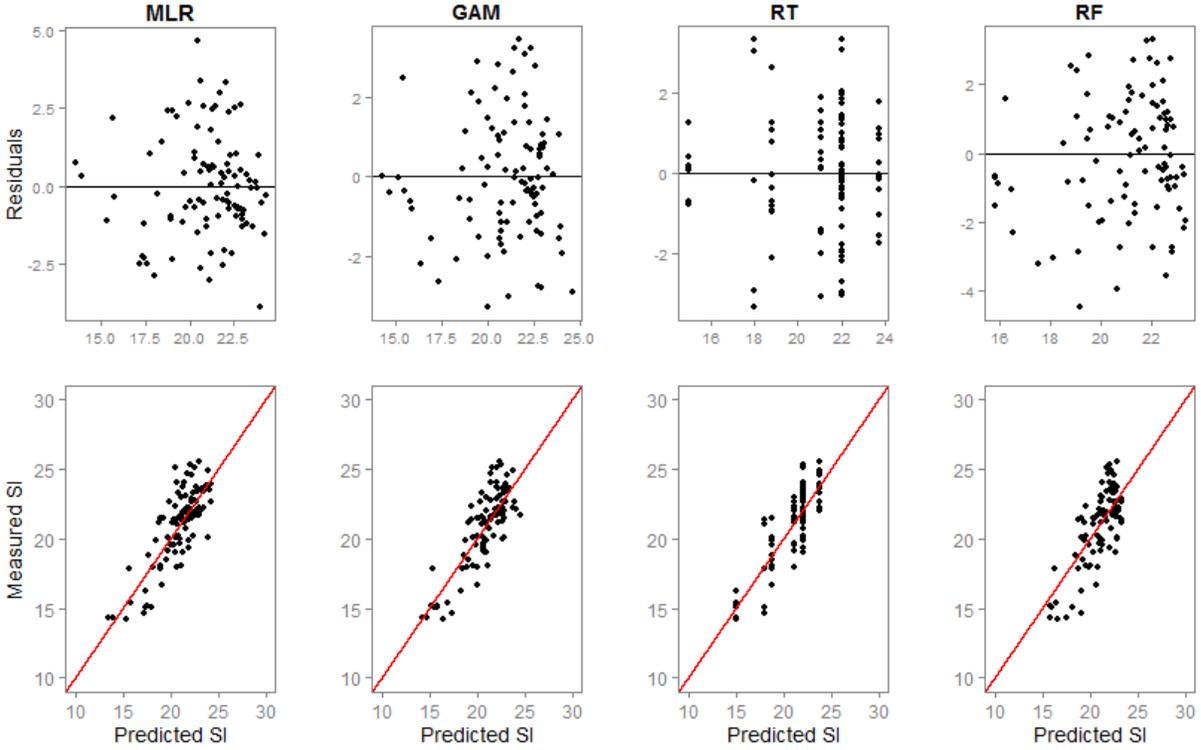
the variance explained for black spruce SI in Quebec based on slope and aspect index derived from DEM. Finally, different data sources were evaluated for estimating *Pinus radiata* SI with an R^2 of 0.61 reported for environmental factors (climate factors and topography) and an R^2 of 0.24 for RapidEye satellite imagery (Watt et al., 2015). My results demonstrate the usefulness of WAM variables for accurate and high resolution estimation of SI, and also demonstrate that DTW index is the strongest LiDAR derived SI predictor and that addition of FA index can help to explain the influence of other important soil properties in lower foothills of central Alberta.

Further improvements in SI predictions might be possible using Full Feature LiDAR. First-return ALS metrics are likely to be more useful in forest plantations than in natural forests since plantations are usually single species, even-aged, with known planting date. For example, Watt et al. (2015) calibrated a SI model for New Zealand's *Pinus radiata* plantations based on 500 sample plot measurements and testing various LiDAR indices, found two variables (H_{99} and stand age) included in the best model ($R^2=0.88$ and RMSE of 1.38 m); and, Packalén et al. (2011) obtained an even more precise SI model using numerous height and density LiDAR metrics and clone identity in a eucalyptus plantation using 195 field sample plots with rRMSE of 2.8%. Although remotely sensed data could be successfully utilized in productivity prediction, field measurements of SI (determination of species, TH trees characteristics, age, height) are still necessary for model calibration. In addition to plantations, stand age can be determined remotely as time-since-disturbance indices from Landsat time series images or fire/harvest history maps and employed in different environmental studies (Nijland et al., 2015; Yeboah et al., 2016; Zald et al., 2016). However, Tompalski et al. (2015) used dominant height estimated from ALS and stand age from Landsat to develop productivity prediction models in natural forests in Vancouver Island, but the model was inaccurate due to the fact that it did not recognise differences between species or effects of regeneration delay.

Several different methods for developing SI prediction models have been suggested in the literature. Four commonly used methods were selected and applied in SI prediction in our study. Aertsen et al. (2010) suggested that three criteria (predictive performance, ecological interpretability and user-friendliness) be considered, with different weights assigned for different applications. Although RT models tended to overfit when only two WAM variables were used, no significant difference is apparent between MLR, GAM, and RT methods in terms of variance

explained (Figure 4.2, bottom right plot). RF gave poorer results due to different statistics being used for comparison, but results are still comparable. Non-parametric models experienced more biases in terms of predicting higher values for less productive sites and lower results at more productive sites (Figures 4.3-4.5). Slightly different fit was found as a result of differences in the representation of nonlinearity between MLR and GAM models, however similarity in covariate behaviour in the models from different methods was confirmed by the form of relationships for each of the species. MLR is a widely used method for this kind of research but is sometimes not able to reveal nonlinear trends and interactions. GAM is a data-driven method but is highly affected by the smoothing method chosen, hence providing ecologically plausible fit, but is not suggested in cases where many variables have different nonlinear patterns. RT provides a user-friendly approach with descriptive results that detect and represent complexity and interactions but does not give a smooth fit since it predicts classes. RF is also an easy-to-apply method designed to work with many variables in large datasets, but is inefficient when applied to small datasets (Sabatia and Burkhart 2014). Overall, all four methods can be used in examining SI-environment relationships. My findings are consistent with recommendations that preference for SI prediction be given to GAM in terms of predictive performance (Wang et al., 2005), as well as to GAM and non-parametric BRT in terms of fit statistics and ecological interpretability (Aertsen et al., 2011, 2010). When spatial autocorrelation issues arise a Sequential Autoregressive (SAR) spatial regression modeling approach may be superior to other methods (Latta et al., 2009; Zald et al., 2016).

Aspen - RS models



Aspen - GB models

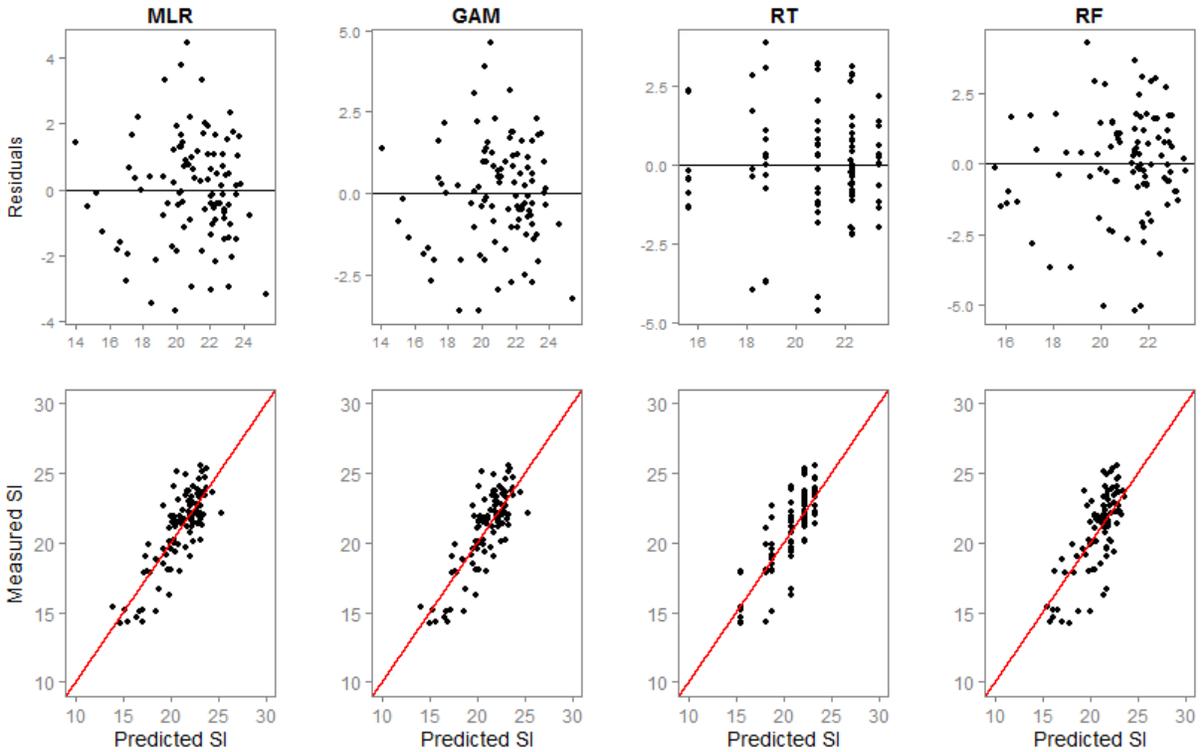
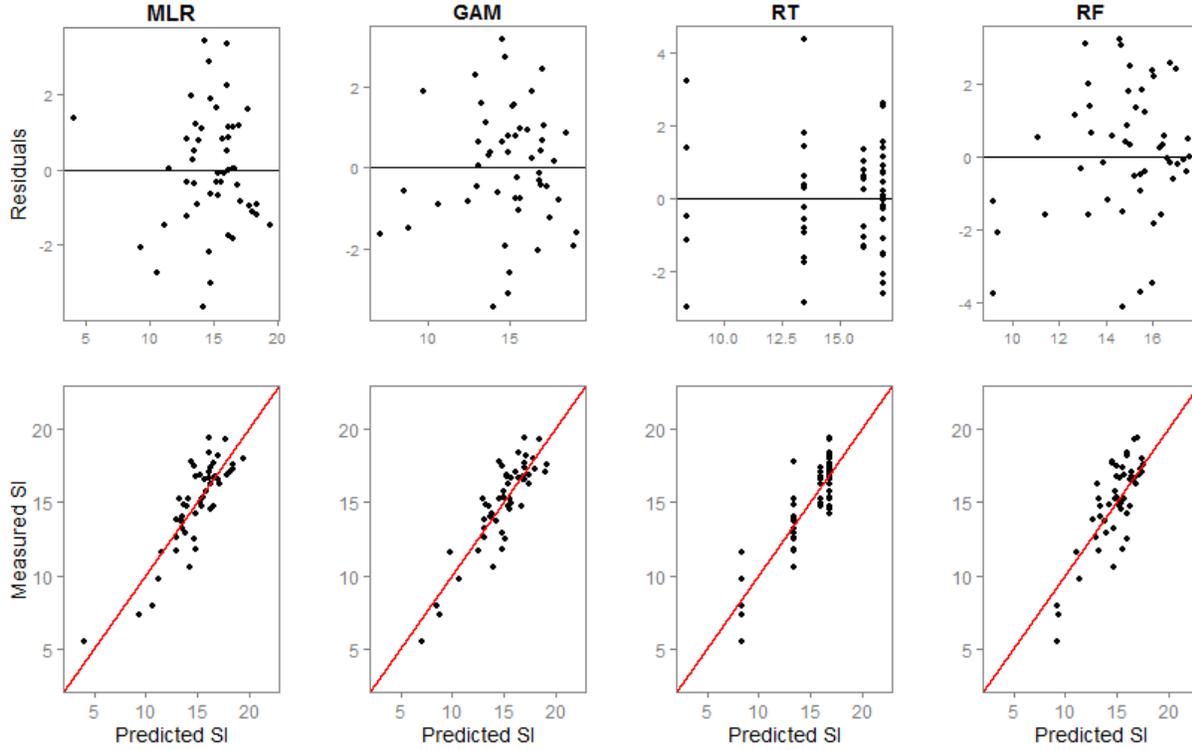


