Norwegian University of Life Sciences • Universitetet for miljø- og biovitenskap
Department of Ecology and Natural Resource Management
Philosophiae Doctor (PhD) Thesis 2007:30

# Uneven-aged forestry in Norway: Inventory and management models

## Ole Martin Bollandsås

# Uneven-aged forestry in Norway: Inventory and management models

Ole Martin Bollandsås
PhD thesis
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences

Ås 2007

ISSN 1503-1667
ISBN 978-82-575-0785-5
Thesis: 2007:30

# Preface

This thesis has been written while I was a PhD-student at the Department of Ecology and Natural Resource Management at the Norwegian University of Life Sciences. Together with a program of formal courses and the dissertation, it is the completion of the doctoral degree Philosophiae Doctor.

The work leading to this thesis started in March 2003. My supervisor for the entire time has been Professor Erik Næsset. In 2006 also Professor Tron Eid and Associate Professor Terje Gobakken became a part of the team of supervisors according to the regulations of the PhD program. Erik, Tron, and Terje has supported me trough discussions, revisions of manuscripts, teaching etc. I owe them great thanks for always being there. The project has been financed by the Research Council of Norway (research project no. 153185/110).

During the spring semester of 2007 I stayed at the University of Wisconsin in Madison, USA. The paper dealing with matrix models was written during this stay in close collaboration with Professor Joseph Buongiorno. It was a great experience for me staying there working with Professor Buongiorno. Thanks also to Terje Gobakken and Lone, Helle, and Henrik Ross Gobakken for letting me stay in their house during my visit to Madison.

I would also like to thank my fellow PhD student Even Bergseng who is an excellent discussion partner of great knowledge that has helped me a lot. My daughter Henriette also deserves great thanks for always being patient and supportive.

# Contents

# Abstract

This thesis presents management models applicable for uneven-aged forest structures in Norway. The framework is a matrix model consisting of several sub-models estimated from National Forest Inventory (NFI). The matrix model operates on diameter classes, projecting future outcomes by means of transition probabilities. Within the thesis, models for diameter distributions, recruitment, diameter growth, mortality, and tree height were developed. These four latter models were implemented in a computer tool and organized as a matrix model.

Diameter distributions are not registered in ordinary forest inventories. Such distributions are necessary as input data for diameter class models such as the one presented in Paper IV of the current thesis. A model for retrieving diameter distributions from laser scanner data was developed in Paper I. The model was adapted to uneven-aged forest structures and was performing well within the range of the model development data. Even if the diameter distribution of other forest structures were not as well predicted, estimated stand volumes derived from the predicted diameter distribution were quite accurate irrespective of forest structure.

Usually, diameter distributions are limited to a lower diameter limit of for example 3 cm. However, information of the magnitude of the trees smaller than this limit may be useful for predictions of future states because the initial state is described in more detail. Paper II deals with identifying variables derived from laser scanner data suitable for explaining the variation in quantity and vitality of young seedlings in a size divers spruce forest. The results indicated that the relationships were strong enough so that prediction models can be developed.

Volumes of single trees are usually determined by means of volume equations that depend on diameter and height. To be able to compute volume of single trees described in a diameter distribution, tree heights have to be modeled. Paper III presents non-linear models that predict the height from diameter and stand variables and enabled prediction of tree heights in the study presented in Paper IV. The models were developed from National Forest Inventory (NFI) data and are therefore valid for all parts of Norway.

Paper IV presents a matrix model for forest management. The matrix model was tested by performing both a short term and a long term validation. The short term validation comprised

tests that compared actual stand growth and model predictions on independent data. The results showed that the model were able to produce unbiased estimates of the number of stems. The long term model validation consisted in simulating stand growth without harvest for 1,000 years to see if the model predicted logical values with respect to steady state stand basal areas and species composition. The results indicated that the model produced estimates that reached steady states according to the hypothesis. The steady state was irrespective of the initial state of the stand being projected, as expected. The species composition was also logic producing large spruce volumes on the good sites and more pine on the poor sites.

# Norsk sammendrag

Denne avhandlingen presenterer en forvaltningsmodell for fleraldret skog i Norge. Rammen er en matrisemodell og denne består av flere delmodeller estimert på data fra landsskogtakseringen i Norge. Matrisemodellen predikerer utviklingen i skog på diameterklassenivå ved hjelp av overgangssannynligheter. I avhandlingen er det utviklet delmodeller for diameterfordelinger, rekruttering, diametertilvekst og mortalitet samt modeller for prediksjon av trehøyde.

Diameterfordelinger blir vanligvis ikke registrert i ordinære skogtakster. Slike fordelinger er nødvendige datagrunnlag for diameterklassevise modeller slik som den presentert i Artikkel IV i denne avhandlingen. En modell for estimering av diameterfordelinger basert på laserdata presenteres i artikkel I. Modellen ble kalibrert for uensaldret skog og resultatene viste at den fungerte godt på data tilsvarende estimeringsdatene. Selv om andre typer diameterfordelinger (andre skogstrukturer) ikke ble predikert like godt, var volum estimert på grunnlag av diameterfordelingen uavhengig skogstrukturen.

Diameterfordelinger begrenses vanligvis til en nedre diameter, for eksempel 3 cm. Informasjon om mengden trær som er mindre enn dette er likevel interessant fordi denne kan brukes til å forbedre prediksjoner av framtidige skogtilstander ved at utgangstilstanden er beskrevet i mer detalj. Artikkel II presenterer en studie der det ble fokusert på å finne laservariable som var egnet til å inngå i modeller som kan forklare variasjon i antall og kvalitet på småplanter i en heterogen granskog. Resultatene indikerer at sammenhengene er sterke nok til at prediksjonsmodeller kan estimeres.

Volum på enkelttrær beregnes vanligvis ved hjelp av funksjoner som er avhengige av diameter og høyde. For å kunne beregne volumet av trær i en diameterfordeling må vi derfor beregne høyden på trærne i en slik fordeling. Artikkel III presenterer ikke-lineære modeller som predikerer trehøyde ved hjelp av diameter og bestandsvariable og muliggjorde prediksjon av trehøyder i studien presentert i Artikkel IV. Modellene ble estimert på data fra landsskogtakseringen i Norge og har derfor hele landet som gyldighetsområde.

Artikkel IV presenterer en matrisemodell for skogforvaltning. Modellen ble testet på kort og lang sikt. På kort sikt ble prediksjoner sammenlignet med observerte verdier og resultatene

viste at modellen ga forventningsrette estimat. Den langsiktige valideringen bestod i å simulere utviklingen i skog uten hogst i en 1000 års perspektiv. Validiteten av denne testen ble vurdert ut fra om modellene ga en såkalt ”steady state”, altså om grunnflate og treslagsfordeling konvergerte mot en stabil tilstand.  Resultatene viste at modellene ga resultater i henhold til denne hypotesen. Den endelige tilstanden som modellen ga, var uavhengig av den initiale tilstanden. Resultatene var også logiske i forhold til at de ga store andeler gran på gode boniteter og mer furu på dårlige.

# List of papers

**Paper I**

Bollandsås, O.M. & Næsset, E. 2007. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data.
*Scandinavian Journal of Forest Research*, *22(1)*, 33-47. http://www.informaworld.com

**Paper II**

Bollandsås, O.M., Hanssen, K.H., Marthiniussen, S. & Næsset, E. In press. Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest. *Forest Ecology and Management.*

**Paper III**

Bollandsås, O.M., Næsset, E. Submitted. Height-diameter models for individual trees in Norway. *Silva Fennica.*

**Paper IV**

Bollandsås, O.M., Buongiorno, J. & Gobakken, T. Submitted. Predicting the growth of stands of trees of mixed species and size: A matrix model for Norway.
*Scandinavian Journal of Forest Research*. http://www.informaworld.com

# Introduction

## Background

In the early 20th century, forest management in Norway became so called even-aged. This means that the forest area was populated by compartments of trees of approximately the same size and mostly one species. Within such silvicultural practices the harvest operations are typically based on clear-fellings with a subsequent planting to regenerate the new compartment or stand, at least for stands of spruce which is the dominating tree species in Norway. Pine stands are often regenerated by means of seed-trees, regularly distributed across the area when the stands are harvested. Irrespective of species, the compartments are quite even with respect to both tree size and species under this silvicultural regime.

Sufficient information for calculating merchandisable volumes and for projecting future growth and yield of these kinds of stands can be given by means of a few variables because most trees are of approximately the same size. These variables are the stand basal area per hectare ($BA$), dominant height, Lorey's mean height, mean diameter, number of stems per hectare, and stand volume. Dominant height ($h_0$) is computed as the arithmetic mean height of the 100 largest trees per hectare according to diameter (Tveite, 1977). Lorey's mean height ($h_L$) is mean height weighted by basal area while mean diameter ($d_g$) is the diameter of the tree with mean basal area. Stand volume ($V$) is computed by means of stand volume equations that are functions of BA and $h_L$ (Brantseg, 1959; Vestjordet 1959; Næsset, 1995; Næsset and Tveite, 1999) or by multiplying the volume of an average tree as described by $h_L$ and $d_g$ by the total number of trees ($N$). Volume of each tree is computed by means of volume equations of individual trees (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967).

Because biodiversity is recognized as important in order to conduct a sustainable forestry, incentives that promote selective cuttings have been put forward. The most prominent of these incentives is the establishment of certification systems based on e.g. the Programme for the Endorsement of Forest Certification schemes (PEFC) or the Forest Stewardship Council (FSC). These are systems that give guidelines for a sustainable forestry. If the guidelines are complied with, the timber producers may be given an environmental label. Because of these certification systems and also public subsidy regulations that promote biological diversity, a larger portion of the forest stands are likely to be of trees of various species and size in the future.

### *Forest inventory and modeling approach*

For an even-aged forest structure the above mentioned stand characteristics are sufficient for volume calculations and projections of the forest and are therefore targeted in inventories of even-aged forest stands. Furthermore, the models used for predicting future growth and yield of such forest structures rely on stand characteristics as input variables. If the forest structure changes and becomes more size and species divers, the targeted descriptors and modeling approach need to be changed accordingly. Even-aged stands comprise trees with similar size and competitive status, whereas uneven-aged stands can have a wide size-range. The traditional stand characteristics may therefore be less useful because they only are average variables for the entire stand as such and do not recognize the growth conditions of the individual tree. Thus, in an uneven-aged stand we need higher resolution variables and models, either related to each tree or related to classes of size. Such variables and models enable prediction of growth for each tree or by size class accounting for differences in competitive status and computation of yield distributed on size classes.

Single-tree models utilize information on single-tree level. Typical variables may be the size of the tree as described by the diameter, basal area, or height; the competitive status of the tree described by a competition index, but also stand-variables to describe the general growth conditions in the stand. Such variables may be *BA*, *N*, or site quality described by a site index. Single tree models are adapted to account for different growth conditions between trees of different sizes in the same stand. However, because the models are single-tree based, they also require input data on the same level. This means that the data collection can be more laborious and that calculations and predictions can be more complex. The input data in such situations may be data comprising every single tree in the forest, but a more realistic data source is a diameter distribution related to an area unit under the assumption that the forest structure repeats itself on the scale equal to this area unit.

In many applications a sufficient resolution of the modeling is classes of size. In principle both the strictly area-based approach and the single tree approach are size class approaches (Vanclay, 1994), where the area-based has only one class and the single tree approach has number of size classes equal to the number of trees. In Norway where a tree of 50 cm of diameter is considered to be large, a plausible number of size classes will be between five and 25 and the size is most often related to diameter. Predictions of future growth and yield in the diameter class approach can be based on single tree models. The models are then applied to the average tree in each diameter class and the prediction of each class is multiplied to the number of trees of each class, assuming a uniform distribution of diameters within each individual class. Furthermore, by utilizing this approach it is also implicitly assumed that little information is lost by using size classes instead of tree level data. Thus, the stand states as described by diameter distributions is assumed to be sufficiently detailed to exploit the rich source of information in individual tree data to predict stand growth.

Species specific diameter distributions are sufficient input data for modeling using either an area-based approach or a diameter class approach. Diameter distributions can be obtained from field measurements or remote sensing (Maltamo et al., 2000; Gobakken and Næsset, 2004; Gobakken and Næsset, 2005; Maltamo et al., 2005; Bollandsås and Næsset, 2007), or with models that predict diameter distributions based on stand data (Vestjordet, 1972; Mønnes, 1982; Holte, 1993). The most prominent remote sensing technique the resent years has been airborne laser scanning.

### *Airborne laser data acquisition*

Airborne laser scanners retrieve 3D information of the vegetation by transmitting geo-referenced laser pulses. These pulses have certain footprints on the ground. The footprint is considered small it the footprint diameter is less that 1 m and large for footprints exceeding 5 m. The xyz coordinate of each pulse's hit on the ground or vegetation is recorded. By filtering out the hits considered to be ground hits, a terrain model can be established and relative height values for the vegetation hits computed. The accuracy of the height values of the terrain model is typically 20 to 30 cm (Kraus and Pfeifer, 1998; Reutebuch et al., 2003; Hodgson and Bresnahan, 2004).

There are two main types of airborne lasers being used. These are the discrete pulse (DP) systems and the continuous waveform (CW) laser systems carried by either a plane or heilcopter. The main difference between these two systems is that CW (Harding et al., 1994, 2001; Blair et al. 1999; Lefsky et al. 1999; Dubayah et al., 2000) registers a complete distribution of the energy being returned from where the pulse hit, whereas DP systems registers heights of a limited number of return echoes defined by major peaks in the return signal (Lefsky et al., 2002). Furthermore, there exist both scanning and profiling lasers. Scanners collect data along a "corridor" on the ground because the laser pulses are distributed out in a certain width. The width of this corridor is dependent on scan angle and flying altitude and is typically from, say, 200 to 1000 meters. Profiling lasers (Nelson et al. 2003, 2004) collect data from a cross-section of the vegetation having a width equal to the footprint of the laser pulse. The profilers have typically been used to cover large areas.

Laser scanners have been in operational use for forestry applications since year 2002 (Næsset et al., 2004; Næsset 2007). The most frequently used sensors in Norway are discrete pulse, scanning lasers with a footprint size of about 20 to 30 centimeters which are operated with scan angles and flying altitudes that yield from <1 to 5 pulses per square meter (Næsset 2002, 2004a, c; Bollandsås and Næsset, 2007). The number of pulses per square meter is also dependent on the pulse repetition rate which on the newest systems is >150 kHz and the speed of the aircraft. From the vegetation heights a number of different variables can be computed. These are variables reflecting the height distribution of the vegetation hits. One type of variables is heights of different percentiles according to maximum height. Furthermore, density variables can be computed by dividing the laser height range in uniform fractions and divide the cumulative number of pulses between the lower limit of each fraction and the maximum laser height, to the total number of pulses (Næsset 2004c).

Laser pulses transmitted from an airborne platform have a variable rate of penetration through the canopy depending on the canopy structure. In boreal forests, the penetration rate can typically be around 40 % (e.g. Næsset, 2004b). Modeling of structural characteristics uneven-aged forest stands (Maltamo et al., 2005) and diameter distributions for uneven-aged forest structures (Bollandsås and Næsset, 2007 (Paper I)) have therefore been successfully carried out in the past. Moreover, laser data (DP and CW) has also been utilized to efficiently derive stand-based estimates of biophysical properties like tree heights, mean diameter, stem number, basal area, and timber volume (e.g. Maclean and Krabill, 1986; Magnussen and Boudewyn, 1998; Nilsson, 1996; Næsset, 1997a, b, 2002, 2004c; Means et al, 2000), leaf area index (Lefsky et al., 1999, Riaño et al., 2004; Solberg et al., 2006) and even forest growth (St-Onge and Vepakomma, 2004; Yu et al., 2004; Næsset and Gobakken, 2005). Laser data has also potential with regard to retrieving structural information of sub-canopy layers (e.g. saplings), either by utilizing variables computed from the laser pulses returned from these layers, or by exploiting relationships between the canopy and the structure of the sub-canopy layers (Bollandsås et al., in press (Paper II)). In the current thesis, laser data is used to model both diameter distributions and to explore how laser data can be utilized to model the quantity and vitality of saplings in an uneven-aged spruce forest.

### *Volume computation*

Computations of timber volumes by diameter classes can be carried out by calculating volumes of the average sized trees in class by means of single tree volume equations, multiplied by the number of trees in each class. Total volume is found by summing over diameter classes. Tree height is a requirement for using the Norwegian single tree volume equations (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967). Tree heights can either be retrieved by field measurements, utilizing models, or by a combination where heights are measured on a sub-sample of trees that enables development of local height-diameter relationships. Heights can also be predicted by height-diameter models fitted for larger regions so that height measurements can be omitted and hence reduce inventory costs.

### *Matrix modeling*

An efficient way to develop a diameter class management model applicable to an uneven-aged forest structure is to express it as a matrix model that project future outcomes by means of transition probabilities. The transition probabilities are related to specific diameter classes, and are probabilities of a tree in a specific class either stays in the same class within the projection period, moves to the next diameter class or dies. These are the only possible outcomes, so these probabilities add up to one. The harvest is exogenously determined. This is an efficient framework because the model operates on a limited number of size classes and because the projections are performed by means of matrix calculations.