Figure 4.3. Residual and measured vs. predicted plots for aspen by data sources and modeling techniques. The 1:1 line is shown in red color.

L. pine - RS models



L. pine - GB models

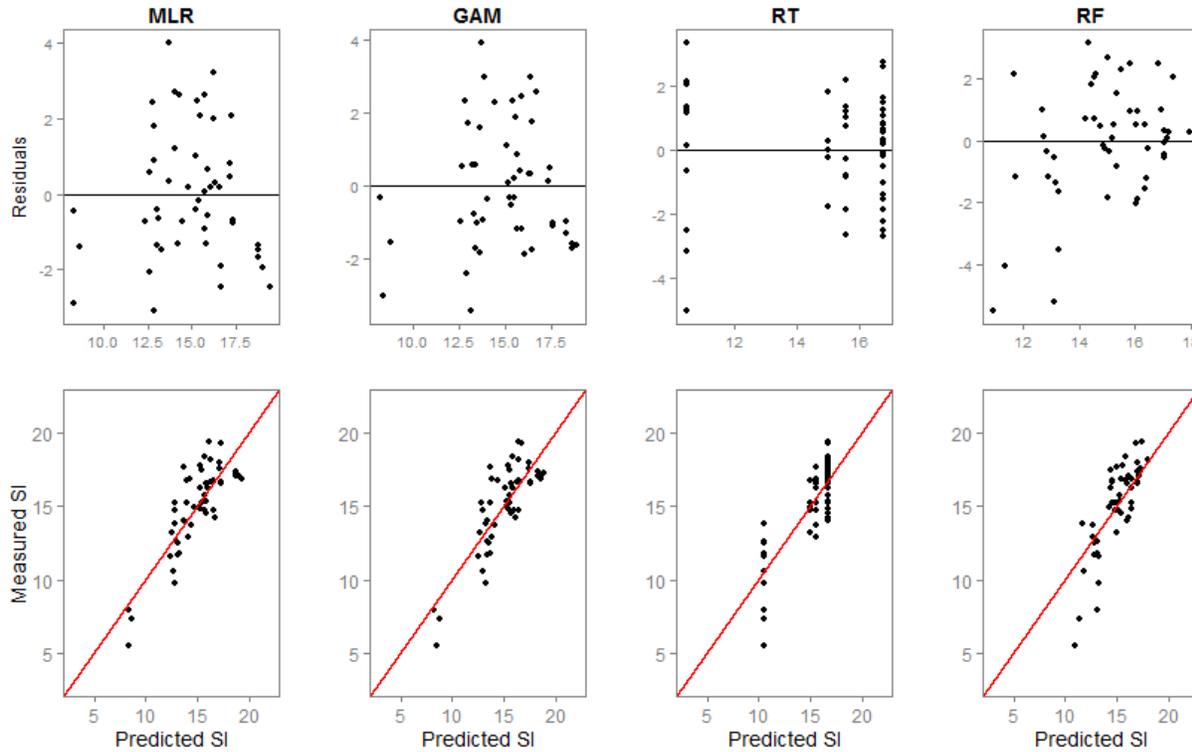
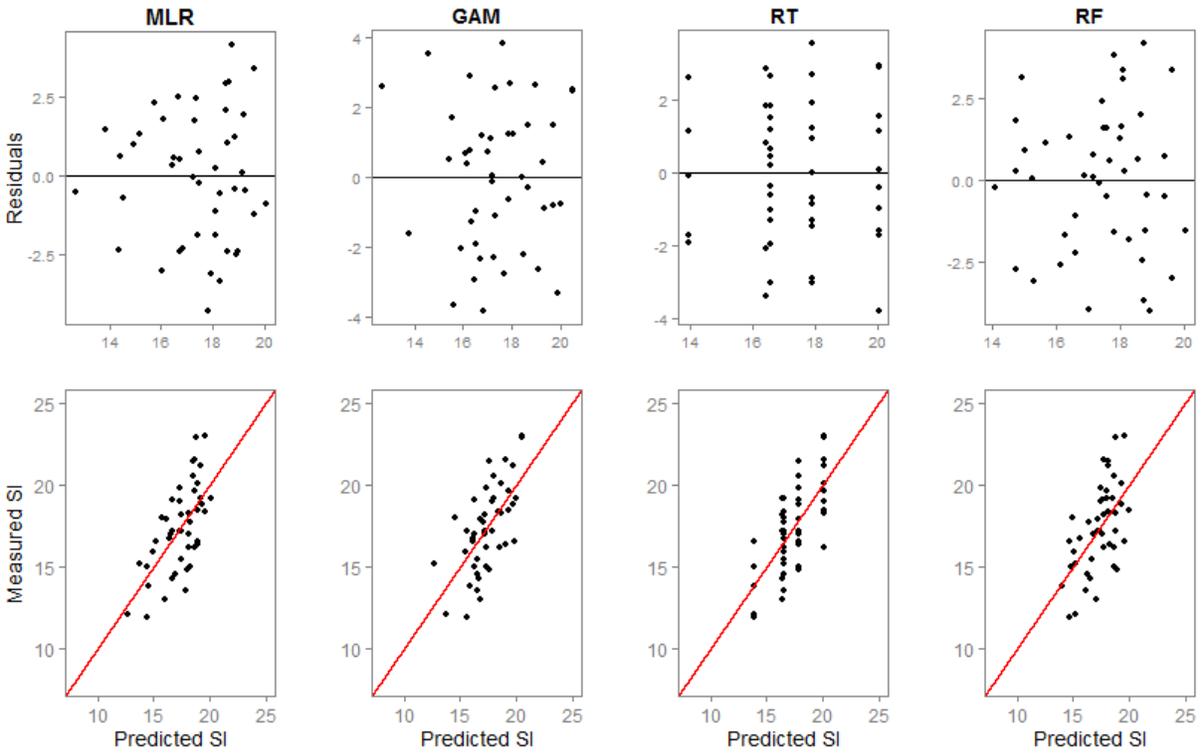


Figure 4.4. Residual and measured vs. predicted plots for pine by data sources and modeling techniques. The 1:1 line is shown in red color.

W. spruce - RS models



W. spruce - GB models

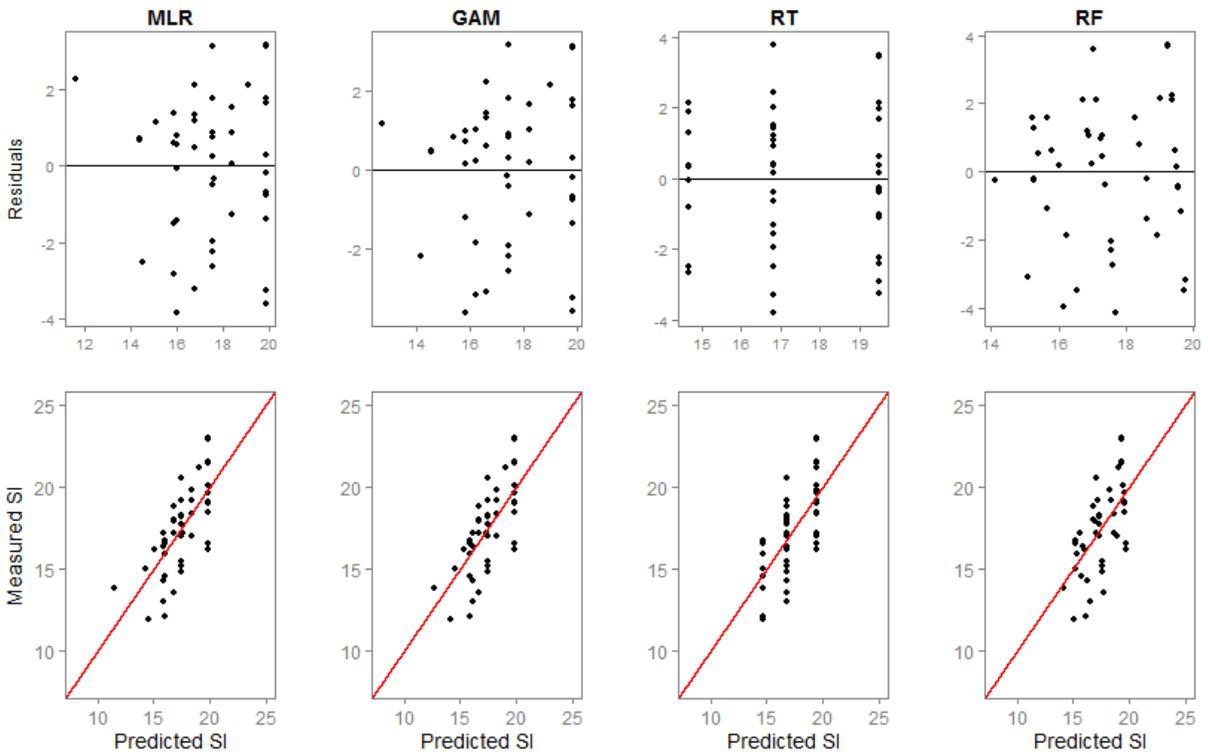


Figure 4.5. Residual and measured vs. predicted plots for spruce by data sources and modeling techniques. The 1:1 line is shown in red color.

Examples of mapped SI variability for all three studied species are shown in Figures 4.6-4.8. Despite some differences in predictions due to model specifics, relationships between terrain attributes and patterns of low and high predicted productivity are generally apparent. However, some issues related to spatial predictions exist. Firstly, extrapolating outside the range of training data as well as outside of study area, is risky and can result in unreliable estimates (McKenney and Pedlar, 2003). Nonparametric regression tree-based methods used in our study are data driven and provide categorical outcomes, so when applied outside of the range of training data they predict the same classes as found in the training dataset (McKenney and Pedlar, 2003). In other words, for both training and external datasets they always predict the same range in response variable. Sabatia and Burkhart (2014) also reported that RF models provided illogical SI predictions when used with climate data outside of the training range. Another difference evident in the aspen SI maps are differences in predictions between MLR and GAM models along streams (out of range for low DTW and high FA) where MLR tends to overpredict SI due to a negative linear fit for DTW and a positive logarithmic fit for FA, while the GAM model exhibits a decrease in SI at low DTW and high FA values. A similar pattern in spruce SI prediction could be noticed at higher elevations (and higher DTW) with the GAM method producing much larger values and area of high SI (>20 m) due to a positive relationships between SI and DTW. In order to develop biologically sound SI models predicted from contemporary and future climate projections Antón-Fernández et al. (2016) applied a modeling technique based on the potential modifier functional form which bounds the response within a specified observed range.

Further, spatial SI predictions illustrate differences resulting from using different types and numbers of explanatory variables in relation to providing even delineations in SI along topographic gradients when using discrete/categorical vs continuous variables and/or different numbers of independent variables. For instance, fine scale differences can be observed between 20-22 and 22-24 m SI classes among aspen based on the MLR model which used continuous surface curvature, and the GAM model which used the discrete variable slope position. Also, differences appear between MLR/GAM models (used only DTW and altitude) and RF model (used DTW, altitude and slope) for pine which result from interactions between DTW and slope at higher topographic positions in the RF model.

For spatial predictions and mapping, MLR and GAM provide continuous fit throughout the range of response and are ecologically plausible. For RT a drawback is that estimates represent classes which result in spatial prediction being fragmented (Wang et al., 2005). When there are small numbers of classes with unequal differences between classes, resulting predictions may not be realistic, as observed for the aspen RT model where application of a DTW threshold of 10.9 m (Figure 3.18) assigns SI values of 18 m and 22 m to neighbouring cells. However, RT might be appropriate for SI predictions at large scales with large datasets (McKenney and Pedlar, 2003), for digital soil type mapping (Illés et al., 2011) or for prediction of other discrete variables (e.g. SMR). Mapping based on RF models provided a more balanced combination of smoothness and classes and no issues with overpredicting.

Scaling and transferability are also important issues in forest productivity prediction. Productivity models are limited to area and scale of development, hence loss of predictability appears when SI models are applied at smaller or larger extents, or transferred to other areas (Aertsen et al., 2012b). In this study variation in potential site productivity is mapped at a fine (detailed) scale and, as a result, is applicable for operational scale forest growth and yield modeling. In addition, predictive SI maps for different species at the same location (Figures 4.6-4.7) may have application in species selection and in defining silvicultural strategies.

Forest productivity is affected by complex interactions between site factors, genetic variability, and management practices and as a result it is difficult to assess (Weiskittel et al., 2011b). Biophysical SI models usually have moderate accuracy and rarely explain more than 80% of the variance, which is also the case in environmental sciences in general. Our models showed promising results, however there are several potential sources of error which could not be accounted for. Brandl et al. (2014) described possible limitations in SI modeling including quality of dataset and missing factors affecting height growth. Computation of SI is subject to measurement error for height and age, but applying provincial SI curves which might not account for regional differences in growth. Some soil properties are associated with short-distance variability (i.e. organic thickness) and subjectivity in determination (eg. SMR, SNR). In addition, biases and limitations in field sampling (sample size and variable distribution) also determine quality of database and potentially bias the inferences. In addition environmental changes over time might be important if age divergence is observed. Growth factors not

accounted for in the models may also include detailed soil attributes, some extreme abiotic or biotic factors, such as pathogens or insects or even clonal diversity for aspen (Latutrie et al., 2015). Finally, SI is maybe not as appropriate an indicator as biomass production (Brandl et al., 2014). While it should also be recognized that biomass production is influenced by environmental factors (climate and site), as well as stand (stocking, composition, stand structure, age) and other factors.

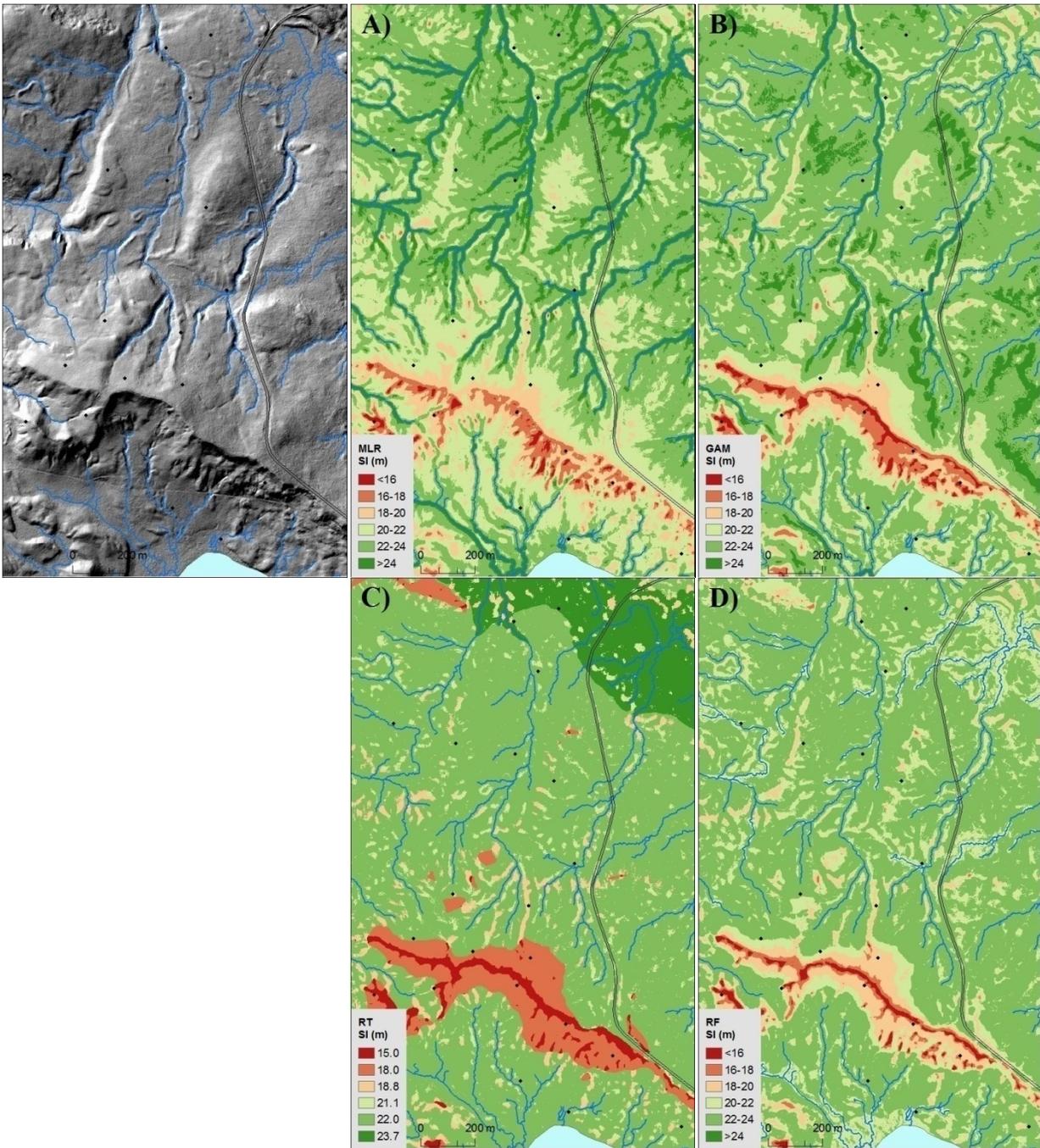


Figure 4.6. Hillshade and mapped aspen SI (detail) generated based on selected: A) MLR model (Table 3.6), B) GAM model (Table 3.7), C) RT model (Table 3.8), and D) RF model (Table 3.9). Dots represent calibration sample plots; blue lines represent WAM predicted stream channels at 1 ha FIA. Map resolution is 1 m.