Forest matrix models have been developed by several authors (e.g. Usher, 1966; Bosch, 1971; Buongiorno and Michie, 1980; Lu and Buongiorno, 1993; Liang et al., 2005). Such models express the stand state as a vector of the number of trees per unit area in predefined species and diameter classes. The model can be expressed by the following general model form:

$$\boldsymbol{y}_{\mathrm{t+p}} = \boldsymbol{G}\,\left(\boldsymbol{y}_t - \boldsymbol{h}_t\right) + \boldsymbol{R} \tag{1}$$

where $\boldsymbol{y}_t = [\boldsymbol{y}_{ijt}]$ is the vector state, in which each entry is the number of live trees per unit area of species group ($i$ =1,…,$m$) and diameter class ($j$ =1,…,$n$) at time $t$, $\boldsymbol{G}$ is a transition probability matrix, $\boldsymbol{h}_t = [\boldsymbol{h}_{ijt}]$, is the harvest vector in which each entry is the number of trees cut per unit area, and $p$ is the projection interval. $\boldsymbol{R}$ is the recruitment vector in which each entry is the number of recruits per unit area by species and diameter class. The transition probability matrix $\boldsymbol{G}$ can have one sub-matrix for each species group with the following structure:

$$\boldsymbol{G}_i = \begin{bmatrix} \boldsymbol{a}_{i1} & 0 & \cdots & 0 & 0 \\ \boldsymbol{b}_{i2} & \boldsymbol{a}_{i2} & \cdots & 0 & 0 \\ 0 & \boldsymbol{b}_{i3} & \ddots & 0 & 0 \\ \vdots & \cdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{b}_{in} & \boldsymbol{a}_{in} \end{bmatrix} \qquad (2)$$

where $\boldsymbol{a}_{ij}$ ($i = 1, \ldots, m, j = 1, \ldots, n$) is the probability that a tree of species $i$ will remain in diameter class $j$ between $t$ and $t + p$, $\boldsymbol{b}_{ij}$ is the probability that a tree in diameter class $j$ grows into diameter class $j$+1 during the time interval $p$.

Similarly, the recruitment vector can consist of one sub-vector for each species group, indicating the number of recruits of a specific species $i$ that enter the smallest diameter class from $t$ to $t + p$:

$$\boldsymbol{R}_i = \begin{bmatrix} \boldsymbol{d}_i \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (3)$$

where $d_i$ is the number of trees of species $i$ that enter the smallest diameter class between $t$ and $t + p$.

The transition probability $\boldsymbol{b}_{ij}$ can be calculated as the ratio of the rate of diameter growth to the width of a diameter class. It can also be modeled directly, but this limits the matrix model to only operate with a certain width of the diameter classes. The probability that a tree stays in a diameter class, $\boldsymbol{a}_{ij}$, is calculated as $\boldsymbol{a}_{it}= 1 - \boldsymbol{b}_{ij}- \boldsymbol{m}_{ij}$, where $\boldsymbol{m}_{ij}$ is the probability that a tree of species group $i$ and diameter class $j$ dies during the interval $t$ to $t+p$. Like the transition probability, the mortality rate is a function of the diameter class of the tree.

The recruitment rate, $\boldsymbol{d}_i$ can be obtained by means of a single tree recruit model. Recruitment models are models that predict the number of individual trees that exceed a certain threshold-size during a specific period. The recruitment can be modeled in two different ways, either by one single equation (e.g. Moser, 1972; Adams and Ek, 1974; Vanclay, 1989) or by a two-stage approach (e.g. Vanclay, 1992; Lexerød, 2005). A two-stage approach first models the probability for recruitment to occur and then conditionally models the number of recruits given the result of the first stage.

# Objectives

The current thesis aims at developing management models adapted to an uneven-aged forest structure and to test methods for efficient collection of inventory data required by such a model framework. The models should be applicable to be implemented in a matrix model for uneven-aged forest management.

# Data

Paper I and II are based on data collected in a nature reserve in the southeastern part of Norway. No silvicultural intervention has been done on this area for at least 75 years (Økland, 1994) and was considered to be representative of uneven-aged forest.

Paper III and IV utilize data from the Norwegian National Forest Inventory (NFI). Data between 1994 and 2006 have been used for either estimation of model parameters or for validation of models. Please refer to each paper for detailed information of each application.

# Major findings

## *Paper I. Diameter distributions*

Paper I deals with the modeling of diameter distributions of the kind that are needed as input data for the matrix model developed in Paper IV. In this specific application, the diameter distribution from a lower threshold of 3 cm was represented by 10 percentiles (Borders et al., 1987) of the basal area distribution (Kangas & Maltamo, 2000). The diameter distribution was obtained by scaling the percentiles with an estimate of stand basal area. Modeling of these percentiles was based on airborne laser scanner data and it successfully reproduced diameter distributions of the same type as the model calibration data (independent validation on uneven-aged forest data). However, the model was not flexible enough to also reproduce normal and uniform diameter distributions. Volume estimates derived from predicted diameter distributions were generally well determined irrespective of the observed distribution in the independent validation. The conclusions of the study were that the model could be used for retrieval of diameter distributions form uneven-aged forest, but since the models was sensitive to the distribution type a stratification with respect to forest structure is needed. This stratification can either be done from existing data or from the laser data itself by means of a classification rule or a model that predict objective measures of stand structure.

### *Paper II. Regeneration*

Paper I limits the diameter distribution in the lower end to 3 cm. However, in Paper II we explored the possibilities of predicting the number and vitality of even smaller trees. Paper II deals with saplings that were between 0.1 and 3 meters of height in the same area as the diameter distribution trial was conducted. The modeling was based on variables derived from airborne laser scanner data. The study aimed at identifying categories of candidate variables for future regeneration prediction models for operational forest management. The conclusions of the study were that the relationship between the laser-depicted canopy and the quantity of small saplings as described by a regeneration success rate (RSR) was so strong that this kind of data can be utilized to develop predictive models. The laser variables that gave best fit a model of the RSR were computed from data originating from the last return echoes. The fit was also improved by using laser data on larger geographical scales than the field plots where the ground truth data was collected. Furthermore, variables affected by variation of the lower fractions of the canopy were better than those only affected by variation in the upper parts. The results also indicated that the density variables were superior to the height variables.

The correlations between laser variables and the variation of sapling vitality as described by apical dominance ratio, leader length, and leader length relative to sapling height, were not as strong as for the RSR. The conclusion was that the correlations would have been stronger if our data set included more variation of forest structure.

Models that can predict the regeneration success may be useful in uneven-aged forestry because information of the whole distribution of size will increase the precision of forecasts due to a better description of the initial stand state. However, for this to be true, models applicable to forecast these small trees have to be used. Furthermore, information of the small sized trees can also be useful when planning harvest operations in uneven-aged stands. Harvest can by using this information be allocated to stands where the regeneration success rate estimate is high.

### *Paper III. Height-diameter models*

To be able to compute volume by means of the volume equations available in Norway (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967), diameter (Paper I) and tree heights are needed as input. Paper III developed non-linear diameter-height models. These are models that predict the heights of single trees by utilizing diameter in breast height together with stand variables as explanatory variables. The stand variables were included to reflect that the height-diameter relationship of a tree is affected with multiple factors. Among others, these factors are site index (Vanclay, 1994), local climate (Mäkinen, 1998), competition (Loetsch et al., 1973), and stand density (Fulton, 1999; Sharma and Zhang 2004).

Two set of models were developed. Model set 1 included stand density and Model set 2 did not. The two sets of models were developed because stand density is easily changed by silvicultural operations. Thus, height estimates of residual trees will be different just before a thinning compared with immediately after the thinning. The model set that included stand density (Model set 1) is therefore most applicable in forest management regimes where stand density is kept relatively constant. Diameter, stand density, site productivity, altitude, and latitude (Model set 1) explained 66 % to 82 % of the variation in tree height for the different tree species included in the study. No serious bias was detected. The random errors assessed by independent tests were 17.1 %, 19.0 %, 19.7 %, and 23.2 % of the observed mean tree height for Norway spruce, Scots pine, birch, and other broadleaves, respectively. The corresponding range of model fit for Model set 2 were 63 % to 79 % and the random errors were 18.1 %, 20.0 %, 20.3%, and 23.8 %.

### *Paper IV. Matrix model*

Paper IV can be regarded as the main paper of this thesis. This paper presents that matrix model as described in the introduction. The paper presents different sub-models of recruitment, diameter growth, and mortality. These sub-models determined the transition probabilities of the matrix model. The probabilities of the matrix-model were dependent on the stand state and accounted for differences in competitive status of each diameter class. The model was tested by comparing actual stand growth and model predictions on 416 randomly selected, independent, plots. The matrix model was applied to the plot data at the time of the first measurement, and two 5-year iterations of the model were performed. Then, the predicted number of trees and basal area by species and diameter class after 10 years was compared to the observed. The results showed that the model were able to produce unbiased estimates of the number of stems of all diameter classes for all species with only one exception. Furthermore, a long term test was also performed. The long term model validation consisted in simulating stand growth without harvest for 1,000 years. The hypothesis was that the predicted stand would reach a steady state corresponding to the ecological climax (Buongiorno et al. 1995). Furthermore, without climate change or other major disturbance that give basis for another level of the steady state, the predicted steady state should be independent of the initial stand state (Stenberg and Siriwardana, 2006). The results of these tests showed that the model produced estimates that reached steady states according to the hypothesis. The steady state was irrespective of the initial state of the stand being projected, as expected. The steady state estimates were also logical according to variation in site index, producing higher volumes with increasing site index. The species composition was also logical producing large spruce volumes on the good sites and more pine on the poor sites.

## Discussion and further work

Paper I showed that the diameter distribution of an uneven-aged spruce forest could be retrieved by means of data from airborne laser scanning. The paper only deals with single species distributions. Future studies should be focused on discriminating species and reproducing species specific distributions. Studies that have focused on tree species classification on single tree level previously are presented by, e.g. Brandtberg et al., 2003; Holmgren and Persson, 2004; Moffiet et al., 2005; Brandtberg, 2007; Ørka et al., 2007. To aid the classification of species with regard to stand variables, different remote sensing techniques could be used in combination with laser scanner data, for example aerial photography or spectral data. For example, Packalen and Maltamo (2006) predicted species specific volumes by combining information form airborne laser scanner data and aerial photographs. The same sources of information were used by Packalen and Maltamo (2007) to predict species specific volume, stem number, basal area, basal area median diameter, and tree height.

Several authors have observed that utilization of the cumulative basal area distributions (Paper I included) scaled by basal area to compute the diameter density distribution, yield a decreasing distribution. Kangas et al. (2007) approached this problem by including new percentiles in the lower end of the distribution, and the behaviour was improved. However, as the percentiles at the lower end get closer, the danger of predicting non-monotonistic diameters increases. They also tried to loosen the strict relationship between number of stems and the basal area of a certain diameter class by estimating a parameter for each diameter class to derive the number of stems from basal area. If such a parameter diverges from the theoretical value, the assumption of linear interpolation between percentiles to derive the number of stems does not hold.

Paper II indicates that laser data potentially may be a good source of information about small saplings. This paper deals with saplings that are smaller than the 3 centimeter diameter threshold used in Paper I. Future work should focus on developing methods for implementing this information in ordinary inventories of uneven-aged forests. The best way of acquiring information about the quantity of saplings is to establish local relationships between the number of saplings and laser data by means of field plots and corresponding laser data. However, if some stratification is made on the basis of forest type, utilization of models valid for larger regions could be possible. The NFI of Norway has the last three years collected data on trees taller than 0.1 m. This data source could be used for developing models for regeneration success if laser data is recorded for these plots. Laser data and precise coordinates of the NFI plots already exist for the county of Hedmark, so development of regeneration success models for this region is already possible.

The individual models developed in Paper III and IV utilize the NFI data of Norway, both at stand and individual tree level. The data encompasses a wide range of growth conditions. However, the data originate from a relatively short period of time and the models are dependent on the growth conditions during this specific period. The cross-sectional variability, however, compensates to some extent for the short time interval between observations, allowing the accurate estimation of parameters without the need of long-term experiments that are necessarily costly and limited in scope. Furthermore, if the data from the period used for model estimation suffers from a systematic climatic offset, the models need to be updated as new data is collected.

In applications of the matrix model, accurate predictions are not to be expected for each and every stand. However, on average for a large number of stands the results indicate that predictions will be unbiased. This is general also for each sub-model which is developed from NFI data collected over a large geographical region. However, given similar conditions, the recruitment, mortality, and growth can vary considerably from stand to stand and tree to tree. Little can be done to reduce this uncertainty, although it can be taken into account in decision making (Lin and Buongiorno, 1998).

In Paper IV, the model was applied only to simulate natural stand growth without human disturbance, to check its validity as a predictor of expected stand growth. The same deterministic model form, or a stochastic extension (e.g. Liang et al. 2006), should also prove useful to predict the economic and ecological effects of continuous-cover/uneven-aged forest management, by simulation, or optimization.

The studies conducted within this thesis have shown that unbiased estimates of forest development of uneven-aged stands can be produced using a matrix models operating on diameter classes. Furthermore, it was also shown that sufficient input data for these models could be retrieved from airborne laser scanning. Moreover, from airborne laser scanner data, estimates of the quantity and vitality of small saplings that can aid traditional recruitment models was retrieved. Thus, the representation of uneven-aged forest structures by number of stems in diameter classes, seem to work well for the tests performed in the current thesis and inventory methods based on airborne laser scanning are considered efficient. Furthermore, for forestry-applications the laser technology is still "young", and there exist a large potential for refinement of the existing range of applications as well as for extended utilization of laser data.

## Acknowledgements

I would like to thank Professor Erik Næsset and Associate Professor Terje Gobakken for useful comments on the synopsis of the current thesis.

## References

Adams, D.M. & Ek, A.R. 1974. Optimizing the management of uneven-aged forest stands. *Canadian Journal of Forest Research 4*, 274-287.

Blair, J.B., Rabine, D.L. & Hofton, M.A. 1999. The laser vegetation imaging sensor (LVIS): a medium-altitude digitization-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing 54*, 115-122.

Bollandsås, O.M., Hanssen, K.H., Marthiniussen, S. & Næsset, E. IN PRESS. Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest. *Forest Ecology and Management.*

Bollandsås, O.M. & Næsset, E. 2007. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. *Scandinavian Journal of Forest Research, 22*(1), 33-47.

Borders, B.E., Souter, R.A., Bailey, R.L. & Ware, K. D. 1987. Percentile-based distributions characterize forest stand tables. *Forest Science, 33*, 570-576.

Bosch, C.A. 1971. Redwoods: a population model. *Science 172*, 345-349.

Braastad, H. 1966. *Volume tables for birch*. Communications of Norwegian Forest Research Institute, 21, 265-365. (In Norwegian with English summary).

Brandtberg, T. 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing, 61*, 325-340

Brandtberg, T., Warner, T.A., Landenberger, R.E. & McGraw, J.B. 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment, 85*, 290-303

Brantseg, A. 1959. *Stand volume functions for pine*. Norsk skoghåndbok. (In Norwegian).

Brantseg, A. 1967. *Volume functions and tables for Scots pine. South Norway*. Communications of Norwegian Forest Research Institute, 22, 695-739. (In Norwegian with English summary).

Buongiorno, J. & Michie, B.R. 1980. A matrix model of uneven-aged forest management. *Forest Science, 26*(4), 609-625.

Buongiorno, J., Peyron, J.L., Houllier, F. & Bruciamacchie, M. 1995. Growth and management of mixed-species, uneven-aged forests in the French Jura: Implications for the economic returns and tree diversity. *Forest Science, 41*(3), 397-429.

Dubayah, R., Knox, R., Hofton, M., Blair, J.B. & Drake, J. 2000. Land surface characterization using lidar remote sensing. *In*: Hill, M. J., Aspinall, R. J. (eds). Spatial information for land use management, pp 25-38. Singapore, International Publishers Direct.

Fulton, M.R. 1999. Patterns in height-diameter relationships for selected tree species and sites in eastern Texas. *Canadian Journal of Forest Research 29*, 1445-1449.

Gobakken, T. & Næsset, E. 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research 19*, 529-542.

Gobakken, T. & Næsset, E. 2005. Weibull and percentile models for LIDAR-based estimation of basal area distribution. *Scandinavian Journal of Forest Research 20*, 490 – 502.

Harding, D. J., Blair, J.B., Garvin, J.B. & Lawrence, W.T. 1994. Laser altimetry waveform measurement of vegetation canopy structure. Proceedings of the International Remote Sensing Symposium 1994, pp 1251-1253. Pasadena (CA): California Institute of Technology.

Hodgson, M. & Bresnahan, P. 2004. Accuracy of airborne lidarderived elevation: Empirical assessment and error budget. *Photogrammetric Engineering and Remote Sensing 70*, 331-339.

Holmgren, J. & Persson, Å. 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment, 90*, 415-423

Holte, A.1993. *Diameter distribution functions for even-aged (Picea abies) stands*. Communications of Norwegian Forest Research Institute 46(1), 1-46.

Kangas, A. & Maltamo, M. 2000. Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. *Silva Fennica, 34*, 371-380.

Kangas, A., Mehtätalo, L. & Maltamo, M. 2007. Modelling percentile based basal area weighted diameter distribution. *Silva Fennica 41*, 425-440.

Kraus, K. & Pfeifer, N. 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing 53*,193-203.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A. & Harding, D. 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment 70*, 339-361.

Lefsky, M.A., Cohen, W.B., Parker, G.G. & Harding, D.J. 2002. Lidar remote sensing for ecosystem studies. *BioScience 52*, 19-30.

Lexerød, N. 2005. Recruitment models for different tree species in Norway. *Forest Ecology and Management 206*, 91-108.

Liang, J., Buongiorno, J. & Monserud, R.A. 2005. Growth and yield of all-aged Douglas fir – western hemlock forest stands: a matrix model with stand diversity effects. *Canadian Journal of Forest Research 35*, 2368-2381.

Liang, J., Buongiorno, J. & Monserud, R.A. 2006. Bootstrap Simulation and Response Surface Optimization of Management Regimes for Douglas-Fir/Western Hemlock Stands. *Forest Science 52*, 579-594.

Lin, C.R. & J. Buongiorno. 1998. Tree diversity, landscape diversity, and economics of maple-birch forests: Implications of a Markovian models. *Management Science 44*(10), 1351-1366.

Loetsch, F., Zöhrer, F. & Haller, K.E. 1973. *Forest inventory*, Volume II. BLV, München, Bern, Wien.

Lu, H-C. & Buongiorno, J. 1993. Long- and short-term effects of alternative cutting regimes on economic returns and ecological diversity in mixed-species forests. *Forest Ecology and Management 58*(3/4), 173-192.

Maclean, G.A. & Krabill, W.B. 1986. Gross-merchantable timber volume estimation using an airborne lidar system. *Canadian Journal of Remote Sensing 12*, 7-18.

Magnussen, S. & Boudewyn, P. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research 28*, 1016-1031.

Mäkinen, H. 1998. The suitability of height and radial increment variation in *Pinus Sylvestris* (L.) for expressing environmental signals. *Forest Ecology and Management 112*, 191-198.

Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T. & Saramäki, J. 2000. Comparison of percentile based prediction methods and Weibull distribution in describing diameter distribution of heterogeneous Scots pine stands. *Forest Ecology and Management, 133*, 263-274.

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J. & Pitkänen, J. 2005. Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management 216*(1/3), 41-50.

Means, J.E., Acker, S.A., Brandon, J.F., Renslow, M., Emerson, L. & Hendrix, C.J. 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engeneering and Remote Sensing 66*, 1367-1371.

Moffiet, T., Mengersen, K., Witte, C., King, R., & Denham, R. 2005. Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing, 59*, 289-309.

Mønness, E. 1982. *Diameter distributions and height curves in even-aged stands of Pinus sylvestris L.* Communications of Norwegian Forest Research Institute 36 (15): 1-43.

Moser, J.W. 1972. Dynamics of uneven-aged forest stands. *Forest Science 18*, 184-191.

Næsset E. & Gobakken, T. 2005. Estimating forest growth using canopy metrics derived from airborne laser scanner data. *Remote Sensing of Environment 96*, 453-465.

Næsset E. & Tveite, B. 1999. Stand volume functions for Picea abies in Eastern, Central and Northern Norway. *Scandinavian Journal of Forest Research 14*, 164-174.

Næsset, E. 1995. Stand volume functions for Picea abiesin Western Norway. *Scandinavian Journal of Forest Research 10*, 42-50.

Næsset, E. 1997a. Determination of means tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing 52*, 49-56.

Næsset, E. 1997b. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment 61*, 246-253.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment 80*, 88-99.

Næsset, E. 2004a. Accuracy of forest inventory using airborne laser-scanning: evaluating the first Nordic full-scale operational project. *Scandinavian Journal of Forest Research 19*, 554-557.

Næsset, E. 2004b. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment 91*, 243-255.

Næsset, E. 2004c. Practical large-scale forest stand inventory using small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research 19*, 164-179.

Næsset, E. 2007. Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research 22*, 433-442.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å. & Söderman, U. 2004. Laser scanning of forest resources: The Nordic experience. *Scandinavian Journal of Forest Research 19*, 482-499.

Nelson, R., Short, A. & Valenti, M. 2004. Measuring biomass and carbon in Delaware using an airborne profiling lidar. *Scandinavian Journal of Forest Research 19*, 500- 511.

Nelson, R., Valenti, M.A., Short, A. & Keller, C. 2003. A multiple resource inventory of Delaware using airborne laser data. *BioScience 53*, 981-992.

Nilsson, M. 1996. Estimating tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment 22*, 1-7.

Økland, B. 1994. Mycetophilidae (Diptera), an insect group vulnerable to forest practices? A comparison of clearcut, managed and semi-natural spruce forests in southeastern Norway. *Biodiversity and Conservation 3*, 68-85.

Ørka, H.O., Næsset, E. & Bollandsås, O.M. 2007. Using airborne laser intensity for tree species classification. IAPRS Volume XXXVI, Part 3/W52, 2007. 300-304.

Packalen, P. & Maltamo, M. 2006. Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science 52*, 611-622.

Packalen, P. & Maltamo, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment 109,* 328-341.

Reutebuch, S. E., McGaughey, R. L., Andersen, H.-E. & Carson, W.W. 2003. Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing 29*, 527-535.

Riaño, D., Valladares, F., Condés, S. & Chuvieco, E. 2004. Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology 124*, 269–275.

Sharma, M. & Zhang, S.Y. 2004. Height-Diameter Models Using Stand Characteristics for *Pinus banksiana* and *Picea mariana*. *Scandinavian Journal of Forest Research 19*, 442-452.

Solberg, S., Næsset, E., Hanssen, K.H. & Christiansen, E. 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment 102*, 364-376

Stenberg, L.C. & Siriwardana, M. 2006. The steady-state treatment of forestry in CGE models. *International Journal of Agricultural Resources, Governance, and Ecology 5*, 1-17.

St-Onge, B. & Vepakomma, U. 2004. Assessing forest gap dynamics and growth using multi-temporal laser-scanner data. In: Thies, M., Koch, B., Spiecker, H. & Weinacker, H. (Eds):

Laser-Scanners for Forest and Landscape Assessment. Proceedings of the ISPRS working group VIII/2. *International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences, Vol XXXVI, Part 8/W2.* (pp. 173-178). Germany: University of Freiburg.

Tveite, B. 1977. *Site index curves for Norway spruce* (*Picea abies* (L.) Karst.). Report of Norwegian Forest Research Institute, 33, 1-84. (In Norwegian with English summary).

Usher, M.B. 1966. A matrix approach to the management of renewable resources, with special application to selection forests. *The Journal of Applied Ecology 3*, 355-367.

Vanclay, J.K. 1989. A growth model for north Queensland rainforests. *Forest Ecology and Management 98*, 1-12.

Vanclay, J.K. 1992. Modelling regeneration and recruitment in tropical rainforest. *Canadian Journal of Forest Research 22*, 1235-1248.

Vanclay, J.K. 1994. *Modelling forest growth and yield: Applications to mixed tropical forests.* Wallingford: CABI.

Vestjordet, E. 1959. *Stand volume functions for spruce and birch.* Norsk skoghåndbok. (In Norwegian).

Vestjordet, E. 1967. *Functions and tables for volume of standing trees. Norway spruce.* Communications of Norwegian Forest Research Institute 22, 543-574. (In Norwegian with English summary).

Vestjordet, E. 1972. *Diameter distribution and height curves for even-aged stands of Norway spruce.* Communications of Norwegian Forest Research Institute 29(8), 469-557.

Yu, X., Hyyppä, J., Kaartinen, H. & Malatamo, M. 2004. Automatic detection of harvested trees and detemination of forest growth using airborne laser scanning. *Remote Sensing of Environment 90*, 451-462.

# Paper I

*Scandinavian Journal of Forest Research*, 2007; 22: 33–47

Taylor & Francis
Taylor & Francis Group

**ORIGINAL ARTICLE**

# Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data

OLE MARTIN BOLLANDSÅS & ERIK NÆSSET

*Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway*

**Abstract**
A model for prediction of stand basal area and diameters at 10 percentiles of a basal area distribution was estimated from small-footprint laser scanner data from primeval conifer forest using partial least squares regression. The regression explained 44–80% and 67% of the variability of the 10 percentiles and stand basal area, respectively. The predicted percentiles, scaled by the predicted stand basal area, were used to compute diameter distributions. A cross-validation showed that the mean differences between the predicted and observed number of stems by diameter class were non-significant ($p > 0.05$) for 22 of 29 diameter classes. Moreover, plot volume was calculated from the predicted diameter distribution and cross-validation revealed a non-significant deviation between predicted and observed volume of −3.3% (of observed volume). An independent validation showed non-significant mean differences for 20 of 21 diameter classes for data corresponding to the model calibration data. Plot volumes calculated from the predicted diameter distributions deviated from observed volume by −4.4%. The model reproduced diameter distributions corresponding to the model calibration data (uneven-sized forest) well. However, the model is not flexible enough to reproduce normal and uniform diameter distributions. Volume estimates derived from predicted diameter distributions were generally well determined, irrespective of the observed distribution.

**Keywords:** *Canopy height, diameter distribution, laser scanning, PLSR, uneven-sized forest.*

## Introduction

Diameter distributions are seldom registered in conventional stand-based forest inventories. Still, the diameter distribution holds significant information about the stand properties and provides input in growth projections and other computations. The diameter distribution is particularly interesting in uneven-sized forest. Traditional stand variables characterizing mean properties of the stand (i.e. mean height, mean diameter and basal area) give a good description of even-sized forests, but in uneven-sized forest some additional information is needed describing the variability in size. Furthermore, optimal timing of harvests and silviculture depend strongly on the diameter distribution. Thus, an observed or predicted diameter distribution is necessary to optimize management strategies. Moreover, uneven-sized stands and forests contribute positively to the preservation of biodiversity (Brokaw & Lent, 1999). On a small spatial scale, the distribution of biomass between different tree sizes may be a key variable (Uuttera & Maltamo, 1995). The diameter distributions may therefore serve as a valuable assessment of the value of a forest stand or area in relation to biodiversity.

Different distribution functions have been proposed to model empirical diameter distributions, where the Weibull is the most frequently used (e.g. Bailey & Dell, 1973; Kangas & Maltamo, 2000*b*). Regardless of the utilized model is used, the distribution parameters must be predicted from some stand characteristics. Methods that use diameters at percentiles of number of stems or basal area to describe diameter distributions have also been proposed, henceforth referred to as the percentile method. Borders et al. (1987) characterized an empirical distribution function with 12 percentiles, according to number of stems. The diameter distributions were obtained by linear interpolation

---

Correspondence: O. M. Bollandsås, Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway. E-mail: ole.martin.bollandsas@umb.no

*(Received 23 December 2005; accepted 21 November 2006)*

ISSN 0282-7581 print/ISSN 1651-1891 online
DOI: 10.1080/02827580601138264

between predicted diameters. A test on 1009 empirical diameter distributions showed that 96% of them were reproduced correctly according to the Kolmogorov-Smirnov two-sample test. They concluded that this was a flexible method to model diameter distributions compared with relying on predefined distribution functions. Others (e.g. Kangas & Maltamo, 2000*a*) have predicted diameters at percentiles of basal area. Kangas and Maltamo's models performed quite well, and the accuracy was much improved by including number of stems per hectare as an explanatory variable. Also multimodal distributions can be reproduced applying the percentile method (Borders et al., 1987). In general, predefined distribution functions do not have this flexibility (Maltamo et al., 2000). Borders et al. (1987) compared their results (96% correctly reproduced distributions) with Bailey (1974), which reproduced 65% of the observed distributions correctly according to the Kolmogorov–Smirnov two-sample test using the parameter recovery method for the two-parameter Weibull distribution. Kangas and Maltamo (2000*b*) compared the percentile method with the parameter prediction of the Weibull distribution on four different data sets and found that the percentile method using number of stems as an explanatory variable was superior to the other tested models. Based on these findings, the percentile method may be most applicable when applied to data originating from multicanopy layered forest.

Field-based inventories of entire forest stands are expensive. However, data from airborne laser scanning have in recent years shown to be efficient in deriving stand-based estimates of tree heights, mean diameter, stem number, basal area and timber volume. Some field data are required to calibrate a model, but the magnitude is much less than for traditional inventories. Laser scanning depicts the canopy surface by measuring, say, from one to 10 vegetation heights per square metre (laser pulse density). This yields a three-dimensional representation of the canopy that can be used to model stand characteristics, exploiting the relationship between the biophysical properties of a stand and canopy structure. Næsset (2002, 2004b) concluded that all the stand parameters mentioned above can be estimated with higher precision by means of laser scanner data than by conventional field-based methods. However, Næsset (2004*b*) found that a high proportion of deciduous tree species mixed with conifer species degraded the accuracy, partly because the crown shapes of deciduous trees are different from and more variable than for conifers. In the current trial, however, where the dominating tree species have regular conical crown shapes, it is likely that there exists a strong relationship between diameter distribution and canopy surface depicted by laser scanning. Maltamo et al. (2004) proposed a method to estimate height distributions from laser data, based on single-tree segmentation. However, regardless of laser pulse density, some trees in the lower canopy storeys will be shaded by the overstorey and hence missed. Maltamo et al. (2004) approached this problem by identifying the visible trees, estimating their heights using the laser data and continuing the height distribution from the smallest segmented height class by a left-truncated Weibull distribution. Compared with just computing timber volume from the laser-observable trees, estimation of an additional height distribution of small trees improved the accuracy considerably.

Segmentation of single trees requires high pulse densities, say more than five pulses per square metre. Alternatively, variables representing an area can be derived from laser data. Such variables can typically be percentiles of the laser measured height distribution over a given area. Gobakken & Næsset (2005) used this area-based approach and found that diameter distributions using both a predefined distribution function and the percentile method could be estimated from metrics derived by laser scanning in a managed even-aged forest. Gobakken & Næsset (2005) used laser data with a density of approximately one pulse per square metre. The aims of this study were to test the accuracy of the percentile method for estimation of diameter distribution using area-based metrics derived from laser scanner data for uneven-sized conifer stands. The estimated models were based on diameters at 10 percentiles of the basal area distribution. The accuracy assessment was carried out by cross-validation and prediction on an independent test data set.

## Materials and methods

### *Model calibration data: inventory A*

The model calibration data originate from 20 circular sample plots of 0.1 ha in a boreal nature reserve in south-eastern Norway (59°50′ N, 11°02′ E, 190–370 m a.s.l., 1400 ha). Large within-stand variation of ages and sizes of trees characterize the forest, and Norway spruce [*Picea abies* (L.) Karst.] and Scots pine (*Pinus sylvestris* L.) are the dominating tree species.

Sample plots in inventory A were subjectively selected. They were spruce dominated and had multicanopy layers. The model calibration data consequently comprised multimodal and reverse-J diameter distributions. Furthermore, each plot was situated on sites with gentle terrain slopes because precise determination of plot perimeters of 0.1 ha circular plots on steep slopes is labour intensive. Gentle slopes were also selected because steep slopes may degrade the precision of the relationship between laser data and the field variables (Næsset, 2004*b*). In the current trial, the terrain slope ranged between 5% and 35%.

Inventory A was carried out during summer 2003. On each sample plot, diameter at breast height ($d_{\mathrm{bh}}$) was callipered for all trees with $d_{\mathrm{bh}} \geq 3$ cm. Tree heights were measured on subsampled trees selected with probability proportional to stem basal area (relascope factor 2) by means of a Vertex III hypsometer. The number of sample trees on each plot ranged from 23 to 42 (not counting plot 1, where all tree heights were measured). A summary of the ground reference data is shown in Table I.

Mean diameter ($d_{\mathrm{BA}}$) was defined as the diameter corresponding to mean stem basal area for trees with $d_{\mathrm{bh}} \geq 3$ cm. Mean height ($h_{\mathrm{L}}$) was defined as the mean height according to basal area. Dominant height ($h_{\mathrm{dom}}$) was defined as the arithmetic mean height of the 100 largest trees per hectare according to diameter (Tveite, 1977). The plot (0.1 ha) $h_{\mathrm{dom}}$ is consequently estimated as the arithmetic mean of the 10 trees with largest diameter within each plot. Stem number ($N$) was defined as number of trees ($d_{\mathrm{bh}} \geq 3$ cm) per hectare. Plot basal area (BA) was computed as basal area per hectare from the stem breast height diameter measurements. As a measure of stand structure, the Gini coefficient (Gini, 1912; Weiner & Solbrig, 1984) was computed for each plot. The Gini coefficient (GC) is an objective measure of the size distribution in a stand and is given by eq. (1):

$$\mathrm{GC} = \frac{\sum_{j=1}^{N}(2j - N - 1)ba_j}{\sum_{j=1}^{N} ba_j(N - 1)} \tag{1}$$

where $j$ is the rank of a tree according to diameter, $ba_j$ is the basal area of a tree with rank $j$, and $N$ denotes the total number of trees. GC is 0 when there is no variation in tree size and approaches 1 as the variation in tree size increases. Stem volume of each tree was estimated using volume models for individual trees (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967) with $d_{\mathrm{bh}}$ and tree height as input variables. Total volume ($V$) was estimated as the sum of individual tree volumes. Missing heights were predicted using multiplicative diameter–height models (Table II) estimated from the subsampled height-measured trees. Models for spruce were estimated separately for each plot. Owing to a low number of deciduous sample trees, only one diameter–height model for deciduous trees was estimated for all plots. The heights of pine trees were estimated using the models developed for spruce, owing to an insufficient number of sampled pines.

Table I. Summary of model calibration data ($n = 20$).

| Characteristic | Range | Mean |
|---|---|---|
| $d_{\mathrm{min}}$ (cm) | 3.0–4.0 | 3.2 |
| $d_{\mathrm{max}}$ (cm) | 33.3–60.6 | 47.0 |
| $d_{\mathrm{max}}$–$d_{\mathrm{min}}$ (cm) | 30.3–57.1 | 43.8 |
| $d_{\mathrm{BA}}$ (cm) | 14.60–30.32 | 21.11 |
| $h_{\mathrm{L}}$ (m) | 15.20–28.90 | 21.65 |
| $h_{\mathrm{dom}}$ (m) | 18.27–32.45 | 25.11 |
| $N$ (ha$^{-1}$) | 630–1780 | 1033 |
| BA (m$^2$ ha$^{-1}$) | 21.6–45.5 | 34.2 |
| $V$ (m$^3$ ha$^{-1}$) | 171.9–634.8 | 360.7 |
| GC (index value) | 0.47–0.69 | 0.58 |
| Tree species distribution (%) | | |
| Spruce | 71–100 | 90 |
| Pine | 0–2 | 0 |
| Deciduous | 0–28 | 10 |

Note: $d_{\mathrm{min}}$ = minimum diameter; $d_{\mathrm{max}}$ = maximum diameter; $d_{\mathrm{BA}}$ = basal area mean diameter; $h_{\mathrm{L}}$ = basal area weighted mean height; $h_{\mathrm{dom}}$ = dominant height; $N$ = stem number; BA = basal area; $V$ = volume; GC = Gini coefficient.