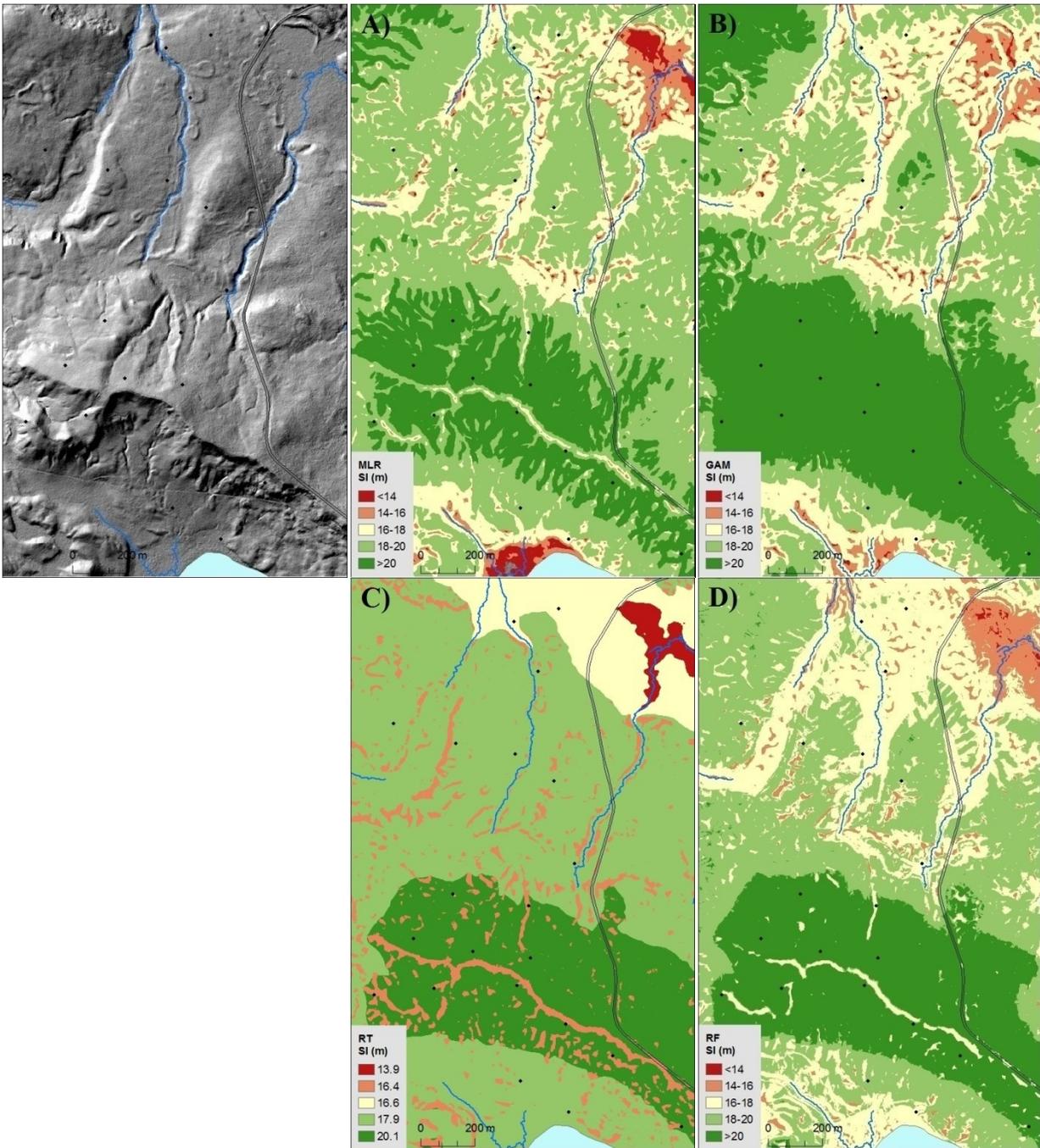


Figure 4.7. Hillshade and mapped white spruce SI (detail) generated based on selected: A) MLR model (Table 3.14), B) GAM model (Table 3.15), C) RT model (Table 3.16), and D) RF model (Table 3.17). Dots represent calibration sample plots; blue lines represent WAM predicted stream channels at 10 ha FIA. Map resolution is 1 m.

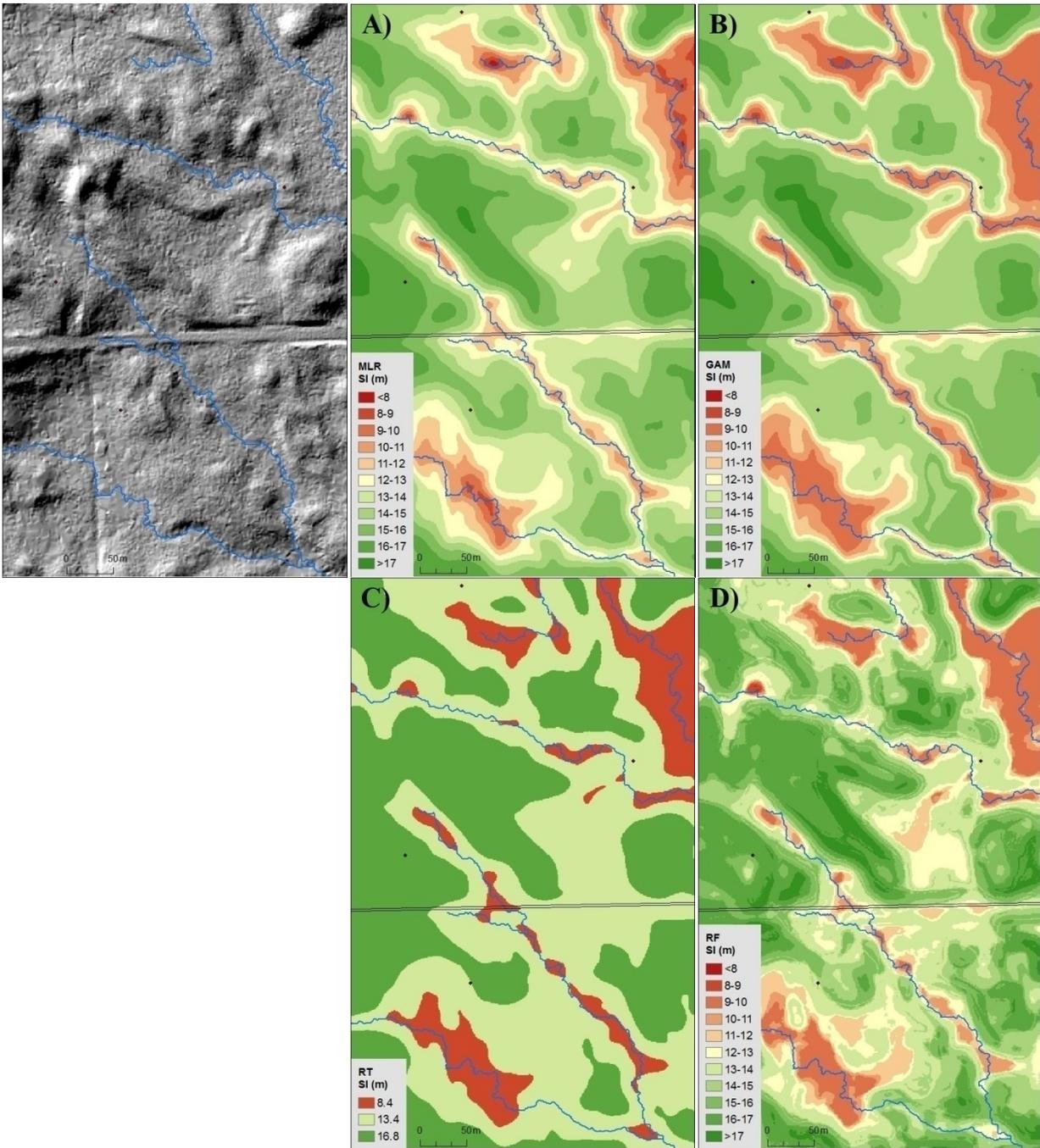


Figure 4.8. Hillshade and mapped lodgepole pine SI (detail) generated based on selected: A) MLR model (Table 3.10), B) GAM model (Table 3.11), C) RT model (Table 3.12), and D) RF model (Table 3.13). Dots represent calibration sample plots; blue lines represent WAM predicted stream channels at 2 ha FIA. Map resolution is 1 m.

5. CONCLUSIONS

Site Index is widely used as a measure of forest site productivity. In this study a geocentric approach to estimating SI was selected in order to utilize digital biophysical data for predicting and mapping fine scale variability in SI. LiDAR's generated Digital Elevation Model (DEM) and Wet Areas Mapping (WAM) provide remotely sensed environmental data at a 1 m resolution for most forested land in Alberta. Relationships between environmental factors and SI of three major commercial tree species in central Alberta (trembling aspen, lodgepole pine, white spruce) were examined in this study. Strong correlations were found between field determined soil properties and topography of the research area. Many studies have also found that soil moisture and vegetation characteristics in western boreal forests are driven by topography. However, this study revealed that forest productivity is also subject to topographic controls and could be explained using remotely sensed environmental data. In addition, different tree species respond differently due to contrasting autoecology.

Six different Flow Initiation Areas, from 0.5 ha to 10 ha, were tested to reveal optimal FIA for calculation of the DTW index for SI prediction. Results show that, for this hilly foothills landscape dominated by glacial till parent materials, DTW based on smaller FIA was better in estimating aspen SI, the largest size of FIA (10 ha) was best for spruce, while the size of FIA did not affect pine SI estimation.

A total of 36 species-specific Site Index models were developed for each of three data sources (DEM+WAM, WAM, field assessment) and four modeling methods (MLR, GAM, RT, RF). Predominance of topographic parameters over soil properties appeared based on examined predictor variables. In addition to revealing different major factors, different strength of relationship was found between species. Models developed in this study accounted for up to 73%, 70% and 50% of variation in SI for pine, aspen and spruce, respectively. Prediction accuracy of obtained models is consistent with other similar SI-environment studies. Poorer results for spruce result from the wide range in ages of sampled spruce stands, the lack of spruce stands across a full range of sites, and the small number of spruce stands where top height trees free of suppression could be sampled.