Planimetric coordinates (Euref89) were measured in the centre of each plot by means of differential Global Navigation Satellite Systems (dGNSS). A Topcon Legacy 20-channel dual-frequency receiver, observing the pseudorange and carrier phase of the Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS), was used as rover equipment. The receiver set-up had a 2 s logging rate, and all satellites below a 15° angle from the antenna (cut-off angle) were disregarded. Logging periods ranged between 0.5 and 1.5 h, with an average antenna height of 4 m.

An identical Topcon Legacy GPS + GLONASS receiver was established as a base station within a distance of < 2.5 km from the sample plots. Næsset (2001) found that the accuracy of coordinates measured by means of dGNSS could be expressed by two times the standard deviation of the corrected single coordinate observations reported by the applied Pinnacle version 1.0 software package (Anon., 1999) used for postprocessing. According to this, planimetric coordinates of the base station had an accuracy of 0.4 cm. Records from the base station were used as reference during postprocessing of the rover coordinates. To ensure that the base station received signals from the same satellites as the rover, the cut-off angle was set to 12°. The postprocessing of all rover records was also undertaken using Pinnacle software (Anon., 1999). Following the same accuracy interpretation as for the base station

Table II. Plotwise diameter-height models[a] for conifer species and the diameter-height model for deciduous species: degrees of freedom (df), intercept estimate ($\beta_0$), parameter estimate of diameter ($\beta_1$), coefficient of determination ($R^2$) and root mean square error (RMSE).

| | | Regression coefficients[c] | | | | |
|---|---|---|---|---|---|---|
| Plot no.[b] | df | $\beta_0$ | $\beta_1$ | $R^2$ | RMSE | $p$ (model) |
| 1 | 105 | 1.297ns | 0.9450*** | 0.91 | 0.30 | <0.0001 |
| 2 | 27 | 2.311ns | 0.7804*** | 0.79 | 0.27 | <0.0001 |
| 3 | 27 | 2.913* | 0.7611*** | 0.81 | 0.32 | <0.0001 |
| 4 | 38 | 23.28*** | 0.4296*** | 0.53 | 0.22 | <0.0001 |
| 5 | 33 | 5.418*** | 0.6623*** | 0.74 | 0.29 | <0.0001 |
| 6 | 20 | 5.114* | 0.6677*** | 0.68 | 0.28 | <0.0001 |
| 7 | 24 | 12.41*** | 0.5084*** | 0.60 | 0.26 | <0.0001 |
| 8 | 19 | 3.629** | 0.7129*** | 0.85 | 0.15 | <0.0001 |
| 9 | 27 | 7.258*** | 0.5928*** | 0.82 | 0.18 | <0.0001 |
| 10 | 22 | 2.073ns | 0.8012*** | 0.86 | 0.18 | <0.0001 |
| 11 | 27 | 2.129ns | 0.8098*** | 0.78 | 0.19 | <0.0001 |
| 12 | 25 | 3.858** | 0.7065*** | 0.80 | 0.15 | <0.0001 |
| 13 | 28 | 6.300*** | 0.6166*** | 0.87 | 0.17 | <0.0001 |
| 14 | 25 | 5.289*** | 0.6376*** | 0.84 | 0.17 | <0.0001 |
| 15 | 32 | 7.886*** | 0.5796*** | 0.74 | 0.18 | <0.0001 |
| 16 | 19 | 3.066*** | 0.7585*** | 0.92 | 0.20 | <0.0001 |
| 17 | 28 | 1.334ns | 0.9153*** | 0.91 | 0.22 | <0.0001 |
| 18 | 25 | 2.324** | 0.8063*** | 0.94 | 0.18 | <0.0001 |
| 19 | 25 | 5.371*** | 0.6735*** | 0.78 | 0.27 | <0.0001 |
| 20 | 25 | 1.647ns | 0.8679*** | 0.90 | 0.24 | <0.0001 |
| All | 73 | 7.473*** | 0.5938*** | 0.79 | 0.35 | <0.0001 |

Note: [a] $h = \beta_0 d_{bh}^{\beta 1}$; [b] models for conifer species are estimated plotwise and denoted by plot number; model for deciduous species is estimated from all deciduous sample trees and denoted "All"; [c] significance level: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns = not significant ($p > 0.05$).

coordinates, the average accuracy of the plot coordinates was 10 cm.

### *Validation data: inventory B*

Validation data (Table III) were collected during summer 2003 in the municipality of Nordre Land in south-eastern Norway (60°50′ N, 10°05′ E, 140–900 m a.s.l.). The data originate from 0.1 ha circular sample plots, which correspond to the model calibration data, and the plots were systematically distributed throughout a forest area of 25,000 ha. In inventory B, all trees ($d_{\mathrm{bh}} \geq 4$ cm) were callipered on a plot inner circle defined by a radius of 8.92 m from the plot centre (250 $\mathrm{m}^2$). Between this circumference and the radius defining the 0.1 ha plot, only trees with $d_{\mathrm{bh}} \geq 10$ cm were callipered. Consequently, the number of trees $4 \leq d_{\mathrm{bh}} < 10$ cm for the entire 0.1 ha plot had to be extrapolated from the inner circle under the assumption that the trees with $4 \leq d_{\mathrm{bh}} < 10$ cm were evenly spatially distributed. Calculations of plot data are in correspondence with inventory A.

The validation data comprised a wider range of size distributions. Both normal and uniform distributions were included in the material, as well as multimodal and reverse-J, because the area partly is influenced by traditional clear-cuttings. The model adequacy checking hence also included effects of forest structure. To separate these effects, the validation data were stratified (three strata) according to their GC value. Lexerød & Eid (2006) calculated GC for 16 simulated diameter distributions and found realistic ranges of different diameter diversity indices for normal, uniform and reverse-J diameter distribution. Based on these findings, stratum 1 was limited to GC values up to 0.30 (normal distributions). Stratum 2 included plots with GC values in the range between 0.31 and 0.46 (approximately uniform distributions). Finally, stratum 3 included plots in the same GC range as the model calibration data (GC >0.47).

### *Laser scanner data*

For inventory A, an ALTM 1233 laser scanning system produced by Optech, Canada, was used. The laser scanner data for inventory A were acquired using a Hughes 500 helicopter on 9 October 2003. The deciduous trees were still foliferous, but had a beginning xanthophyll colouring. The average flying altitude was approximately 600 m above the ground with an average speed of 35 $\mathrm{ms}^{-1}$. Twenty-one flightlines were flown, with an overlap between adjacent stripes of about 20%. The pulse repetition frequency was 33 kHz and the scan frequency was 50 Hz. Maximum scan angle was 11°, which corresponded to an average swath width of about 230 m.

Table III. Summary of validation data ($n = 18$).

| Characteristic[a] | Range | Mean |
|---|---|---|
| Stratum 1 ($n = 3$) | | |
| $d_{\text{min}}$ (cm) | 10.0–10.0 | 10.0 |
| $d_{\text{max}}$ (cm) | 30.0–40.0 | 33.3 |
| $d_{\text{max}}$–$d_{\text{min}}$ (cm) | 20.0–30.0 | 23.3 |
| $d_{\text{BA}}$ (cm) | 18.7–25.7 | 21.4 |
| $h_{\text{L}}$ (m) | 16.5–25.7 | 19.8 |
| $h_{\text{dom}}$ (m) | 20.4–26.9 | 22.9 |
| $N$ (ha$^{-1}$) | 590–1060 | 843 |
| BA (m$^2$ ha$^{-1}$) | 27.3–30.7 | 29.0 |
| $V$ (m$^3$ ha$^{-1}$) | 217.0–386.6 | 281.0 |
| GC (index value) | 0.27–0.30 | 0.28 |
| Tree species distribution (%) | | |
| Spruce | 96–100 | 99 |
| Pine | 0–0 | 0 |
| Deciduous | 0–4 | 1 |
| Stratum 2 ($n = 9$) | | |
| $d_{\text{min}}$ (cm) | 4.0–10.0 | 5.3 |
| $d_{\text{max}}$ (cm) | 30.0–50.0 | 39.1 |
| $d_{\text{max}}$–$d_{\text{min}}$ (cm) | 26.0–46.0 | 34.9 |
| $d_{\text{BA}}$ (cm) | 16.5–24.0 | 21.2 |
| $h_{\text{L}}$ (m) | 18.9–24.0 | 21.6 |
| $h_{\text{dom}}$ (m) | 21.0–26.9 | 23.9 |
| $N$ (ha$^{-1}$) | 760–2020 | 1172 |
| BA (m$^2$ ha$^{-1}$) | 32.4–46.3 | 39.2 |
| $V$ (m$^3$ ha$^{-1}$) | 303.0–519.4 | 397.1 |
| GC (index value) | 0.36–0.46 | 0.42 |
| Tree species distribution (%) | | |
| Spruce | 93–100 | 99 |
| Pine | 0–0 | 0 |
| Deciduous | 0–7 | 1 |
| Stratum 3 ($n = 6$) | | |
| $d_{\text{min}}$ (cm) | 4.0–4.0 | 4.0 |
| $d_{\text{max}}$ (cm) | 28.0–44.0 | 36.3 |
| $d_{\text{max}}$–$d_{\text{min}}$ (cm) | 24.0–40.0 | 32.3 |
| $d_{\text{BA}}$ (cm) | 13.5–22.0 | 17.8 |
| $h_{\text{L}}$ (m) | 14.5–24.0 | 21.4 |
| $h_{\text{dom}}$ (m) | 15.0–27.2 | 23.7 |
| $N$ (ha$^{-1}$) | 790–2270 | 1333 |
| BA (m$^2$ ha$^{-1}$) | 11.3–40.6 | 32.7 |
| $V$ (m$^3$ ha$^{-1}$) | 72.1–441.9 | 338.9 |
| GC (index value) | 0.51–0.64 | 0.55 |
| Tree species distribution (%) | | |
| Spruce | 67–100 | 89 |
| Pine | 0–33 | 8 |
| Deciduous | 0–11 | 3 |

Note: [a] for description, see Table I.

Pulses transmitted at scan angles that exceeded 10.5° were excluded from the final data set. The average footprint diameter for individual plots was approximately 18 cm. The mean number of pulses transmitted was 5.0 m$^{-2}$. First and last returns were recorded.

The laser scanner data for inventory B were acquired during summer 2003. The same laser scanner system as for inventory A was used. The laser was carried by a Piper PA31-310 aircraft. The average flying altitude was approximately 800 m above ground, and the sampling density was about 0.7 pulses m$^{-2}$. First and last pulse data were supplied.

The initial processing of the laser data was carried out by the contractor (Blom Geomatics, Norway). This processing includes computation of planimetric coordinates and ellipsoidal height values for all first and last returns. Furthermore, a matching between swaths was performed to remove orientation errors. The last return data were used to model the terrain surface. In a filtering operation undertaken by the contractor using the Terrascan software package (Anon., 2005), local maxima assumed to represent vegetation hits were discarded. A triangulated irregular network (TIN) was generated from the planimetric coordinates and corresponding height values of the individual terrain ground points retained in the last pulse data. The ellipsoidal height accuracy of the TIN model was expected to be around 20–30 cm (Kraus & Pfeifer, 1998; Reutebuch et al., 2003; Hodgson & Bresnahan, 2004).

All transmitted pulses that were classified as first or last returns from the vegetation were geographically registered. The first step in this process was to spatially register all first and last return observations (points) to the TIN according to their coordinates. Secondly, terrain surface height values were computed for each point by bilinear interpolation from the TIN. The height above ground of each point was computed as the difference between the height of the first or last return and the terrain surface height. Furthermore, observations with a height value less than 2 m were excluded from the two first data sets to eliminate the effect of stones, shrubs, etc., from the tree canopy data sets (Nilsson, 1996). Thus, the data retained for analysis were geographically registered data on canopy heights derived from the first and last returns. The data were spatially registered to the field plots. Pulses that hit outside the plots were excluded from further analysis.

### *Data preparation*

First and last pulse height distributions were created from the laser height values ($>2$ m, see above) of each 0.1 ha sample plot. From these distributions a total of 46 variables was derived. Ten percentiles (0%, 10%, …, 90%) characterized both first and last pulse laser heights. Accordingly, canopy density was expressed by cumulative proportions of laser hits, in 10 uniform fractions across the range of the laser heights ($>2$ m), of the total number of pulses. Moreover, for the first and last pulse laser height values ($>2$ m), means, standard deviations and coefficients of variation were derived. Further details are provided by Næsset (2004*b*).

Empirical diameter distributions of each plot were derived as the number of callipered trees in predefined 2 cm diameter classes, where the specific value of each class corresponds to class centre. The diameter measurements were used to compute stand BA. Diameters at 10 percentiles of BA defined an empirical cumulative probability density function of each plot. A model for the 10 percentiles ($d_{10}$, $d_{20}$, ..., $d_{100}$) and BA was estimated with the 46 laser variables as potential regressors. Numerous applications (e.g. Nelson et al., 1988, 1997; Næsset, 1997, 2002, 2004*b*; Nelson, 1997; Lefsky et al., 1999; Magnussen et al., 1999; Means et al., 1999, 2000; Lim et al., 2003) have successfully used such laser variables for modelling biophysical forest properties. The present choice of variables was based on these empirical findings.

### Statistical methods

When modelling diameter distributions using the percentile method, previous authors have most frequently used seemingly unrelated regression (SUR) (Zellner, 1962) to estimate model parameters (Kangas & Maltamo, 2000*a*, *b*; Maltamo et al., 2000; Gobakken & Næsset, 2005). Since the 46 laser variables are derived from the same height distributions, there is a high level of collinearity among them. Because of the collinearity, only a few variables can be selected for the final model and variables that hold significant information may be excluded. In the current trial partial least squares regression (PLSR) (Wold et al., 1983; Martens, 2001) was used to estimate the relationships between the 10 percentiles, basal area and the laser data. Another study that used PLSR to predict biophysical properties of forest stands using laser data was presented by Næsset et al. (2005). This study compared PLSR, SUR and ordinary least squares (OLS) for modelling $h_L$, $h_{dom}$, $d_{BA}$, $N$, BA, and V. PLSR is an alternative to SUR, as it is based on extracting linear combinations from the explanatory variables, compressed into a few orthogonal principal components or latent variables. These latent variables represent the main variance structures among the explanatory variables that maximize the correlation to the variance structures of the response variables. The emphasis of this method is prediction, rather than understanding the causal relationships between predictors and responses. PLSR is related to principal component regression (PCR). The main difference is that PCR does not extract latent variables that are dependent on the variance structure of the dependent variables. Like SUR, the PLSR procedure also allows for many responses to be modelled simultaneously. Another important rationale for using PLSR in the current trial was that the data at hand contained fewer observations than explanatory variables. The resulting negative degrees of freedom make the fitting of an OLS model impossible, or at least makes the variable selection process difficult.

### Model calibration and validation

For all data analyses, the PLS procedure of SAS (Tobias, 1995) including additional macros (Anon., 1997) was used. Initially, the best number of latent variables ($A_{opt}$) was estimated by leave-one-out cross-validation of a preliminary estimation. A common way of using cross-validation for determining $A_{opt}$, is to select the number of components that minimizes RMSE. It is, however, indicated that this approach has a tendency to overfit the model (Nørgaard et al., 2000). In the present work, $A_{opt}$ was found by adding a punishing factor of 3% to the lowest root mean square error ($RMSE_{min}$), as suggested by Westad (1999). Thus, the lowest number of components with $RMSE < 1.03 \times RMSE_{min}$ was selected. Outliers and variables representing merely statistical noise were detected in this step. Outlying observations were considered in terms of their Euclidian distance to the model. The variable importance for the projection (VIP) criterion was considered when deciding on elimination of variables. Variables that were of little importance for explaining the variation in the response values, i.e. having a VIP value $<0.8$ (Wold, 1995), were excluded from the model. Finally, a re-estimation without outliers and insignificant variables was carried out.

When applying the model for prediction, the cumulative percentage of BA in each diameter class was found by linear interpolation between percentiles, therefore assuming a uniform distribution between neighbouring percentiles. The 0% percentile was truncated to 3 cm (lower limit of the smallest diameter class). To avoid negative frequencies, the values of following predicted percentiles have to increase ($d_0 < d_{10} < \ldots < d_{100}$). In the current trial, the range of diameters was sufficiently wide and evenly distributed so that decreasing values were avoided. If decreasing values occur, the decreasing interval can be set to zero by adjusting the larger percentile to the level of the adjacent lower percentile. Furthermore, the number of trees in each diameter class was found by scaling the relative basal area to the predicted stand basal area.

For descriptive inventory purposes, an accurately predicted stand volume is *per se* independent of the predicted diameter distribution. If underpredicted diameter classes are compensated by an overprediction in other diameter classes, the result will still be adequate. However, if the predicted diameter distribution is to be used as input in a forecast and the different diameter classes are to be projected for a given period, accurate prediction of stem number in each diameter class becomes important. If the number of stems is poorly determined, the planning of harvests and silvicultural treatments will have a poor basis, as the projection forecasts harvestable volumes and need for treatments at incorrect points in time. Thus, the current trial appraised the accuracy of the diameter distribution model by focusing on both the number of stems and the volume.

The derived model was both cross-validated and validated on an independent data set. Since the validation data included a wide range of diameter distribution types, this data set was stratified to separate the effect of stand structures. Mean differences between observed and predicted number of stems per hectare were calculated for each diameter class. The statistical significance of the mean differences was assessed by two-tailed $t$-tests. Other authors (e.g. Gobakken & Næsset, 2005) have evaluated their model by means of an error index (2) proposed by Reynolds et al. (1988). To compare the present work with Gobakken and Næsset (2005) the same index was computed. The error index $e$ was computed as the sum of the absolute deviations of the predicted minus the observed number of trees in each diameter class relative to the total number of observed trees, i.e.

$$e = \frac{\sum_{j=1}^{k} |n_{pj} - n_{oj}|}{N} 100 \qquad (2)$$

where $n_{pj}$ and $n_{oj}$ are the predicted and observed number of trees, respectively, in diameter class $j$, $j = 1, 2, \ldots, k$, and $N$ is the total number of trees according to the field inventory. As used here, the error index gives an overall description of the absolute deviations between the predicted and observed number of trees in all diameter classes. Finally, the difference between observed volume per hectare and the predicted volume per hectare was assessed. Plot volume was derived from the resulting diameter distribution by using single-tree volume models to compute the volume of the centre tree in each diameter class, multiplied by the number of trees in each diameter class. Tree height in each diameter class was computed by the same diameter-height models used to derive the ground-truth volume. Moreover, the species distribution observed in the field was used to discriminate between conifer and deciduous tree volume. The statistical significance of the mean differences was assessed by two-tailed $t$-tests.

## Results

Initially, the percentiles ($d_{10}$, $d_{20}$, $\ldots$, $d_{100}$), were derived from the empirical distributions of each plot's field values, and BA was computed from individual tree diameters. The percentiles and BA were then modelled simultaneously from the laser variables by means of PLSR.

The derived model included three latent variables (Table IV), derived as linear combinations of 43 of the 46 potential laser variables. Hence, three laser variables were excluded from the model on the basis of the VIP criterion. These variables were the 0% percentile of the height distribution for the first pulse data and the proportion of laser hits in the uppermost fraction of the laser heights, for both the first and last pulse data. The $R^2$ ranged between 0.44 and 0.80 for the percentiles and was 0.67 for BA. For the percentiles the RMSE ranged between 2.34 cm ($d_{60}$) and 6.01 cm ($d_{100}$) and was 3.09 m$^2$ ha$^{-1}$ for BA.

The cross-validation revealed a non-significant ($p > 0.05$) difference between the predicted and observed number of stems for 22 of 29 diameter classes (Table V). The number of stems of the two smaller diameter classes was greatly overpredicted, but for the rest of the diameter classes the differences were relatively small. The signs of the differences were positive for all diameter classes up to 30 cm, and negative for the rest. The mean difference between observed and predicted volume was −3.3% ($p > 0.05$). The standard deviation of the differences was 11% of observed volume (see Table X). The mean value of the error index was 78.1.

Table IV. Estimation results from the partial least squares modelling: $R^2$ and root mean square error (RMSE) for the 10 percentiles of the basal area distribution and basal area (BA).

| Dependent variable[a] | $R^2$ | RMSE (cm) |
|---|---|---|
| $d_{10}$ | 0.57 | 2.39 |
| $d_{20}$ | 0.69 | 2.58 |
| $d_{30}$ | 0.62 | 2.90 |
| $d_{40}$ | 0.70 | 2.82 |
| $d_{50}$ | 0.73 | 2.71 |
| $d_{60}$ | 0.80 | 2.34 |
| $d_{70}$ | 0.67 | 3.20 |
| $d_{80}$ | 0.75 | 3.20 |
| $d_{90}$ | 0.75 | 3.21 |
| $d_{100}$ | 0.44 | 6.01 |
| BA | 0.67 | 3.09 |

Note: [a] $d_i = Xb_i$; $X$ = laser data matrix; $b_i$ = PLS regression coefficient vector. The model includes three latent variables that are linear combinations of 43 derived laser variables.

Table V. Differences between predicted and observed mean number of trees for 2 cm diameter classes based on cross-validation.

| Diameter class | Observed mean no. of trees | Mean difference[a] | Min. difference | Max. difference | SD |
|---|---|---|---|---|---|
| 4 | 131.5 | 289.5*** | 127.1 | 433.6 | 82.3 |
| 6 | 102.0 | 85.1*** | −1.3 | 159.4 | 48.4 |
| 8 | 90.5 | 15.8ns | −127.4 | 79.7 | 51.6 |
| 10 | 78.0 | 0.7ns | −97.9 | 83.3 | 41.5 |
| 12 | 68.5 | 8.9ns | −74.5 | 89.6 | 40.7 |
| 14 | 59.5 | 8.5ns | −71.1 | 72.1 | 37.1 |
| 16 | 54.5 | 5.9ns | −65.7 | 40.0 | 25.6 |
| 18 | 55.5 | 1.4ns | −97.0 | 43.6 | 29.5 |
| 20 | 49.0 | 15.9* | −23.8 | 100.7 | 29.8 |
| 22 | 37.5 | 16.0* | −33.0 | 68.5 | 26.8 |
| 24 | 42.0 | 13.7* | −14.6 | 66.2 | 22.4 |
| 26 | 42.0 | 6.4ns | −52.7 | 46.7 | 25.5 |
| 28 | 38.5 | 1.0ns | −90.2 | 41.3 | 27.7 |
| 30 | 35.0 | 1.2ns | −51.0 | 33.3 | 22.6 |
| 32 | 32.0 | −2.4ns | −31.5 | 34.0 | 15.6 |
| 34 | 31.0 | −8.5* | −31.8 | 13.2 | 13.8 |
| 36 | 19.0 | −1.2ns | −28.2 | 19.5 | 12.3 |
| 38 | 22.0 | −6.3ns | −44.0 | 9.8 | 14.0 |
| 40 | 14.0 | −4.2* | −18.0 | 13.3 | 9.0 |
| 42 | 8.0 | −0.8ns | −18.9 | 11.9 | 8.8 |
| 44 | 5.5 | 0.2ns | −11.9 | 9.2 | 5.7 |
| 46 | 6.0 | −2.8ns | −22.7 | 7.2 | 6.9 |
| 48 | 5.0 | −2.6ns | −20.0 | 6.3 | 6.4 |
| 50 | 2.0 | −1.2ns | −18.2 | 10.0 | 5.7 |
| 52 | 1.0 | −0.6ns | −10.0 | 8.4 | 3.7 |
| 54 | 0.5 | −0.1ns | −10.0 | 7.1 | 2.8 |
| 56 | 1.5 | −1.2ns | −10.0 | 0.0 | 3.1 |
| 58 | 0.5 | −0.5ns | −10.0 | 0.0 | 2.2 |
| 60 | 0.5 | −0.5ns | −10.0 | 0.0 | 2.2 |

Note: [a] significance level: ***$p<0.001$, *$p<0.05$, ns = not significant ($p>0.05$).

Independent validation on the full range of the validation data (Table VI) revealed that the model performed quite well. However, significant overprediction of the number of stems for the smallest diameter classes was found. Non-significant differences between the predicted and observed number of stems were found for 20 of 24 diameter classes. The mean difference between the predicted and observed volume for all validation plots was $-8.3\%$ ($p<0.05$) of mean observed volume, with a corresponding standard deviation for the differences of 14.2%. The error index indicates similar absolute difference between predicted and observed number of stems as for the cross-validation.

The validation data were stratified according to the GC value of each plot. Stratum 1 included data from the plots with GC value $<0.30$, indicating normal distributions. Table VII shows that the differences between predicted and observed number of stems were non-significant for 17 of 20 diameter classes. The table also shows that the number of stems of small diameter classes was significantly overpredicted. The mean difference between predicted and observed volume was non-significant (see Table X). Moreover, validation on this part of the data set revealed the largest error index value.

Plots where the GC value were between 0.31 and 0.46 constituted stratum 2. This range of the GC is assumed to indicate uniform distributions. Table VIII shows that 20 of 24 diameter classes had non-significant differences between predicted and observed number of stems. Similar to the results in Tables V–VII, small diameter classes were overpredicted with respect to number of stems. The mean of the volume differences (see Table X) indicates a significant underprediction of volume.

Finally, stratum 3 contained plots where the GC values were in the same range as the model calibration data ($\text{GC}>0.47$). Table IX shows that the difference between predicted and observed number of stems was significant for only one diameter class. Contrary to the validation on the other strata, there seems to be no significant overprediction for the small diameter classes. The mean difference between predicted and observed volume was non-significant (Table X). The error index for this subset was lower than for all other subsets.

Table VI. Differences between predicted and observed mean number of trees for 2 cm diameter classes based on full range of validation data ($n=18$).

| Diameter class | Observed mean no. of trees | Mean difference$^{a}$ | Min. difference | Max. difference | SD |
|---|---|---|---|---|---|
| 5 | 133.3 | 154.5*** | −247.0 | 309.8 | 155.7 |
| 7 | 88.9 | 62.6* | −173.2 | 158.1 | 91.7 |
| 9 | 62.2 | 30.5ns | −136.5 | 155.9 | 71.8 |
| 11 | 97.8 | −8.6ns | −107.1 | 100.1 | 54.3 |
| 13 | 76.1 | −8.4ns | −124.0 | 47.3 | 41.2 |
| 15 | 82.8 | 2.4ns | −129.3 | 172.9 | 61.9 |
| 17 | 95.0 | −18.2ns | −132.5 | 131.5 | 62.1 |
| 19 | 80.0 | −7.9ns | −115.6 | 99.2 | 54.7 |
| 21 | 87.2 | −34.1ns | −164.0 | 48.4 | 48.6 |
| 23 | 82.8 | −16.9ns | −142.0 | 94.7 | 51.7 |
| 25 | 68.9 | −5.4ns | −66.3 | 87.1 | 42.8 |
| 27 | 55.0 | 6.0ns | −63.1 | 112.1 | 47.0 |
| 29 | 55.0 | −15.9* | −65.5 | 70.4 | 30.9 |
| 31 | 37.8 | −10.2ns | −78.5 | 28.5 | 23.7 |
| 33 | 26.7 | −6.5ns | −84.1 | 23.0 | 29.4 |
| 35 | 15.0 | 1.0ns | −50.0 | 39.8 | 20.4 |
| 37 | 7.2 | 4.2ns | −40.0 | 23.7 | 15.8 |
| 39 | 10.0 | −1.1ns | −21.4 | 14.6 | 11.1 |
| 41 | 6.1 | 0.4ns | −16.2 | 10.8 | 7.7 |
| 43 | 1.7 | 3.6** | −1.4 | 11.0 | 4.7 |
| 45 | 1.1 | 1.9* | −2.6 | 8.7 | 3.2 |
| 47 | 0.0 | 0.2ns | 0.0 | 4.0 | 0.9 |
| 49 | 0.0 | 0.0ns | 0.0 | 0.0 | 0.0 |
| 51 | 0.6 | −0.6ns | −10.0 | 0.0 | 2.4 |

Note: $^{a}$ significance level: ***$p<0.001$, **$p<0.01$, *$p<0.05$, ns = not significant ($p>0.05$).

## Discussion

The current trial predicted diameters at 10 percentiles of BA, scaled to the predicted BA to estimate diameter distributions. Results from the cross-validation (Table V) showed that the predicted number of stems was insignificantly different from the observed for most diameter classes, but that the number of small trees was poorly predicted. This can

Table VII. Differences between predicted and observed mean number of trees for 2 cm diameter classes based on validation data (stratum 1, $n=3$).

| Diameter class | Observed mean no. of trees | Mean difference$^{a}$ | Min. difference | Max. difference | SD |
|---|---|---|---|---|---|
| 5 | 0.0 | 272.4** | 246.7 | 309.8 | 33.2 |
| 7 | 0.0 | 139.0** | 125.8 | 158.1 | 16.9 |
| 9 | 0.0 | 84.1** | 76.1 | 95.6 | 10.2 |
| 11 | 66.7 | 16.4ns | 3.4 | 41.0 | 21.3 |
| 13 | 56.7 | 11.9ns | −5.8 | 36.5 | 22.0 |
| 15 | 100.0 | −35.0ns | −129.3 | 24.6 | 82.6 |
| 17 | 113.3 | −45.5ns | −118.9 | 22.7 | 70.9 |
| 19 | 133.3 | −78.8ns | −115.6 | −29.8 | 44.2 |
| 21 | 103.3 | −41.6ns | −73.4 | 6.5 | 42.4 |
| 23 | 73.3 | −21.1ns | −39.6 | 0.0 | 19.9 |
| 25 | 50.0 | 12.7ns | 6.8 | 23.0 | 9.0 |
| 27 | 56.7 | 6.4ns | −18.8 | 25.5 | 22.8 |
| 29 | 36.7 | 21.1ns | −27.5 | 70.4 | 48.9 |
| 31 | 30.0 | −6.9ns | −19.0 | 5.0 | 12.0 |
| 33 | 3.3 | 13.7ns | 10.4 | 19.6 | 5.1 |
| 35 | 10.0 | 3.7ns | −13.2 | 13.5 | 14.7 |
| 37 | 3.3 | 4.3ns | −5.5 | 12.1 | 9.0 |
| 39 | 3.3 | 0.0ns | −10.0 | 9.9 | 10.0 |
| 41 | 3.3 | −0.3ns | −10.0 | 9.0 | 9.5 |
| 43 | 0.0 | 1.7ns | 0.0 | 5.1 | 2.9 |

Note: $^{a}$ significance level: **$p<0.01$, ns = not significant ($p>0.05$).

Table VIII. Differences between predicted and observed mean number of trees for 2 cm diameter classes based on validation data (stratum 2, $n = 9$).

| Diameter class | Observed mean no. of trees | Mean difference$^a$ | Min. difference | Max. difference | SD |
|---|---|---|---|---|---|
| 5 | 66.7 | 231.8*** | 135.3 | 305.2 | 54.6 |
| 7 | 62.2 | 90.0*** | 16.1 | 155.7 | 49.8 |
| 9 | 62.2 | 29.9ns | −136.5 | 94.2 | 71.9 |
| 11 | 118.9 | −13.7ns | −82.7 | 92.7 | 50.5 |
| 13 | 80.0 | −7.1ns | −50.5 | 47.3 | 31.0 |
| 15 | 86.7 | 9.9ns | −42.7 | 172.9 | 63.6 |
| 17 | 102.2 | −11.5ns | −132.5 | 131.5 | 74.7 |
| 19 | 77.8 | 3.4ns | −51.0 | 49.1 | 33.2 |
| 21 | 98.9 | −45.2* | −164.0 | 21.7 | 57.1 |
| 23 | 92.2 | −28.4ns | −142.0 | 94.7 | 66.0 |
| 25 | 84.4 | −19.9ns | −66.3 | 87.1 | 45.2 |
| 27 | 57.8 | −6.8ns | −63.1 | 88.6 | 44.7 |
| 29 | 53.3 | −16.9ns | −65.5 | 5.1 | 22.1 |
| 31 | 48.9 | −18.5ns | −78.5 | 5.9 | 26.7 |
| 33 | 32.2 | −11.5ns | −84.1 | 23.0 | 34.7 |
| 35 | 15.6 | 2.9ns | −30.0 | 39.8 | 20.9 |
| 37 | 8.9 | 4.4ns | −20.0 | 22.0 | 12.1 |
| 39 | 12.2 | −0.9ns | −21.4 | 11.3 | 11.4 |
| 41 | 5.6 | 2.4ns | −10.0 | 10.8 | 6.8 |
| 43 | 3.3 | 3.2ns | −1.4 | 9.8 | 4.8 |
| 45 | 1.1 | 2.7* | −0.6 | 7.8 | 3.1 |
| 47 | 0.0 | 0.4ns | 0.0 | 4.0 | 1.3 |
| 49 | 0.0 | 0.0ns | 0.0 | 0.0 | 0.0 |
| 51 | 1.1 | −1.1ns | −10.0 | 0.0 | 3.3 |

Note: $^a$ significance level: ***$p < 0.001$, *$p < 0.05$, ns = not significant ($p > 0.05$).

be explained by the use of basal area distributions. Since the cumulative basal area increases at a slow rate for the small diameters, the interval between the smallest diameter and the 10th percentile will tend to be wide. Interpolation between these points of the basal area distribution will consequently yield less

Table IX. Differences between predicted and observed mean number of trees for 2 cm diameter classes based on validation data (stratum 3, $n = 6$).

| Diameter class | Observed mean no. of trees | Mean difference$^a$ | Min. difference | Max. difference | SD |
|---|---|---|---|---|---|
| 5 | 300.0 | −20.2ns | −247.0 | 173.4 | 146.7 |
| 7 | 173.3 | −16.6ns | −173.2 | 145.6 | 110.0 |
| 9 | 93.3 | 4.7ns | −91.1 | 155.9 | 81.8 |
| 11 | 81.7 | −13.4ns | −107.1 | 100.1 | 72.9 |
| 13 | 80.0 | −20.6ns | −124.0 | 25.3 | 60.0 |
| 15 | 68.3 | 9.8ns | −58.1 | 65.6 | 53.2 |
| 17 | 75.0 | −14.7ns | −74.9 | 50.6 | 40.3 |
| 19 | 56.7 | 10.5ns | −92.6 | 99.2 | 62.9 |
| 21 | 61.7 | −13.6ns | −43.0 | 48.4 | 36.6 |
| 23 | 73.3 | 2.3ns | −31.2 | 64.7 | 35.5 |
| 25 | 55.0 | 7.3ns | −44.8 | 70.1 | 46.5 |
| 27 | 50.0 | 25.1ns | −47.5 | 112.1 | 58.1 |
| 29 | 66.7 | −32.7** | −49.2 | −2.4 | 18.8 |
| 31 | 25.0 | 0.5ns | −26.1 | 28.5 | 20.8 |
| 33 | 30.0 | −9.2ns | −59.2 | 19.7 | 26.4 |
| 35 | 16.7 | −3.3ns | −50.0 | 20.8 | 24.3 |
| 37 | 6.7 | 4.0ns | −40.0 | 23.7 | 24.0 |
| 39 | 10.0 | −1.8ns | −20.0 | 14.6 | 13.0 |
| 41 | 8.3 | −2.3ns | −16.2 | 10.0 | 8.8 |
| 43 | 0.0 | 5.0ns | 0.0 | 11.0 | 5.5 |
| 45 | 1.7 | 1.5ns | −2.6 | 8.7 | 3.9 |

Note: $^a$ significance level: **$p < 0.01$, ns = not significant ($p > 0.05$).