No significant differences in variation explained were observed between RS and GB models, while WAM data by itself explained most of the total amount of SI variation. All four statistical methods could be used in examining SI-environment relationship but with some advantages and disadvantages for each of them related to data and application specifics. In terms of best predictors, among RS variables DTW was selected by each statistical method for each species and in the most cases it is the strongest predictor in the model, while among GB variables different variables appeared as the most important for different species according to silvics specifics.

Potential application and limitations of the research were examined. SI maps for all species were produced and, despite some general issues, plausible relationships between terrain attributes and patterns of low and high predicted productivity are generally apparent. Therefore, obtained maps appear to adequately portray variation in productivity over short distances and are potentially applicable to forest growth and yield modeling and silviculture planning.

Our results demonstrated successful application of WAM variables in both accurate and high resolution SI estimation, and also indicate that DTW index is the strongest LiDAR derived SI predictor and that FA index is useful for explaining variation in soil properties in this study area. However, further refinements in application of WAM and LiDAR data in productivity prediction are possible. Using additional explanatory variables such as mapped soil properties, and other abiotic or biotic factors might improve SI model performance. Inclusion of first-return LiDAR metrics (Watt et al., 2015) could also improve the models. In future studies I suggest focussing on potential to utilize different kinds of digital data, such as combining active and passive remote sensing technology, in determination of the response side (eg. SI) of productivity models. Directly modeling height growth as a function of tree age and environmental data (Brandl et al., 2014) without using SI curves should also be considered. Using other productivity indicators might also help overcome some weakness of SI method, i.e. mapping aboveground live biomass by employing LiDAR indices and disturbance history (Zald et al., 2016). Finally, ecological niche is important for contribution to landscape productivity assessment, and application of WAM data could be very useful in predicting fine scale spatial distribution of major tree species, as done by Dunckel et al. (2015).

REFERENCES

- Aertsen, W., Kint, V., de Vos, B., Deckers, J., van Orshoven, J., Muys, B., 2012a. Predicting forest site productivity in temperate lowland from forest floor, soil and litterfall characteristics using boosted regression trees. *Plant Soil* 354, 157–172. doi:10.1007/s11104-011-1052-z
- Aertsen, W., Kint, V., Muys, B., Van Orshoven, J., 2012b. Effects of scale and scaling in predictive modelling of forest site productivity. *Environ. Model. Softw.* 31, 19–27. doi:10.1016/j.envsoft.2011.11.012
- Aertsen, W., Kint, V., Van Orshoven, J., Muys, B., 2011. Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). *Environ. Model. Softw.* 26, 929–937. doi:10.1016/j.envsoft.2011.01.003
- Aertsen, W., Kint, V., van Orshoven, J., Özkan, K., Muys, B., 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecol. Modell.* 221, 1119–1130. doi:10.1016/j.ecolmodel.2010.01.007
- Ågren, A.M., Lidberg, W., Strömberg, M., Ogilvie, J., Arp, P.A., 2014. Evaluating digital terrain indices for soil wetness mapping – a Swedish case study. *Hydrol. Earth Syst. Sci. Discuss.* 11, 4103–4129. doi:10.5194/hessd-11-4103-2014
- Albert, M., Schmidt, M., 2010. Climate-sensitive modelling of site-productivity relationships for Norway spruce (*Picea abies* (L.) Karst.) and common beech (*Fagus sylvatica* L.). *For. Ecol. Manage.* 259, 739–749. doi:10.1016/j.foreco.2009.04.039
- Andraž, Č., Matevski, V., Juvan, N., Kostadinovski, M., Košir, P., Marinšek, A., Paušič, A., Šilc, U., 2015. Transition along gradient from warm to mesic temperate forests evaluated by GAMM. *J. Plant Ecol.* 1–13. doi:10.1093/jpe/rtv069
- Antón-Fernández, C., Mola-Yudego, B., Dalsgaard, L., 2016. Climate sensitive site index models for Norway. *Can. J. For. Res.* 33, 1–33. doi:10.1139/cjfr-2015-0155
- Anyomi, K.A., Lorenzetti, F., Bergeron, Y., Leduc, A., 2015. Stand Dynamics, Humus Type and Water Balance Explain Aspen Long Term Productivity across Canada. *Forests* 6, 416–432. doi:10.3390/f6020416
- Anyomi, K.A., Raulier, F., Bergeron, Y., Mailly, D., 2013. The predominance of stand composition and structure over direct climatic and site effects in explaining aspen (*Populus tremuloides* Michaux) site index within boreal and temperate forests of western Quebec, Canada. *For. Ecol. Manage.* 302, 390–403. doi:10.1016/j.foreco.2013.03.035
- Anyomi, K.A., Raulier, F., Bergeron, Y., Mailly, D., Girardin, M.P., 2014. Spatial and temporal heterogeneity of forest site productivity drivers: a case study within the eastern boreal forests of Canada. *Landsc. Ecol.* 29, 905–918. doi:10.1007/s10980-014-0026-y

- Applequist, M.B., 1958. A simple pith locator for use with off-center increment cores. *J. For.* 56, 141.
- Beckingham, J.D., Corns, I.G.W., Archibald, J.H., 1996. *Field Guide to Ecosites of West-Central Alberta*. Nat.Resour. Can., Can. For. Serv. North.For.Cent., Edmonton, Alberta.
- Bergeron, Y., Chen, H.Y.H., Kenkel, N.C., Leduc, A.L., Macdonald, S.E., 2014. Boreal mixedwood stand dynamics: Ecological processes underlying multiple pathways. *For. Chron.* 90, 202–213. doi:10.5558/tfc2014-039
- Berrill, J., O’Hara, K.L., 2014. Estimating site productivity in irregular stand structures by indexing the basal area or volume increment of the dominant species. *Can. J. For. Res.* 44, 92–100. doi:10.1139/cjfr-2013-0230
- Berrill, J.-P., O’Hara, K.L., 2016. How do biophysical factors contribute to height and basal area development in a mixed multiaged coast redwood stand? *Forestry* 89, 170–181. doi:10.1093/forestry/cpv049
- Bokalo, M., Stadt, K., Comeau, P., Titus, S., 2013. The Validation of the Mixedwood Growth Model (MGM) for Use in Forest Management Decision Making. *Forests* 4, 1–27. doi:10.3390/f4010001
- Bontemps, J.-D., Bouriaud, O., 2014. Predictive approaches to forest site productivity: recent trends, challenges and future perspectives. *Forestry* 87, 109–128. doi:10.1093/forestry/cpt034
- Bourg, N.A., Mcshea, W.J., Gill, D.E., 2005. Putting a CART before the search: Successful habitat prediction for a rare forest herb. *Ecology* 86, 2793–2804. doi:10.1890/04-1666
- Brandl, S., Falk, W., Klemmt, H.-J., Stricker, G., Bender, A., Rötzer, T., Pretzsch, H., 2014. Possibilities and Limitations of Spatially Explicit Site Index Modelling for Spruce Based on National Forest Inventory Data and Digital Maps of Soil and Climate in Bavaria (SE Germany). *Forests* 2626–2646. doi:10.3390/f5112626
- Bravo-Oviedo, A., Roig, S., Bravo, F., Montero, G., del-Rio, M., 2011. Environmental variability and its relationship to site index in Mediterranean maritime pine. *For. Syst.* 20, 50–64. doi:10.5424/fs/2011201-9106
- Breiman, L., 2001. Random forests. *Mach. Learn.* 5–32. doi:10.1023/A:1010933404324
- Bueis, T., Bravo, F., Pando, V., Turrión, M., 2016. Relationship between environmental parameters and *Pinus sylvestris* L. site index in forest plantations in northern Spain acidic plateau. *iForest - Biogeosciences For.* 9, 394–401. doi:10.3832/ifor1600-008
- Burns, R.M., Honkala, B.H., 1990. *Silvics of North America: 1. Conifers; 2. Hardwoods*. Agric. Handb. 654. U.S. Dep. Agric. For. Serv. Washington, DC. vol.2, 877 p.
- Campbell, D.M.H., White, B., Arp, P.A., 2013. Modeling and mapping soil resistance to penetration and rutting using LiDAR-derived digital elevation data. *J. Soil Water Conserv.* 68, 460–473. doi:10.2489/jswc.68.6.460

- Canfor/WeyCo Report, 2008. Regenerated stand productivity in north central Alberta. Report #2. Canadian Forest Products.
- Caouette, J.P., Steel, E.A., Hennon, P.E., Cunningham, P.G., Pohl, C.A., Schrader, B.A., 2016. Influence of elevation and site productivity on conifer distributions across Alaskan temperate rainforests. *Can. J. For. Res.* 261, 249–261.
- Chen, H.Y., Klinka, K., Kabzems, R.D., 1998. Site index, site quality, and foliar nutrients of trembling aspen: relationships and predictions. *Can. J. For. Res.* 28, 1743–1755. doi:10.1139/x98-154
- Chen, H.Y., Krestov, P. V, Klinka, K., 2002. Trembling aspen site index in relation to environmental measures of site quality at two spatial scales. *Can. J. For. Res.* 32, 112–119. doi:10.1139/x01-179
- Chen, H.Y., Popadiouk, R. V, 2002. Dynamics of North American boreal mixedwoods. *Environ. Rev.* 10, 137–166. doi:10.1139/a02-007
- Comeau, P.G., Kimmins, 1989. Above- and below-ground biomass and production of lodgepole pine on sites with differing soil moisture regimes. *Can. J. For. Res.* 19, 447–454. doi:10.1017/CBO9781107415324.004
- Coops, N., Tompaski, P., Nijland, W., 2016. A forest structure habitat index based on airborne laser scanning data. *Ecol. Indic.* 67, 346–357. doi:10.1016/j.actamat.2015.02.029
- Coops, N.C., Gaulton, R., Waring, R.H., 2011. Mapping site indices for five Pacific Northwest conifers using a physiologically based model. *Appl. Veg. Sci.* 14, 268–276. doi:10.1111/j.1654-109X.2010.01109.x
- Coops, N.C., Waring, R.H., 2011. A process-based approach to estimate lodgepole pine (*Pinus contorta* Dougl.) distribution in the Pacific Northwest under climate change. *Clim. Change* 105, 313–328. doi:10.1007/s10584-010-9861-2
- Danescu, A., Albrecht, A.T., Bauhus, J., 2016. Structural diversity promotes productivity of mixed, uneven-aged forests in southwestern Germany. *Oecologia* 1–15. doi:10.1007/s00442-016-3623-4
- DeRose, J.R., Gardner, R.S., 2010. Technique to Improve Visualization of Elusive Tree-Ring Boundaries in Aspen (*Populus tremuloides*). *Tree-Ring Res.* 66, 75–78. doi:10.3959/2009-11.1
- Dilts, T.E., 2015. Topography Tools for ArcGIS 10.1. Available at: <http://www.arcgis.com/home/item.html?id=b13b3b40fa3c43d4a23a1a09c5fe96b9>.
- Dunckel, K., Weiskittel, A., Fiske, G., Sader, S.A., Latty, E., Arnett, A., 2015. Linking remote sensing and various site factors for predicting the spatial distribution of eastern hemlock occurrence and relative basal area in Maine, USA. *For. Ecol. Manage.* 358, 180–191. doi:10.1016/j.foreco.2015.09.012
- Eckhart, T., Hintsteiner, W., Lair, G., Van Loo, M., Hasenauer, H., 2014. The impact of soil

- conditons on the growth of douglas-fir in Austria. *Austrian J. For. Sci.* 131, 107–128.
- Eisenbies, M.H., Burger, J.A., Aust, W.M., Patterson, S.C., 2006. Assessing Change in Soil-Site Productivity of Intensively Managed Loblolly Pine Plantations. *Soil Sci. Soc. Am. J.* 70, 1037. doi:10.2136/sssaj2004.0065er
- Farrelly, N., Dhubhain, A.N., Nieuwenhuis, M., 2011. Sitka spruce site index in response to varying soil moisture and nutrients in three different climate regions in Ireland. *For. Ecol. Manage.* 262, 2199–2206. doi:10.1016/j.foreco.2011.08.012
- Filipescu, C.N., Comeau, P.G., 2007. Aspen competition affects light and white spruce growth across several boreal sites in western Canada. *Can. J. For. Res.* 37, 1701–1713. doi:10.1139/X07-011
- Fries, A., Lindgren, D., Ying, C.C., Ruotsalainen, S., Lindgren, K., Elfving, B., Karlmat, U., 2000. The effect of temperature on site index in western Canada and Scandinavia estimated from IUFRO Pinus contorta provenance experiments. *Can. J. For. Res.* 30, 921–929.
- Gray, L.K., Hamann, A., 2013. Tracking suitable habitat for tree populations under climate change in western North America. *Clim. Change* 117, 289–303. doi:10.1007/s10584-012-0548-8
- Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecol. Modell.* 157, 89–100. doi:10.1016/S0304-3800(02)00204-1
- Hamann, A., Wang, T., Spittlehouse, D.L., Murdock, T.Q., 2013. A comprehensive, high-resolution database of historical and projected climate surfaces for western North America. *Bull. Am. Meteorol. Soc.* 94, 1307–1309. doi:10.1175/BAMS-D-12-00145.1
- Hamel, B., Bélanger, N., Paré, D., 2004. Productivity of black spruce and Jack pine stands in Quebec as related to climate, site biological features and soil properties. *For. Ecol. Manage.* 191, 239–251. doi:10.1016/j.foreco.2003.12.004
- Hijmans, R., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O., Bevan, A., Racine, E., Shortridge, A., 2016. Package “raster”. CRAN -R.2.5-8. Available at: <http://cran.r-project.org/package=raster>.
- Hiltz, D., Gould, J., White, B., Ogilvie, J., Arp, P., 2012. Modeling and mapping vegetation type by soil moisture regime across boreal landscapes, in: *Restoration and Reclamation of Boreal Ecosystems: Attaining Sustainable Development*. Cambridge Univ. Press, New York, NY., pp. 56–75.
- Hogg, E.H., Barr, A.G., Black, T.A., 2013. A simple soil moisture index for representing multi-year drought impacts on aspen productivity in the western Canadian interior. *Agric. For. Meteorol.* 178-179, 173–182. doi:10.1016/j.agrformet.2013.04.025
- Hogg, E.H., Brandt, J.P., Michaelian, M., 2008. Impacts of a regional drought on the productivity, dieback, and biomass of western Canadian aspen forests. *Can. J. For. Res.* 38, 1373–1384. doi:10.1139/X08-001

- Hogg, E.H., Hart, M., Lieffers, V.J., 2002. White tree rings formed in trembling aspen saplings following experimental defoliation. *Can. J. For. Res.* 32, 1929–1934. doi:10.1139/x02-114
- Hostin, P.J., Titus, S.J., 1996. Indirect site productivity models for white spruce in Alberta's boreal mixedwood forest. *For. Chron.* 72, 73–79.
- Huang, S., Meng, S.X., Yang, Y., 2009. A Growth and Yield Projection System (GYPSY) for Natural and Post-harvest Stands in Alberta. Alberta Sustain. Resour. Dev. Tech. Rep. T/216, 1–22.
- Huang, S., Titus, S.J., 1993. An index of site productivity for uneven aged or mixed species stands. *Can. J. For. Res.* 23, 558–562. doi:10.1139/x93-074
- Illés, G., Kovács, G., Heil, B., 2011. Comparing and evaluating digital soil mapping methods in a Hungarian forest reserve. *Can. J. Soil Sci.* 91, 615–626. doi:10.4141/cjss2010-007
- Iverson, L.R., Dale, M.E., Scott, C.T., Prasad, A., 1997. A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests (U.S.A.). *Landsc. Ecol.* 12, 331–348. doi:10.1023/a:1007989813501
- Jenness, J., 2007. Some Thoughts on Analyzing Topographic Habitat Characteristics Surface Area. Available at: http://www.jennessent.com/downloads/topographic_analysis_online.pdf.
- Jiang, H., Radtke, P.J., Weiskittel, A.R., Coulston, J.W., Guertin, P.J., 2015. Climate- and soil-based models of site productivity in eastern US tree species. *Can. J. For. Res.* 45, 325–342. doi:10.1139/cjfr-2014-0054
- Kabzems, R., Klinka, K., 1987. Initial quantitative characterization of soil nutrient regimes. II. Relationships among soils, vegetation, and site index. *Can. J. For. Res.* 17, 1565–1571. doi:10.1017/CBO9781107415324.004
- Kane, V.R., Lutz, J.A., Alina Cansler, C., Povak, N.A., Churchill, D.J., Smith, D.F., Kane, J.T., North, M.P., 2015. Water balance and topography predict fire and forest structure patterns. *For. Ecol. Manage.* 338, 1–13. doi:10.1016/j.foreco.2014.10.038
- Kayahara, G.J., Klinka, K., Marshall, P.L., 1998. Testing site index-site-factor relationships for predicting *Pinus contorta* and *Picea engelmannii* x *P. glauca* productivity in central British Columbia, Canada. *For. Ecol. Manage.* 110, 141–150. doi:10.1016/S0378-1127(98)00279-5
- Laamrani, A., Valeria, O., Bergeron, Y., Fenton, N., Cheng, L.Z., Anyomi, K., 2014. Effects of topography and thickness of organic layer on productivity of black spruce boreal forests of the canadian clay belt region. *For. Ecol. Manage.* 330, 144–157. doi:10.1016/j.foreco.2014.07.013
- Landhausser, S.M., Deshaies, D., Lieffers, V.J., 2010. Disturbance facilitates rapid range expansion of aspen into higher elevations of the Rocky Mountains under a warming climate. *J. Biogeogr.* 37, 68–76. doi:10.1111/j.1365-2699.2009.02182.x
- Latta, G., Temesgen, H., Adams, D., Barrett, T., 2010. Analysis of potential impacts of climate change on forests of the United States Pacific Northwest. *For. Ecol. Manage.* 259, 720–729.

doi:10.1016/j.foreco.2009.09.003

- Latta, G., Temesgen, H., Barrett, T.M., 2009. Mapping and imputing potential productivity of Pacific Northwest forests using climate variables. *Can. J. For. Res.* 39, 1197–1207. doi:10.1139/X09-046
- Latutrie, M., Mérian, P., Picq, S., Bergeron, Y., Tremblay, F., 2015. The effects of genetic diversity, climate and defoliation events on trembling aspen growth performance across Canada. *Tree Genet. Genomes* 11, 96. doi:10.1007/s11295-015-0925-3
- Laubhann, D., Sterba, H., Reinds, G.J., De Vries, W., 2009. The impact of atmospheric deposition and climate on forest growth in European monitoring plots: An individual tree growth model. *For. Ecol. Manage.* 258, 1751–1761. doi:10.1016/j.foreco.2008.09.050
- Liaw, A., Wiener, M., Andy Liaw, M., 2015. Package “randomForest.” Available at: <https://www.stat.berkeley.edu/~breiman/RandomForests/>.
- Lieffers, V.J., Macmillan, R.B., MacPherson, D., Branter, K., Stewart, J.D., 1996. Semi-natural and intensive silvicultural systems for the boreal mixedwood forest. *For. Chron.* 72, 286–292.
- Lieffers, V.J., Messier, C., Burton, P.J., Ruel, J.-C., Grover, B.E., 2003. Nature-based silviculture for sustaining a variety of boreal forest values, in: Chapter 13: Towards Sustainable Management of the Boreal Forest. Edited by P.J. Burton, C. Messier. D.W. Smith and W.L. Adamowicz. NRC Research Press. Ottawa, ON, Canada.
- Little, 2001. Soil Productivity of White Spruce - Aspen Mixedwood Ecosystems along Toposequences in West-Central Alberta. MSc thesis. Dep. Renew. Resour. Univ. Alberta.
- Logan, M., 2010. Biostatistical Design and Analysis Using R: a practical guide. Wiley - Blackwell, Oxford, UK.
- Loh, W.-Y., 2011. Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 1, 14–23. doi:10.1002/widm.8
- Lumley, T., 2015. Package “leaps.” Available at: <http://cran.r-project.org/web/packages/leaps/leaps.pdf>.
- Ma, M., Jiang, H., Liu, S., Zhu, C., Liu, Y., Wang, J., 2006. Estimation of forest-ecosystem site index using remote-sensed data. *Acta Ecol. Sin.* 26, 2810–2815. doi:10.1016/S1872-2032(06)60045-0
- Macdonald, S.E., Lecomte, N., Bergeron, Y., Brais, S., Chen, H.Y.H., Comeau, P., Drapeau, P., Lieffers, V.J., Quideau, S., Spence, J., Work, T., 2010. Ecological implications of changing the composition of boreal mixedwood forests. A State of Knowledge Report. Sustainable Forest Management Network, Edmonton, Alberta.
- Mah, S., Nigh, G., 2015. SIBEC First and Second-generation Approximation Estimates : backgrounder. Available at: <http://www2.gov.bc.ca/gov/content/environment/research-monitoring-reporting/research/sibec>.