Table X. Mean difference ($\overline{D}$) between predicted volume according to the predicted distributions and observed plot volume ($V$), and standard deviation for the differences (SD); from cross-validation (CV) and independent validation (IV).

| Validation method | No. of observations | Observed mean $V$ | $(\overline{D})^a$ (%) | SD (%) | Error index | |
|---|---|---|---|---|---|---|
| | | | | | Range | Mean |
| CV | 20 | 360.7 | −3.3ns | 11.0 | 41.5–126.8 | 78.1 |
| $\text{IV}_{\text{Stratum 1}}$ | 3 | 281.0 | 0.6ns | 13.4 | 95.9–131.8 | 111.3 |
| $\text{IV}_{\text{Stratum 2}}$ | 9 | 397.1 | −12.6* | 13.7 | 49.4–106.2 | 77.0 |
| $\text{IV}_{\text{Stratum 3}}$ | 6 | 338.9 | −4.4ns | 13.1 | 26.5–97.6 | 59.7 |
| $\text{IV}_{\text{Full range}}$ | 18 | 358.3 | −8.3* | 14.2 | 26.5–131.8 | 77.0 |

Note: $^a$ significance level: $^*p < 0.05$, ns = not significant ($p > 0.05$).

accurate estimates for number of stems compared with the larger diameters where the intervals between percentiles are narrower. Thus, the basal area distribution is giving more weight to the larger stems. This can explain the insignificant difference (−3.3% of observed volume) between $V$ estimated from the predicted diameter distribution and observed $V$ (Table X). Since the number of stems of the diameter classes that hold the large volumes is accurately determined, total volume will also tend to be more accurately estimated than number of stems. Error index values of the cross-validation ranged between 41.5 and 126.8, with a mean value of 78.1. This result is similar to the findings in Gobakken & Næsset (2005) for even-aged, managed forest. Maltamo et al. (2004) reported a bias of −6.3% of observed volume, while Gobakken & Næsset (2005) reported a bias of −0.3%. However, differences in methodology and forest structure of the test areas must be taken into account when comparing these results. Although the forest structures in the trial area studied by Maltamo et al. (2004) and the current experiment are similar, Maltamo et al. (2004) focused on segmentation of large individual trees and determining the small trees by fitting either a complete Weibull distribution or a left-truncated Weibull distribution. Still, the results presented in the current study imply that using laser data on a plot level, and statistically computing variables that represent the stand structure, may give predictions of comparable accuracy to utilization of single-tree segmentation combined with theoretical probability density functions. In both approaches it is the small trees that are the main challenge because they are shaded by the dominant canopy layer. However, by including laser-derived variables from the last pulse data, some variation in the biological matter of the sublayers is likely to be accounted for (Næsset, 2004*a*; Maltamo et al., 2005). Næsset (2004*a*) found that the penetration rate of the last pulse was about 40% in spruce-dominated mature forest on good sites, whereas the first pulse penetration in the same forest was only 15%. Furthermore, Maltamo et al. (2005) were able to classify 24 out of 28 plots correctly with respect to forest structure classes using the HistMod algorithm (Lloyd, 1982; Maltamo et al., 2005) in a heterogeneous forest.

Results from the validation showed that the predicted number of stems was fairly accurate for stratum 3 (Table IX). An example from one of the sample plots of stratum 3 is displayed in Figure 1. Because the model was calibrated to reproduce multimodal distributions with a large proportion of small trees, this result should not be surprising. However, the model also predicted distributions of similar shapes even if the observed diameters in fact were normally or uniformly distributed. Figure 2 shows the observed and predicted number of stems of a stratum 1 sample plot. Because of the somewhat rigid shapes of the predicted distributions, the small trees were overpredicted for strata 1 and 2 (Tables VII and VIII, respectively). Furthermore, the predicted number of small trees had a somewhat narrow range. This implies that the variation in the lower canopy storeys has too little impact on the laser variables, yielding little variation in the predicted number of small trees. Moreover, most of the timber volume is associated with the dominating canopy layer in boreal forest (Maltamo et al., 2004). This was reflected by the fairly small differences between predicted and observed volume (Table X) and the fact that there seemed to be no trend between the error index and volume differences. The only significant volume difference was found for stratum 2. However, the table should be read with caution because of the limited number of observations in each stratum. With the exception of stratum 1, all volume estimates were slightly negatively biased, and it can be read from Table X that the bias increased with increasing observed volume. This increasing bias was probably due to the fact that the relative above-ground biomass distributed among leaves, needles, branches and

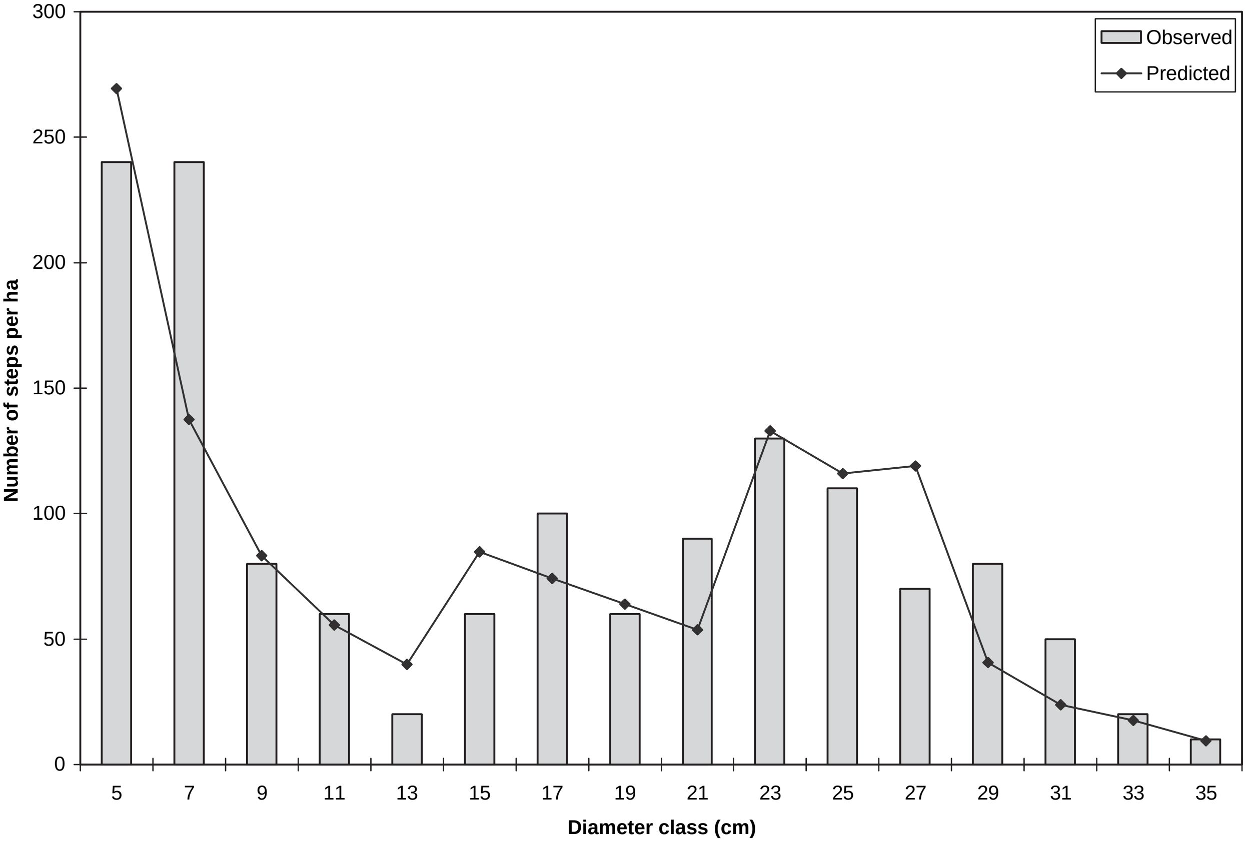

Figure 1. Predicted diameter distribution (line) and observed diameter distribution (bars) for stratum 3 of the validation data. The value of each diameter class refers to class centre.

stems will be variable owing to stand vigour (Solberg, 1999) for forest that reaches a relatively old age and large timber volumes. Since laser variables are affected by the penetration rate of the laser pulses into the canopy, loss of needles and leaves will yield lower volume estimates based on laser data if the model is calibrated on more vigorous forest. In addition, the proportion of large dead standing trees is likely to vary relative to total stand volume. Thus, the bias and random errors of predictions will tend to be larger for this type of forest because the laser-depicted canopy surface may vary considerably.

The validation was affected by different flying altitudes and laser pulse sampling densities between inventories A and B. These effects were, however, likely to be small. Næsset (2004*a*) compared small-footprint data collected from flying altitudes of 540 and 850 m, and found that both the laser variables and volume estimates were robust with respect to alterations in flying altitude. The differences in flying altitude in the present study were even smaller than in the trial by Næsset (2004*a*). Furthermore, Holmgren (2003) reported only minor differences in volume estimated from low-density laser data (0.1 pulse $m^{-2}$) and high-density laser data (4.3 pulses $m^{-2}$). However, the results in Holmgren (2003) were obtained by using a medium-footprint (3.68 m) laser scanner, and the footprints of the high-density data were thus overlapping. The effect of higher density on the precision of the volume estimates will thus be reduced, and it is likely that the effect will be more pronounced for small-footprint data with a footprint diameter of, say, 10–30 cm.

As indicated previously, an accurately predicted diameter distribution is important when the predicted distribution is to be the basis of a forecast, projecting each diameter class for a given period. Because this trial has demonstrated that the model is rigid in relation to reproducing diameter distributions other than multimodal, reverse-J, there is a need for some *a priori* information about the forest structure before the model can be used. In most cases, such information can be acquired from current forest management plans. Furthermore, the method proposed here requires that some field sample plots are established to be able to calibrate the model, and the required forest structure information can thus be derived directly from the model calibration data. However, to be independent of

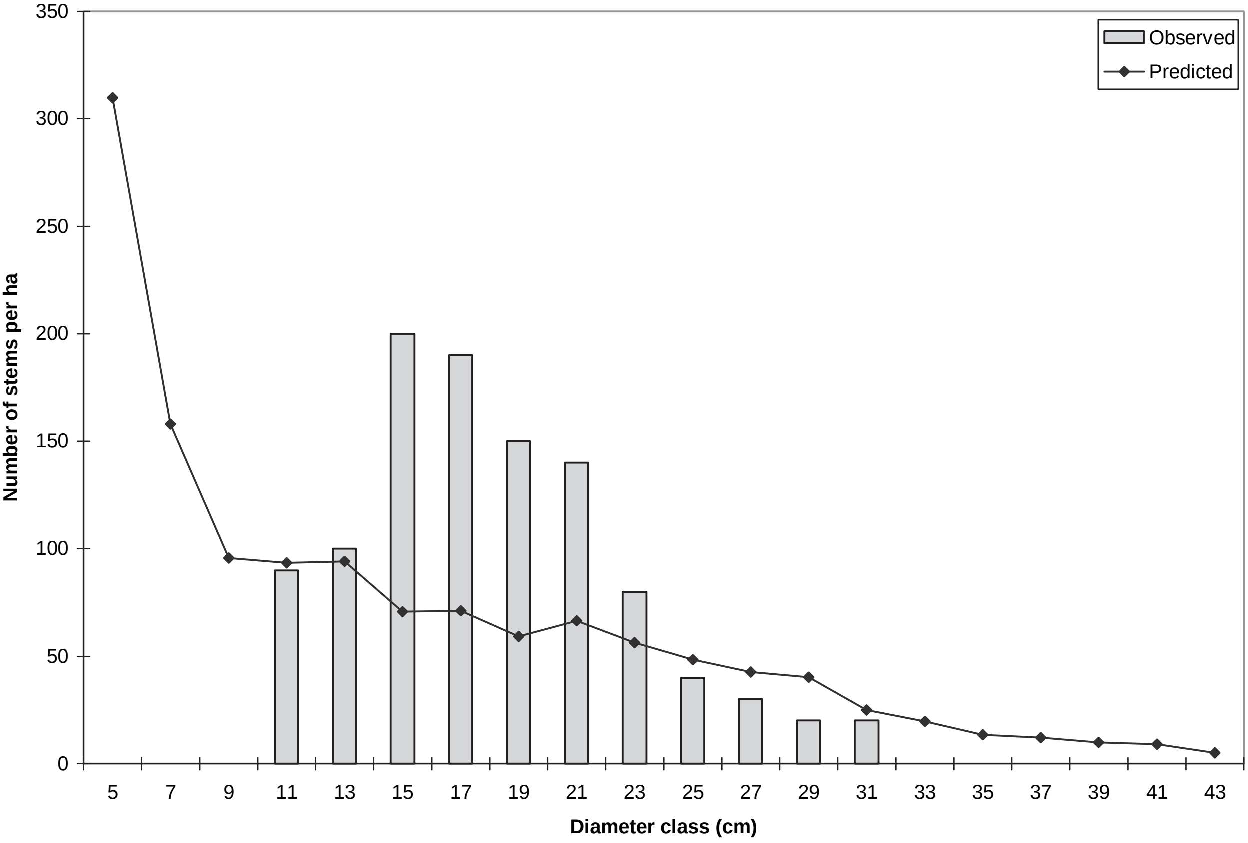

Figure 2. Predicted diameter distribution (line) and observed diameter distribution (bars) for stratum 1 of the validation data. The value of each diameter class refers to class centre.

such data sources, a strategy where the structure information is extracted directly from laser data can be developed. This can be done either by developing a classification rule that discriminates forest of different distribution types, or by estimating a model that predicts the GC value, which in turn can be used as a basis for stratification.

This trial has demonstrated that multimodal diameter distributions of fixed sample plots can be reproduced using the proposed method. Future work should focus on procedures to implement this method in operational forest inventories where entire forest stands are the target units.

## Acknowledgements

This research was founded by the Research Council of Norway (research project no. 153185/110). We wish to thank Associate Professor Terje Gobakken for help with SAS programming and two anonymous referees for valuable comments. Thanks also to Blom Geomatics AS for collection and processing of the laser scanner data, and Prevista AS for providing field data and laser data used as validation data in this study.

## References

Anon. (1997). *Examples using the PLS procedure*. Cary, NC: SAS Institute.

Anon. (1999). *Pinnacle user's manual*. San Jose, CA: Javad Positioning Systems.

Anon. (2005). *TerraScan user's guide*. Helsinki: Terrasolid. From http://terrasolid.fi. Accessed 1.1.2006.

Bailey, R. L. (1974). Weibull model for *Pinus radiata* diameter distributions. In *Proceedings of the Fourth Conference of the Advisory Group of Forest Statistics* (pp. 51–59). IUFRO Subject Group S6.02.

Bailey, R. L. & Dell, T. R. (1973). Quantifying diameter distribution with the Weibull function. *Forest Science*, *19*, 97–104.

Borders, B. E., Souter, R. A., Bailey, R. L. & Ware, K. D. (1987). Percentile-based distributions characterize forest stand tables. *Forest Science*, *33*, 570–576.

Braastad, H. (1966). Volume tables for birch. *Communications of Norwegian Forest Research Institute*, *21*, 265–365. (In Norwegian with English summary.)

Brantseg, A. (1967). Volume functions and tables for Scots pine. South Norway. *Communications of Norwegian Forest Research Institute*, *22*, 695–739. (In Norwegian with English summary.)

Brokaw, N. V. L. & Lent, R. A. (1999). Vertical structure. In M. L. Hunter Jr (Ed.), *Maintaining biodiversity in forest ecosystems* (pp. 373–399). Cambridge: Cambridge University Press.

Gini, C. (1912). Variabilita e mutabilita: Contributo allo studio delle distribuzioni e relazioni stastistiche. *Studi Economico–Giuridici dell' Universita di Cagliari*, *3*, 1–158.

Gobakken, T. & Næsset, E. (2005). Weibull and percentile models for lidar-based estimation of basal area distribution. *Scandinavian Journal of Forest Research*, *20*, 490–502.

Hodgson, M. & Bresnahan, P. (2004). Accuracy of airborne lidar-derived elevation: Empirical assessment and error budget. *Photogrammetric Engineering and Remote Sensing*, *70*, 331–339.

Holmgren, J. (2003). Estimation of forest variables using airborne laser scanning. Umeå: Department of Forest Resource Management and Geomatics, Swedish University of Agricultural Sciences. *Acta Universitatis Agriculturae Sueciae, Silvestria*, *278*.

Kangas, A. & Maltamo, M. (2000*a*). Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. *Silva Fennica*, *34*, 371–380.