- Maindonald, J., Braun, W.J., 2003. *Data Analysis and Graphics Using R - an Example-Based Approach* (3rd edition). Cambridge Univ. Press 1–565.
- McCune, B., Keon, D., 2002. Equations for potential annual direct incident radiation and heat load. *J. Veg. Sci.* 13, 603–606. doi:10.1111/j.1654-1103.2002.tb02087.x
- McKenney, D.W., Pedlar, J.H., 2003. Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. *For. Ecol. Manage.* 175, 497–507. doi:10.1016/S0378-1127(02)00186-X
- Milborrow, S., 2015. Package “plotmo.” Available at: <http://www.milbo.users.sonic.net>.
- Mitsuda, Y., 2016. Relationship between the Parameter of Photosynthetic Rate of a Process-Based Growth Model for Sugi (*Cryptomeria japonica*) Planted Forests and Topographic Factors. *Formath* 15, 10–19. doi:10.15684/formath.15.002
- Mitsuda, Y., Ito, S., Sakamoto, S., 2007. Predicting the site index of sugi plantations from GIS-derived environmental factors in Miyazaki Prefecture. *J. For. Res.* 12, 177–186. doi:10.1007/s10310-007-0004-1
- Monserud, R.A., Huang, S., Yang, Y., 2006. Predicting lodgepole pine site index from climatic parameters in Alberta. *For. Chron.* 82, 562–571.
- Monserud, R.A., Rehfeldt, G.E., 1990. Genetic and environmental components of variation of site index in inland Douglas-fir. *For. Sci.* 36, 1–9.
- Monserud, R.A., Yang, Y., Huang, S., Tchebakova, N., 2008. Potential change in lodgepole pine site index and distribution under climatic change in Alberta. *Can. J. For. Res.* 38, 343–352. doi:10.1139/X07-166
- Murphy, P.N.C., Ogilvie, J., Arp, P., 2009. Topographic modelling of soil moisture conditions: A comparison and verification of two models. *Eur. J. Soil Sci.* 60, 94–109. doi:10.1111/j.1365-2389.2008.01094.x
- Murphy, P.N.C., Ogilvie, J., Connor, K., Arp, P.A., 2007. Mapping wetlands: A comparison of two different approaches for New Brunswick, Canada. *Wetlands* 27, 846–854. doi:10.1672/0277-5212(2007)27[846:MWACOT]2.0.CO;2
- Murphy, P.N.C., Ogilvie, J., Meng, F.R., White, B., Bhatti, J.S., Arp, P.A., 2011. Modelling and mapping topographic variations in forest soils at high resolution: A case study. *Ecol. Modell.* 222, 2314–2332. doi:10.1016/j.ecolmodel.2011.01.003
- Natural Resources Canada, 2015. *The State of Canada’s Forests: Annual Report 2015*. Available at: <http://www.nrcan.gc.ca/forests/report/16496>.
- Newton, P., 2015. Occurrence of Density-Dependent Height Repression within Jack Pine and Black Spruce Populations. *Forests* 6, 2450–2468. doi:10.3390/f6072450
- Nigh, G., 2002. Site Index Conversion Equations for Mixed Trembling Aspen and White Spruce Stands in Northern British. *Silva Fenn.* 36, 789–797.

- Nigh, G., Love, B.A., 1997. Site Index Adjustment for Old-growth Coastal Western Hemlock Stands in the Kalum Forest District. BC Min. For., Res. Br., Victoria, BC Work. Pap. 27.
- Nigh, G.D., 1995. Variable growth intercept models for lodgepole pine in the Sub-Boreal Spruce biogeoclimatic zone, British Columbia. BC Min. For., Victoria, BC. Res. Rep. 02. 20.
- Nigh, G.D., Ying, C.C., Qian, H., 2004. Climate and Productivity of Major Conifer Species in the Interior of British Columbia, Canada. *For. Sci.* 50, 659–671.
- Nijland, W., Coops, N.C., Macdonald, S.E., Nielsen, S.E., Bater, C.W., White, B., Ogilvie, J., Stadt, J., 2015. Remote sensing proxies of productivity and moisture predict forest stand type and recovery rate following experimental harvest. *For. Ecol. Manage.* 357, 239–247. doi:10.1016/j.foreco.2015.08.027
- Nijland, W., Nielsen, S.E., Coops, N.C., Wulder, M. a., Stenhouse, G.B., 2014. Fine-spatial scale predictions of understory species using climate- and LiDAR-derived terrain and canopy metrics. *J. Appl. Remote Sens.* 8, 083572. doi:10.1117/1.JRS.8.083572
- Nothdurft, A., Wolf, T., Ringeler, A., Böhner, J., Saborowski, J., 2012. Spatio-temporal prediction of site index based on forest inventories and climate change scenarios. *For. Ecol. Manage.* 279, 97–111. doi:10.1016/j.foreco.2012.05.018
- Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biol.* 83, 171–193. doi:10.1086/587826
- Oltean, G.S., 2015. Estimation of site index and soil properties using the topographic depth-to-water index. MSc thesis. Dep. Renew. Resour. Univ. Alberta.
- Oltean, G.S., Comeau, P.G., White, B., 2016. Linking the Depth-to-Water Topographic Index to Soil Moisture on Boreal Forest Sites in Alberta. *For. Sci.* 62, 1–12. doi:10.5849/forsci.15-054
- Osika, D.E., Stadt, K.J., Comeau, P.G., MacIsaac, D.A., 2013. Sixty-year effects of deciduous removal on white spruce height growth and site index in the Western Boreal. *Can. J. For. Res.* 43, 139–148. doi:10.1139/cjfr-2012-0169
- Packalén, P., Mehtätalo, L., Maltamo, M., 2011. ALS-based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed-effect model that accounts for the clone effect. *Ann. For. Sci.* 68, 1085–1092. doi:10.1007/s13595-011-0124-9
- Paré, D., Bergeron, Y., Longpré, M.-H., 2001. Potential Productivity of Aspen Cohorts Originating from Fire, Harvesting, and Tree-fall Gaps on Two Deposit Types in Northwestern Quebec. *Can. J. For. Res.* 31, 1067–1073. doi:10.1139/cjfr-31-6-1067
- Piedallu, C., Gégout, J.C., Bruand, A., Seynave, I., 2011. Mapping soil water holding capacity over large areas to predict potential production of forest stands. *Geoderma* 160, 355–366. doi:10.1016/j.geoderma.2010.10.004
- Pinno, B.D., Bélanger, N., 2011. Estimating trembling aspen productivity in the boreal transition ecoregion of saskatchewan using site and soil variables. *Can. J. Soil Sci.* 91, 661–669.

doi:10.4141/CJSS10082

- Pinno, B.D., Paré, D., Guindon, L., Bélanger, N., 2009. Predicting productivity of trembling aspen in the Boreal Shield ecozone of Quebec using different sources of soil and site information. *For. Ecol. Manage.* 257, 782–789. doi:10.1016/j.foreco.2008.09.058
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199. doi:10.1007/s10021-005-0054-1
- Quinn, G.P., Keough, M.J., 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge Univ. Press 1–537.
- Robichaud, Methven, 1993. The effect of site quality on the timing of stand breakup, tree longevity, and the maximum attainable height of black spruce. *Can. J. For. Res.* 23, 1514–1519. doi:10.1017/CBO9781107415324.004
- Ryu, S.R., Chen, J., Zheng, D., Bresee, M.K., Crow, T.R., 2006. Simulating the effects of prescribed burning on fuel loading and timber production (EcoFL) in managed northern Wisconsin forests. *Ecol. Modell.* 196, 395–406. doi:10.1016/j.ecolmodel.2006.02.013
- Sabatia, C.O., Burkhart, H.E., 2014. Predicting site index of plantation loblolly pine from biophysical variables. *For. Ecol. Manage.* 326, 142–156. doi:10.1016/j.foreco.2014.04.019
- Seynave, I., Gégout, J.-C., Hervé, J.-C., Dhôte, J.-F., Drapier, J., Bruno, É., Dumé, G., 2005. Site Index Prediction By Environmental Factors and Understorey Vegetation: a Two-Scale Approach Based on Survey Databases. *Can. J. For. Res.* 35, 1669–1678. doi:10.1139/x05-088
- Sharma, R.P., Brunner, A., Eid, T., 2012. Site index prediction from site and climate variables for Norway spruce and Scots pine in Norway. *Scand. J. For. Res.* 27, 619–636. doi:10.1080/02827581.2012.685749
- SIBEC Sampling and Data Standards, 2009. . B.C. Minist. For. , Res. Branch. Available <http://www2.gov.bc.ca/assets/gov/environment/research-monitoring-and-reporting/research/sibec-documents/standards.pdf>.
- Silva, C.A., Klauberg, C., Hudak, A.T., Vierling, L.A., Liesenberg, V., Carvalho, S.P.C. e, Rodriguez, L.C.E., 2016. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry* 0, 1–12. doi:10.1093/forestry/cpw016
- Skovsgaard, J.P., Vanclay, J.K., 2013. Forest site productivity: A review of spatial and temporal variability in natural site conditions. *Forestry* 86, 305–315. doi:10.1093/forestry/cpt010
- Skovsgaard, J.P., Vanclay, J.K., 2008. Forest site productivity: A review of the evolution of dendrometric concepts for even-aged stands. *Forestry* 81, 13–31. doi:10.1093/forestry/cpm041
- Smith, D.M., Larson, B.C., Kelty, M.J., Ashton, P.M.S., 1997. *The practice of silviculture:*