Kangas, A. & Maltamo, M. (2000*b*). Performance of percentile based diameter distribution prediction and Weibull method in independent data sets. *Silva Fennica*, *34*, 381–398.

Kraus, K. & Pfeifer, N. (1998). Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *53*, 193–203.

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G. & Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, *67*, 83–98.

Lexerød, N. L. & Eid, T. (2006). An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management*, *222*, 17–28.

Lim, K., Treitz, P., Baldwin, K., Morrison, I. & Green, J. (2003). Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, *29*, 658–678.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Magnussen, S., Eggermont, P. & LaRiccia, V. N. (1999). Recovering tree heights from airborne laser scanner data. *Forest Science*, *45*, 407–422.

Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T. & Saramäki, J. (2000). Comparison of percentile based prediction methods and Weibull distribution in describing diameter distribution of heterogeneous Scots pine stands. *Forest Ecology and Management*, *133*, 263–274.

Maltamo, M., Eerikäinen, K., Pitkäinen, J., Hyyppä, J. & Vehmas, M. (2004). Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment*, *90*, 319–330.

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J. & Pitkänen, J. (2005). Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*, *216*, 41–50.

Martens, H. (2001). Reliable and relevant modelling of real world data: A personal account of the development of PLS Regression. *Chemometrics and Intelligent Laboratory Systems*, *58*, 85–95.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., et al. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sensing of Environment*, *67*, 298–308.

Means, J. E., Acker, S. A., Brandon, J. F., Renslow, M., Emerson, L. & Hendrix, C. J. (2000). Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, *66*, 1367–1371.

Næsset, E. (1997). Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *52*, 49–56.

Næsset, E. (2001). Effects of differential single- and dual-frequency GPS and GLONASS observations on point accuracy under forest canopies. *Photogrammetric Engineering and Remote Sensing*, *67*, 1021–1027.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, *80*, 88–99.

Næsset, E. (2004*a*). Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment*, *91*, 243–255.

Næsset, E. (2004*b*). Practical large-scale forest stand inventory using small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, *19*, 164–179.

Næsset, E., Bollandsås, O. M. & Gobakken, T. (2005). Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment*, *94*, 541–554.

Nelson, R. (1997). Modeling forest canopy heights: The effects of canopy shape. *Remote Sensing of Environment*, *60*, 327–334.

Nelson, R., Krabill, W. & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, *24*, 247–267.

Nelson, R., Oderwald, R. & Gregoire, G. (1997). Separating the ground and airborne laser sampling phases to estimate tropical forest basal area, volume, and biomass. *Remote Sensing of Environment*, *60*, 311–326.

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, *56*, 1–7.

Nørgaard, L., Saudland, A., Wagner, J., Pram Nielsen, J., Munck, L. & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy*, *54*, 413–419.

Reutebuch, S. E., McGaughey, R. L., Andersen, H.-E. & Carson, W. W. (2003). Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing*, *29*, 527–535.

Reynolds, M. R., Jr, Burk, T. E. & Huang, W.-C. (1988). Goodness-of-fit tests and model selection procedures for diameter distribution models. *Forest Science*, *34*, 373–399.

Solberg, S. (1999). Crown density changes of Norway spruce and the influence from increased age on permanent monitoring plots in Norway during 1988–97. *European Journal of Forest Pathology*, *29*, 219–230.

Tobias, R. D. (1995). An introduction to partial least squares regression. *Proceedings of the Twentieth Annual SAS Users' Group International Conference* (pp. 1250–1257). Cary, NC: SAS Institute.

Tveite, B. (1977). Site index curves for Norway spruce (*Picea abies* (L.) Karst.). *Report of Norwegian Forest Research Institute*, *33*, 1–84. (In Norwegian with English summary.)

Uuttera, J. & Maltamo, M. (1995). Impact of regeneration method on stand structure prior to first thinning. Comparative study North Karelia, Finland vs. Republic of Karelia, Russian Federation. *Silva Fennica*, *29*, 267–285.

Vestjordet, E. (1967). Functions and tables for volume of standing trees. Norway spruce. *Communications of Norwegian Forest Research Institute*, *22*, 543–574. (In Norwegian with English summary.)

Weiner, J. & Solbrig, O. T. (1984). The meaning of measurement of size hierarchies in plant populations. *Oecologica*, *61*, 334–336.

Westad, F. O. (1999). *Relevance and parsimony in multivariate modelling*. Dr Ing. Thesis 129, Norwegian University of Science and Technology, Trondheim.

Wold, S. (1995). PLS for multivariate linear modeling. In H. van de Waterbeemd (Ed.), *Chemometric methods in molecular design* (pp. 195–218). Weinheim: VCH.

Wold, S., Martens, H. & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. In A. Ruhe, & B. Kågström (Eds.), *Matrix pencils* (pp. 286–293). Heidelberg: Springer.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348–368.

# Paper II

Available online at www.sciencedirect.com
ScienceDirect
Forest Ecology and Management xxx (2007) xxx–xxx
Forest Ecology and Management
www.elsevier.com/locate/foreco

# Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest

Ole Martin Bollandsås [a,*], Kjersti Holt Hanssen [b], Solfrid Marthiniussen [a], Erik Næsset [a]

[a] *Norwegian University of Life Sciences, Department of Ecology and Natural Resource Management, P.O. Box 5003, N-1432 Ås, Norway*
[b] *Norwegian Forest and Landscape Institute, P.O. Box 115, N-1431 Ås, Norway*

Received 20 December 2006; received in revised form 20 September 2007; accepted 4 October 2007

---

## Abstract

The relationships between measures of forest structure as derived from airborne laser scanner data and the variation in quantity (Q) and vitality (V) of young trees in a size-diverse spruce forest were analyzed. A regeneration success rate (Q), leader length (V), relative leader length (V), and apical dominance ratio (V) were regressed against 27 different laser-derived explanatory variables representing three different spatial scales. The resulting 81 different models for each response variable were ranked according to their Akaike information criterion score and significance level. Each laser variable was then associated with four categories. These were scale, return, fraction, and type. Within the scale category, laser variables were grouped according to the spatial scale from which they originated. Similarly, within the return, fraction, and type categories, the variables were grouped according to if they originated from first or last return echoes; if they originated from lower, middle, or upper fraction of the range of laser heights or values derived from the full range of laser pulses, and if they were canopy height or canopy density metrics. The results show that the laser variables were strongest correlated with the quantity of small trees and that these variables could be attributed to large-scale, last return, lower fraction, and density metrics. The correlations with the vitality responses were weaker, but the results indicate that variables derived from a smaller scale than for the quantity were better in order to explain variation in leader length, relative leader length, and apical dominance ratio.

*Keywords:* Regeneration; Uneven-aged forest; Laser scanner data; Canopy structure

---

## 1. Introduction

Recent years increasing focus on environmental issues related to commercial utilization of forest resources has resulted in various certification systems whose goal is to conserve biodiversity. One of the most focused issues in this respect has been the negative effects on habitat and species diversity caused by clearfellings and the establishment of monocultures. Thus, various incentives and regulations are introduced to favor uneven-aged, multi-species forests, and the forest area comprising even-aged structures is consequently most likely to decrease. Continuous cover forestry (CCF) is a silvicultural system designed to maintain an uneven-aged forest structure. The main idea of CCF is to maintain a continuous canopy cover so that the forest floor is not exposed and so that the disturbance of natural forest dynamics is kept at a minimum. Furthermore, CCF produces a wide range of tree sizes and a structurally diverse forest. This diversity is a result of harvesting schemes restricted to removal of single trees or small groups of trees. Whereas the regeneration after a clearfelling is based on planting, CCF usually relies on natural regeneration. In a well functioning uneven-aged forest, there is a continuous process of establishment and ingrowth. Thus, the future timber stock is already established under the dominating canopy at the time the cutting is performed.

From a forest planning point of view, heterogeneous forest structures are more challenging with respect to both the inventory and the projections of growth and yield. The basic unit needs to be at single-tree or diameter class level because mean forest characteristics at stand level do not give a sufficient description of an uneven-aged forest, nor sufficient input for projections. Computer simulators based on single-tree models are now being developed in many countries. In the US, the first single-tree model was developed as early as 1964 (Newham,

---

[*] Corresponding author. Tel.: +47 64 96 57 53; fax: +47 64 96 58 02.
*E-mail address:* olebo@umb.no (O.M. Bollandsås).

0378-1127/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.foreco.2007.10.017

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

1964) followed by others (e.g., Mitchell, 1969, 1975) through the 1960s and 1970s and have been in operational use since the 1980s (Pretzsch et al., 2002). In the recent years there has been focus on the single-tree approach also in Europe, and single-tree models and simulators are being developed (e.g., Pretzsch et al., 2002; Gobakken et al., 2005). Generally, such simulators rely on both single-tree and site information as input data, and utilizes single-tree models for growth, mortality (e.g., Monserud and Sterba, 1999; Eid and Tuhus, 2001), and recruitment. Recruitment models are models that predict the number of individual trees that exceed a certain threshold-size during a specific period. Furthermore, recruitment can be modeled in two different ways, either by one single equation (e.g., Moser, 1972; Adams and Ek, 1974; Vanclay, 1989) or by a two-stage approach (e.g., Vanclay, 1992; Lexerød, 2005). The concept of the two-stage approach is to first model the probability for recruitment to occur and then conditionally model the number of recruits given the result of the first stage. All other models implemented in a simulator are fitted for tree sizes exceeding this threshold and the recruitment models predict the trees added to the system during each projection period.

The quantity and vitality of the young growth in uneven-aged forest types is influenced by several factors. The first essential requirement for regeneration is a source of seeds. Second, the establishment of a seedling from a seed is dependent on the properties of the humus layer, competition from other plants, nutrient availability, and microclimate (moisture and light/heat) at the specific site. When establishment has been successful, growth depends largely on the availability of light and soil resources. Many of the factors important for both establishment and growth are influenced by stand structure. For instance, stand structure will affect below canopy light levels, which not only determine energy input but also influence temperature, the composition of the bottom and field layer species, humus layer processes and so on. Thus, under varying forest structure, the quantity and vitality of the young growth will be expected to vary accordingly.

Small footprint airborne laser scanning has shown to produce good data for reproducing forest structures. The laser depicts the canopy by transmissions of geo-referenced laser pulses, recording vegetation heights at the hit point of each pulse. Typical densities of these laser hits on the ground are between 1 and 5 pulses per square meter. Structural characteristics have been modeled from discrete laser returns by several authors (e.g., Maltamo et al., 2004; Parker and Russ, 2004; Tickle et al., 2006). One specific application in this respect is the retrieval of size (tree height and diameter) distributions of forest stands. For inventory purposes, these distributions are particularly interesting for uneven-aged forests because of the description aspect and that it is necessary as input in a single-tree simulator. Examples of such applications are presented by Maltamo et al. (2005) and Bollandsås and Næsset (2007). In addition to retrieval of size distributions the laser scanner is also utilized to efficiently derive stand-based estimates of leaf area index, tree heights, mean diameter, stem number, basal area, and timber volume (e.g., Næsset, 2002, 2004b; Solberg et al., 2006). Laser pulses can penetrate at least 40% of maximum canopy height (Næsset, 2004a) and account for much of the variation in canopy structure. However, the retrieval of small trees (say diameter less than 5 cm in breast height) under a dominating canopy by means of laser scanning is still challenging because laser data reflected from deep down in the vertical canopy structure both can be scarce and noisy. The data may be scarce because most laser pulses are reflected from the dominating canopy, and noisy because laser echoes from the lowest layer also are influenced by shrubs and stones (Nilsson, 1996). In this study we, therefore, wanted to use data from the dominating canopy layers in order to explain variations in the quantity and vitality of seedlings. Even though there are several factors that influence establishment and growth that are not, or only partly, affected by the stand structure, it is likely that there exist some relationship between the laser-depicted canopy and the variation in young growth. We believe that utilization of laser data describing canopy structure to detect young growth could be a valuable contribution for improving existing recruitment models or constructing new ones based solely on laser variables.

The objective of the present study was to analyze the relationship between measures of forest structure as derived from airborne laser scanner data and variation in the quantity and vitality of young trees in the height range of 0.1–3 m in a size-diverse spruce-dominated forest. The focus was on exploration and identification of laser-derived variables that have a potential for development of future prediction models that might be used in operational forest management. Furthermore, the study investigated potential scale effects on the relationship between the laser-derived variables and the quantity of young trees.

## 2. Materials and methods

### 2.1. Study area

The data originate from a boreal nature reserve in southeastern Norway (59°50′N, 11°02′E, 190–370 m a.s.l). The size of the reserve is approximately 1400 ha, and is comprised mainly by Norway spruce (*Picea abies* (L.) Karst.) and Scots pine (*Pinus sylvestris* L.). Within the reserve, no clear-cuttings have been executed, but some selective cuttings have been carried out before 1940 (Økland, 1994). However, the area in question is considered as primeval forest.

### 2.2. Field inventory

The data were collected on 72 circular field plots of 25 $m^2$ each. The plots were located in 18 clusters of four plots each. Stand characteristics are displayed in Table 1. The location of each cluster was subjectively determined in the field according to dominant tree species and age distribution. The criteria were that the plot should be spruce-dominated and have a multi-layered canopy. A center for each cluster was determined and a plot of 25 $m^2$ was located 12 m from this center in each cardinal direction using a tripod-mounted compass and a measuring tape

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

Table 1
Forest data by clusters

| Characteristic[a] | $n$ | Mean | S.T.D. | Range |
|---|---|---|---|---|
| $d_{\min}$ (cm) | 18 | 3.2 | 0.3 | 3.0–4.0 |
| $d_{\max}$ (cm) | 18 | 47.0 | 7.3 | 33.3–60.6 |
| $d_{\max}$–$d_{\min}$ (cm) | 18 | 43.8 | 7.3 | 30.3–57.1 |
| $d_g$ (cm) | 18 | 21.11 | 3.8 | 14.60–30.32 |
| $h_L$ (m) | 18 | 21.65 | 3.7 | 15.20–28.90 |
| $h_{\text{dom}}$ (m) | 18 | 26.27 | 3.3 | 19.80–32.00 |
| $N$ (ha$^{-1}$) | 18 | 1033 | 308 | 630–1780 |
| $G$ (m$^2$ ha$^{-1}$) | 18 | 34.2 | 5.7 | 21.6–45.5 |
| $V$ (m$^3$ ha$^{-1}$) | 18 | 360.7 | 110.4 | 171.9–634.8 |
| Tree species distribution (%) | | | | |
| Spruce | 18 | 90 | | 71–100 |
| Pine | 18 | 0 | | 0–2 |
| Deciduous | 18 | 10 | | 0–28 |

[a] $d_{\min}$: minimum diameter, $d_{\max}$: maximum diameter, $d_g$: mean diameter by basal area, $h_L$: Lorey's mean height, $N$: stem number, $G$: basal area, $V$: volume.

to determine the location. Each plot was split into four by two perpendicular lines through the plot center in a north/south and east/west direction. In each of these resulting 6.25 m$^2$ quadrants, the number of seedlings between 0.1 and 3 m were recorded. Apical dominance ratio (ADR), total height (TH), and absolute and relative leader length (LL and RLL) were recorded for the tallest seedling in each quadrant. LL was computed as the average of the leader lengths in the two last years, RLL as LL relative to TH, while ADR was computed as LL relative to the length of the longest lateral branch on the first node. Mean values of ADR, TH, and LL were computed for each plot to depress micro-scale variation in the occurrence and vitality of the seedlings. Furthermore, a regeneration success rate (RSR) was computed. First, the number of seedlings in each quadrant ($n_i$) less or equal to a sufficient number of each quadrant ($n_{\text{suf}}$), was registered. The sufficient number of seedlings on an area of 6.25 m$^2$ was defined to be three, which corresponds to 4800 seedlings ha$^{-1}$. The RSR was then computed as the sum of the registrations on each quadrant relative to the sufficient number of the entire plot ($4n_{\text{suf}}$) as displayed in Eq. (1).

$$\text{RSR} = \left[\frac{\sum_{i=1}^{4} n_i}{4n_{\text{suf}}}\right] \times 100 \tag{1}$$

where $n_{\text{suf}} = 3$, $n_i \leq n_{\text{suf}}$, $i$ = quadrant.

A summary of the data on plot level is displayed in Table 2.

### 2.3. Response variables

This study aims at explaining variation in both the quantity and vitality of seedling establishment in an uneven-aged spruce forest by means of laser scanner data. The quantity is in the current study represented by RSR, rather than the actual number. The reasons for this are that above a certain number, the establishment is most likely dependent on growth factor variations on a very small spatial scale, for instance the occurrence of partly decomposed downed logs or bare mineral soil, but also that RSR will be more representative of the number of seedlings needed for the regeneration to be successful. The vitality of the regeneration is represented by ADR, RLL, and LL. All these variables are in fact related to the leader length. The leader length is influenced by both light conditions and edaphic factors, and its derived variables may thus be seen as vitality indicators. Total height is strongly dependent on seedling age and was, therefore, not used as a vitality indicator.

### 2.4. Determination of plot coordinates

To ensure that the field data and the laser data were geographically co-registered, precise field coordinates were determined in the center of each cluster by means of differential Global Navigation Satellite Systems (dGNSS). A Topcon Legacy 20-channel dual-frequency receiver, observing pseudorange and carrier phase of Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) was used as rover equipment. The receiver setup had a 2-s logging rate, and all satellites below a 15° angle (cutoff angle) from the antenna were disregarded. The logging period for each cluster ranged between 0.5 and 1.5 h and the average antenna height was 4 m.