- Applied forest ecology. Ninth. Ninth Ed. ed. John Wiley Sons, Inc., New York.
- Socha, J., 2008. Effect of topography and geology on the site index of *Picea abies* in the West Carpathian, Poland. *Scand. J. For. Res.* 23, 203–213. doi:10.1080/02827580802037901
- Speer, H.J., 2010. *Fundamentals of Tree-Ring Research*. Univ. Arizona Press 1–333.
- Subedi, S., Fox, T.R., 2016. Predicting Loblolly Pine Site Index from Soil Properties Using Partial Least-Squares Regression. *For. Sci.* 62, 1–8.
- Swenson, J.J., Waring, R.H., Fan, W., Coops, N., 2005. Predicting site index with a physiologically based growth model across Oregon, USA. *Can. J. For. Res.* 35, 1697–1707. doi:10.1139/x05-089
- Szwaluk, K.S., Strong, W.L., 2003. Near-surface soil characteristics and understory plants as predictors of *Pinus contorta* site index in southwestern Alberta, Canada. *For. Ecol. Manage.* 176, 13–24. doi:10.1016/S0378-1127(02)00228-1
- Therneau, T., Atkinson, B., Ripley, B., Ripley, M.B., 2015. rpart: Recursive Partitioning and Regression Trees. R Packag. version 4.1-10. Available <https://cran.r-project.org/web/packages/rpart/rpart.pdf>.
- Thornley, J.H.M., Cannell, M.G.R., 2004. Long-term effects of fire frequency on carbon storage and productivity of boreal forests: a modeling study. *Tree Physiol.* 24, 765–773.
- Tompalski, P., Coops, N.C., White, J.C., Wulder, M.A., 2015a. Augmenting site index estimation with airborne laser scanning data. *For. Sci.* 61, 861–873. doi:10.5849/forsci.14-175
- Tompalski, P., Coops, N.C., White, J.C., Wulder, M.A., Pickell, P.D., 2015b. Estimating Forest Site Productivity Using Airborne Laser Scanning Data and Landsat Time Series. *Can. J. Remote Sens.* 41, 232–245. doi:10.1080/07038992.2015.1068686
- Touw, W.G., Bayjanov, J.R., Overmars, L., Backus, L., Boekhorst, J., Wels, M., van Hijum, S.A.F.T., 2013. Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? *Brief. Bioinform.* 14, 315–326. doi:10.1093/bib/bbs034
- Travis Swaim, J., Dey, D.C., Saunders, M.R., Weigel, D.R., Thornton, C.D., Kabrick, J.M., Jenkins, M.A., 2016. Predicting the height growth of oak species (*Quercus*) reproduction over a 23-year period following clearcutting. *For. Ecol. Manage.* 364, 101–112. doi:10.1016/j.foreco.2016.01.005
- Ung, C.H., Bernier, P.Y., Raulier, F., Fournier, R.A., Lambert, M.C., Régnière, J., 2001. Biophysical site indices for shade tolerant and intolerant boreal species. *For. Sci.* 47, 83–95.
- Wang, G., 1998. Is height of dominant trees at a reference diameter an adequate measure of site quality? *For. Ecol. Manage.* 112, 49–54. doi:10.1016/S0378-1127(98)00315-6
- Wang, G.G., 1995. White spruce site index in relation to soil, understory vegetation, and foliar nutrients. *Can. J. For. Res.* 25, 29–38. doi:10.1017/CBO9781107415324.004

- Wang, G.G., Huang, S., Monserud, R.A., Klos, R.J., 2004. Lodgepole pine site index in relation to synoptic measures of climate, soil moisture and soil nutrients. *For. Chron.* 80, 678–686. doi:10.5558/tfc80678-6
- Wang, G.G., Klinka, K., 1996. Use of synoptic variables in predicting white spruce site index. *For. Ecol. Manage.* 80, 95–105. doi:10.1016/0378-1127(95)03630-X
- Wang, T., Campbell, E.M., O'Neill, G. a., Aitken, S.N., 2012a. Projecting future distributions of ecosystem climate niches: Uncertainties and management applications. *For. Ecol. Manage.* 279, 128–140. doi:10.1016/j.foreco.2012.05.034
- Wang, T., Hamann, A., Spittlehouse, D.L., Murdock, T.Q., 2012b. ClimateWNA-high-resolution spatial climate data for western North America. *J. Appl. Meteorol. Climatol.* 51, 16–29. doi:10.1175/JAMC-D-11-043.1
- Wang, T., Wang, G., Innes, J., Nitschke, C., Kang, H., 2015. Climatic niche models and their consensus projections for future climates for four major forest tree species in the Asia–Pacific region. *For. Ecol. Manage.* doi:10.1016/j.foreco.2015.08.004
- Wang, Y., LeMay, V.M., Baker, T.G., 2007. Modelling and prediction of dominant height and site index of *Eucalyptus globulus* plantations using a nonlinear mixed-effects model approach. *Can. J. For. Res.* 37, 1390–1403. doi:10.1139/X06-282
- Wang, Y., Raulier, F., Ung, C.H., 2005. Evaluation of spatial predictions of site index obtained by parametric and nonparametric methods - A case study of lodgepole pine productivity. *For. Ecol. Manage.* 214, 201–211. doi:10.1016/j.foreco.2005.04.025
- Waring, R.H., Milner, K.S., Jolly, W.M., Phillips, L., McWethy, D., 2006. Assessment of site index and forest growth capacity across the Pacific and Inland Northwest U.S.A. with a MODIS satellite-derived vegetation index. *For. Ecol. Manage.* 228, 285–291. doi:10.1016/j.foreco.2006.03.019
- Watt, M.S., Dash, J.P., Bhandari, S., Watt, P., 2015. Comparing parametric and non-parametric methods of predicting Site Index for radiata pine using combinations of data derived from environmental surfaces, satellite imagery and airborne laser scanning. *For. Ecol. Manage.* 357, 1–9. doi:10.1016/j.foreco.2015.08.001
- Watt, M.S., Palmer, D.J., Kimberley, M.O., Hock, B.K., Payn, T.W., Lowe, D.J., 2010a. Development of models to predict *Pinus radiata* productivity throughout New Zealand. *Can. J. For. Res. Can. Rech. For.* 40, 488–499. doi:10.1139/x09-207
- Watt, M.S., Palmer, D.J., Kimberley, M.O., Höck, B.K., Payn, T.W., Lowe, D.J., 2010b. Development of models to predict *Pinus radiata* productivity throughout New Zealand. *Can. J. For. Res.* 40, 488–499. doi:10.1139/X09-207
- Weiskittel, Crookston, N., Radtke, P., 2011a. Linking climate, gross primary productivity, and site index across forests of the western United States. *Can. J. For. Res.* 41, 1710–1721. doi:10.1139/x11-086
- Weiskittel, Hann, D.W., Kershaw, J., Vanclay, J.K., 2011b. Forest growth and yield modeling.

John Wiley & Sons, Hoboken, N.J.

- Weiss, A., 2001. Topographic position and landforms analysis. Poster Present. ESRI User Conf. San Diego, CA 64, 227 – 245.
- White, B., Ogilvie, J., Campbell, D.M.H., Hiltz, D., Gauthier, B., Chisholm, H.K., Wen, H.K., Murphy, P.N.C., Arp, P.A., 2012. Using the Cartographic Depth-to-Water Index to Locate Small Streams and Associated Wet Areas across Landscapes. *Can. Water Resour. J.* 37, 333–347. doi:10.4296/cwrj2011-909
- Wickham, M.H., 2014. Package “ggplot2.” Available at: <http://ggplot2.org>, <https://github.com/hadley/ggplot2>.
- Wood, A.S., Wood, M.S., 2013. Package “mgcv.” Available at: <https://cran.r-project.org/web/packages/mgcv/index.html>.
- Wood, S.N., 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, Boca Rat. 1–391.
- Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., White, J.C., 2008. The role of LiDAR in sustainable forest management. *For. Chron.* 84, 807–826. doi:10.5558/tfc84807-6
- Yeboah, D., Chen, H.Y.H., Kingston, S., 2016. Tree species richness decreases while species evenness increases with disturbance frequency in a natural boreal forest landscape. *Ecol. Evol.* 6, 842–850. doi:10.1002/ece3.1944
- Zald, H.S.J., Spies, T.A., Seidl, R., Pabst, R.J., Olsen, K.A., Steel, E.A., 2016. Complex mountain terrain and disturbance history drive variation in forest aboveground live carbon density in the western Oregon Cascades, USA. *For. Ecol. Manage.* 366, 193–207. doi:10.1016/j.foreco.2016.01.036
- Zellweger, F., Braunisch, V., Morsdorf, F., Baltensweiler, A., Abegg, M., Roth, T., Bugmann, H., Bollmann, K., 2015. Disentangling the effects of climate, topography, soil and vegetation on stand-scale species richness in temperate forests. *For. Ecol. Manage.* 349, 36–44. doi:10.1016/j.foreco.2015.04.008
- Ziegler, A., König, I.R., 2014. Mining data with random forests: current options for real-world applications. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 4, 55–63. doi:10.1002/widm.1114
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A. a, Smith, G.M., 2009. *Mixed Effects Models and Extensions in Ecology with R*. Stat. Biol. Heal. Springer New York. 1–579. doi:10.1007/978-0-387-87458-6

APPENDIX

Soil Moisture Regime (SMR) maps produced using Random Forest model applied at areas representing aspen blocks (left map) and pine blocks (right map). Map Legend: 2 - xeric, 4 - submesic, 5 - mesic, 6 - subhygric, 7 - hygric, 8 - subhydric SMR. Black outlined circles representing sample plots with colors assigned based on observed SMR values. The Random Forest model was built using five predictor variables (DTW_1, FA, SLO, PLCURV, TWI) and has 30.4% OOB estimate of error rate (misclassification rate based on validation samples).