On the roof of a building within a distance of <2.5 km from the sample plots, an identical Topcon Legacy GPS + GLONASS receiver was established as base station. Næsset (2001) found that the accuracy of coordinates measured by means of dGNSS could be expressed by two times the standard deviation of the corrected single coordinate observations reported by the applied Pinnacle version 1.0 software package (Anon, 1999a) used for post-processing. According to this, planimetric coordinates of the base station had an accuracy of 0.4 cm. Records from the base station were used as reference during post-processing of the rover coordinates. To ensure that the base station received signals from the same satellites as the rover, the cutoff angle was set to 12°. The post-processing of all rover records was also undertaken by means of the Pinnacle software (Anon, 1999a). Following the same accuracy interpretation as for the base station coordinates, average accuracy of the cluster coordinates was 10 cm.

The field plot locations were determined relative to the cluster center by means of a compass attached to a tripod for angle and measuring tape for distance. However, determination

Table 2
Summary of data on plot level

| Variable[a] | $n$[b] | Mean | S.T.D. | Range |
|---|---|---|---|---|
| $N$ | 72 | 2.4 | 4.0 | 0–19 |
| RSR (%) | 72 | 14.6 | 19.4 | 0–83.3 |
| LL (cm) | 37 | 6.3 | 2.8 | 2.5–17.0 |
| ADR | 37 | 0.6 | 0.3 | 0.2–1.5 |
| TH (dm) | 37 | 12.4 | 8.0 | 1.8–30.0 |

[a] $N$: number of seedlings per plot, RSR: regeneration success rate, LL: mean leader length of the tallest seedlings by quadrants, ADR: mean apical dominance ratio by quadrants, TH: mean total height of the tallest seedlings by quadrants.
[b] For 35 plots there where no saplings to observe.

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

of directions relative to grid-north is dependent on the regional time-specific difference between absolute north and compass measured north. During the time of data collection, this regional difference was 0.2° for the study area. To also correct for local magnetic disturbance, a correction factor ($k$) was calculated for the study area based on 14 control measurements of the compass between 14 dGNSS-determined pairs of points spread throughout the study area. The correction factor $k$ (Eq. (2)) was calculated as the mean difference between the 14 point-pairs, weighted by the point-pairs' inter-distance relative to the precision of the point-pairs.

$$k = \frac{\sum_{i=1}^{14} \mathrm{OD}_i (D_i/\mathrm{SDD}_i)}{\sum_{i=1}^{14} D_i/\mathrm{SDD}_i} \qquad (2)$$

where $\mathrm{OD}_i$ is the observed difference between the compass-measured angle and the computed angle on the basis of the dGNSS measurements for point-pair $i$, $\mathrm{D}_i$ is the distance between point-pair $i$, and $\mathrm{SDD}_i$ is the mean standard deviation of the coordinates of point-pair $i$ as reported by the post-processing reports. The calculated $k$ was used to correct the measured angles and the field plot coordinates could then be calculated relative to the corresponding dGNSS determined cluster coordinate.

### 2.5. Laser scanner data

A Hughes 500 helicopter carried the ALTM 1233 laser scanning system produced by Optech, Canada. The laser scanner data were acquired in October 2003. The leaves were then still retained on the deciduous trees, but had a beginning xanthophyll coloring. The average flying altitude was approximately 600 m above the ground with an average speed of 35 $\mathrm{ms}^{-1}$. Twenty-one flight lines were flown with an overlap between adjacent stripes of about 20%. The pulse repetition frequency was 33 kHz and the scan frequency was 50 Hz. Maximum scan angle was 11°, which corresponded to an average swath width of about 230 m. Pulses transmitted at scan angles that exceeded 10.5° were excluded from the final dataset. The average footprint diameter for individual plots was approximately 18 cm. The mean number of pulses transmitted was 5.0 $\mathrm{m}^{-2}$. First and last returns echoes were recorded.

The initial processing of the laser data was accomplished by the contractor (Blom Geomatics, Norway). This processing includes computation of planimetric coordinates and ellipsoidal height values for all first and last returns. Furthermore, a matching between swaths was performed in order to remove orientation errors. The last return data were used to model the terrain surface. In a filtering operation undertaken by the contractor using the Terrascan software package (Anon, 2004), local maxima assumed to represent vegetation hits were discarded. A triangulated irregular network (TIN) was generated from the planimetric coordinates and corresponding height values of the individual terrain ground points retained in the last pulse data. The ellipsoidal height accuracy of the TIN model was expected to be around 20–30 cm (Kraus and Pfeifer, 1998; Reutebuch et al., 2003; Hodgson and Bresnahan, 2004).

Two different datasets were derived from the first and last pulse data for further analysis. All first and last return observations (points) were spatially registered to the TIN according to their coordinates. Terrain surface height values were computed for each point by linear interpolation from the TIN. The relative height of each point was computed as the difference between the height of the first or last return and the terrain surface height. Thus, the two datasets retained for analysis were geographically registered data for all transmitted pulses that were classified as first and last returns, respectively.

These datasets were spatially registered to three separate concentric circles around the field plots. The radii of these three circles were 2.82, 5.64, and 8.46 m, respectively, which correspond to (1) the area of the field plot itself (25 $\mathrm{m}^2$), (2) 100 $\mathrm{m}^2$, and (3) the maximum possible area (225 $\mathrm{m}^2$) around each plot center without overlap between adjacent plots within cluster. Pulses that hit outside the radius of 8.46 m were excluded from further analysis.

### 2.6. Computations

First and last pulse height distributions were created for each circle ($r$ = 2.82, 5.64, 8.46 m) around each sample plot center from the laser echoes considered to be reflected from the tree canopy, i.e., echoes with height values of >3 m. The tree canopy threshold value of 3 m was set to correspond to the maximum height of trees belonging to the understorey. From these distributions a total of 27 variables were derived. Three percentiles of 10, 50, and 90% of maximum height characterized both first and last return laser heights. Accordingly, measures of canopy density were derived by dividing the range between the lowest laser canopy height (>3 m) and the maximum canopy height into four uniform fractions. Cumulative canopy densities, henceforth called density variables, were then computed as the proportions of first and last pulse laser hits between the lower limit of each fraction and maximum laser height to total number of pulses. The canopy density for the lower fraction is, therefore, computed as the number of laser hits having a height value larger than 3 m, proportional to the total number of pulses. Canopy density for the second fraction is computed as the number of pulses above the lower limit of the second fraction proportional to total number of pulses, and so on (Fig. 1). Moreover, maximum and mean height values, standard deviations and coefficients of variation, including the standard deviation for the density variable for the lower fraction between quadrants, were derived. Slope and aspect (topographic variables) were computed from the terrain surface height values. Further details are provided by Næsset (2004b).

### 2.7. Data analysis

Because the data originate from clustered plots, there exists spatial dependency between plots within clusters. Thus, data analysis was carried out by means of the PROC MIXED procedure of the SAS statistical software package (Anon, 1999b), estimating random coefficient models. Each variable

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

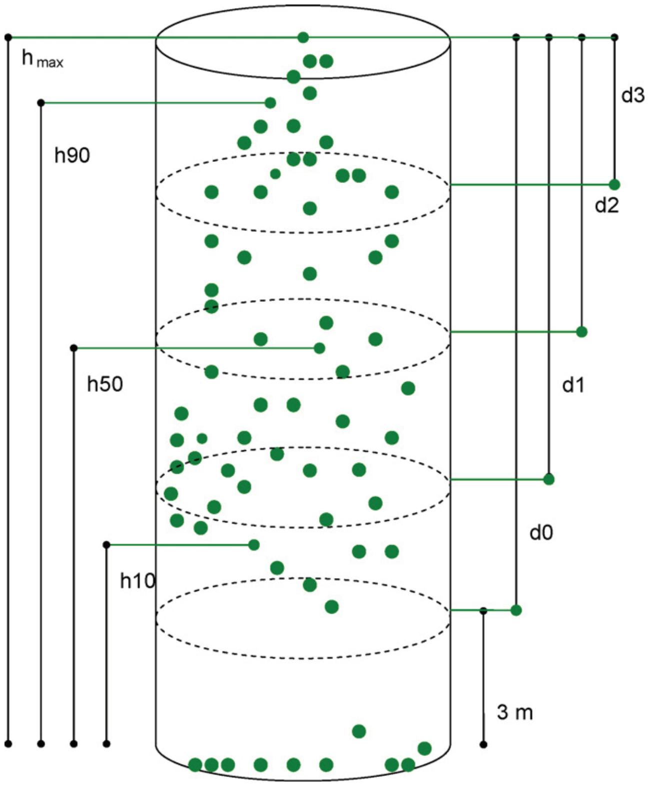

Fig. 1. Principle sketch for computation of the canopy density and percentile height variables. Dots represent laser hits. Ellipsoids (dotted lines) discriminate fractions. Vertical lines on the left indicate the height of the height variables. Vertical lines on the right indicate the range of laser hits from which the density variables are computed. The canopy density for each fraction was computed as the proportions of first and last pulse laser hits between the lower limit of each fraction and maximum laser height, to total number of pulses.

extracted from the laser data were regressed against each of the response variables. Subsequently, each of the models was ranked by their Akaike information criterion (AIC) (Akaike, 1974) score, separately for each response variable. This yielded a rank of each laser variable according to the goodness of fit of the model for each response variable. Similarly, each model was also ranked according to the significance level of the explanatory variable. The two different rankings were carried out to see if the top ranked variables were robust when the ranking criterion changed. Furthermore, each laser variable was attributed to four categories. These were scale, return, fraction, and type. Within the scale category, laser variables were grouped according to if the variables were extracted from the 25 m$^2$ circles (small), 100 m$^2$ circles (medium), or 225 m$^2$ circles (large). Similarly, within the return, fraction, and type categories, the variables were grouped according to if they originated from first or last return echoes; if they originated from lower, middle, or upper fraction of the range of laser heights or values derived from the full range of laser pulses, and if they were height or density metrics. Density variables from the lowest fraction (from 3 m to max canopy height) and the 10th percentile height variables constitute the lower fraction category. The middle fraction category includes the density variables from fraction number two and three, and the 50th percentile height variables. Finally, the upper fraction category then comprises the upper fraction density metrics and the 90th percentile height variables. Means and variation metrics for the height variables cannot be attributed to any of these fractions and are thus assigned to a full range category.

## 3. Results

Tables 3 and 5 display the results from the ranking of the laser variables with AIC and *p*-values, respectively. The tables show the modus group (most frequent group of variables within

Table 3
The most frequent group of variables (modus group) of the best 5, 10, and 15 category (superscript ‘a–d’) assigned, AIC-ranked, laser variables after modeling regeneration success rate (RSR), apical dominance ratio (ADR), relative leader length (RLL), and leader length (LL)

| Response | *n* | Scale[a] | | Return[b] | | Fraction[c] | | Type[d] | |
|---|---|---|---|---|---|---|---|---|---|
| | | Modus group | Freq. (%) | Modus group | Freq. (%) | Modus group | Freq. (%) | Modus group | Freq. (%) |
| RSR | 5 | Large | 80 | Last | 80 | Lower | 80 | Density | 100 |
| RSR | 10 | Large | 50 | Last | 70 | Middle | 60 | Density | 100 |
| RSR | 15 | Large | 53 | Last | 60 | Lower | 40 | Density | 100 |
| ADR | 5 | Small | 40 | Last | 80 | Middle | 60 | Density | 100 |
| ADR | 10 | Large | 40 | Last | 50 | Lower | 50 | Density | 90 |
| ADR | 15 | Medium | 40 | Last | 53 | Middle | 53 | Density | 93 |
| RLL | 5 | Small | 80 | Last | 60 | Middle | 60 | Density | 100 |
| RLL | 10 | Small | 60 | Last | 60 | Middle | 70 | Density | 100 |
| RLL | 15 | Small | 47 | Last | 67 | Middle | 60 | Density | 100 |
| LL | 5 | Medium | 80 | Last | 80 | Upper | 40 | Density | 100 |
| LL | 10 | Medium | 70 | Last | 80 | Middle | 50 | Density | 100 |
| LL | 15 | Medium | 60 | Last | 60 | Middle | 47 | Density | 100 |

[a] Area from which laser variables originate. Small: 25 m$^2$, medium: 100 m$^2$, large: 225 m$^2$.
[b] First or last laser echo.
[c] Fraction (upper, middle, lower or full range) of the range of laser heights from which laser variables originate.
[d] Type of laser variable (height, density, or topographic variable).

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

Table 4
The 5 highest ranked laser variables according to AIC value after modeling regeneration success rate (RSR), apical dominance ratio (ADR), relative leader length (RLL), and leader length (LL)

| Response | AIC rank | Laser variable[a] | Relationship | Scale[b] | Effect of cluster[c] | $p$-value |
|---|---|---|---|---|---|---|
| RSR | 1 | STDd02 | Positive | Large | Yes | 0.011 |
| RSR | 2 | d02 | Negative | Large | Yes | 0.001 |
| RSR | 3 | d02 | Negative | Medium | Yes | 0.004 |
| RSR | 4 | CVd02 | Positive | Large | Yes | 0.007 |
| RSR | 5 | STDd01 | Positive | Large | Yes | 0.167 |
| ADR | 1 | STDd02 | Negative | Small | Yes | 0.020 |
| ADR | 2 | d02 | Negative | Large | Yes | 0.191 |
| ADR | 3 | d02 | Negative | Medium | Yes | 0.218 |
| ADR | 4 | d12 | Negative | Large | Yes | 0.255 |
| ADR | 5 | d11 | Positive | Small | Yes | 0.302 |
| RLL | 1 | d32 | Positive | Small | Yes | 0.011 |
| RLL | 2 | d31 | Positive | Small | Yes | 0.070 |
| RLL | 3 | d22 | Positive | Small | Yes | 0.064 |
| RLL | 4 | STDd01 | Negative | Large | No | 0.359 |
| RLL | 5 | STDd02 | Negative | Small | No | 0.348 |
| LL | 1 | d32 | Negative | Medium | Yes | 0.685 |
| LL | 2 | d02 | Positive | Medium | Yes | 0.456 |
| LL | 3 | d31 | Negative | Medium | Yes | 0.768 |
| LL | 4 | d02 | Positive | Large | Yes | 0.484 |
| LL | 5 | d22 | Negative | Medium | Yes | 0.743 |

[a] Variable name convention: S.T.D.: standard deviation; CV: coefficient of variation; lower case d: density metrics; first number: fraction number across the range of laser heights (0: lower, 1and 2: middle, 3: upper, 4: full range); last number: first (1) or last (2) laser return echo.

[b] Area from which laser variables originate. Small = 25 $m^2$, medium = 100 $m^2$, large = 225 $m^2$.

[c] Indicates if the variable is correlated within clusters.

the category) with the corresponding frequency for all variable categories and responses. The best explanatory variables for RSR according to these rankings are attributed to large-scale, last return echo, lower fraction, and density metrics. Of the five highest AIC-ranked variables 80% is attributed to large-scale, 80% to last return echo, 80% to lower fraction, and 100% to density metrics for RSR. The modus groups are the same for the $p$-value ranking. For the other responses the table shows similar results except for that small or medium scale variables seem better for the responses representing vitality, ADR, RLL, and LL. Furthermore, the results for the vitality responses are not as unambiguous as for RSR. The frequencies are generally lower for these three responses.

Tables 4 and 6 show the five highest AIC and $p$-value ranked laser variables and their relationship to each response (positive or negative) and the corresponding significance level and also if

Table 5
The most frequent group of variables (modus group) of the best 5, 10, and 15 category (superscript ‘a–d’) assigned, significance-ranked, laser variables after modeling regeneration success rate (RSR), apical dominance ratio (ADR), relative leader length (RLL), and leader length (LL)

| Response | $n$ | Scale[a] | | Return [b] | | Fraction[c] | | Type[d] | |
|---|---|---|---|---|---|---|---|---|---|
| | | Modus group | Freq. (%) | Modus group | Freq. (%) | Modus group | Freq. (%) | Modus group | Freq. (%) |
| RSR | 5 | Large | 80 | Last | 80 | Lower | 60 | Density | 60 |
| RSR | 10 | Large | 70 | Last | 70 | Upper | 60 | Height | 60 |
| RSR | 15 | Large | 73 | Last | 67 | Upper | 40 | Height | 60 |
| ADR | 5 | Small | 80 | First | 80 | Middle | 40 | Density | 60 |
| ADR | 10 | Small | 50 | Last | 50 | Middle | 30 | Density | 60 |
| ADR | 15 | Small | 40 | Last | 60 | Middle | 40 | Density | 67 |
| RLL | 5 | Small | 100 | Last | 80 | Upper | 40 | Density | 80 |
| RLL | 10 | Small | 90 | Last | 80 | Middle | 30 | Density | 70 |
| RLL | 15 | Small | 87 | Last | 67 | Lower | 33 | Density | 60 |
| LL | 5 | Medium | 60 | Last | 60 | Middle | 80 | Density | 80 |
| LL | 10 | Small | 60 | First | 60 | Middle | 60 | Density | 90 |
| LL | 15 | Small | 47 | First | 60 | Middle | 40 | Density | 80 |

[a] Area from which laser variables originate. Small = 25 $m^2$, medium = 100 $m^2$, large = 225 $m^2$.

[b] First or last laser echo.

[c] Fraction (upper, middle, lower or full range) of the range of laser heights from which laser variables originate.

[d] Type of laser variable (height, density, or topographic variable).

Table 6
The 5 highest ranked laser variables according to *p*-value value after modeling regeneration success rate (RSR), apical dominance ratio (ADR), relative leader length (RLL), and leader length (LL)

| Response | *p*-value rank | Laser variable[a] | Relationship | Scale[b] | Effect of cluster[c] | *p*-value |
|---|---|---|---|---|---|---|
| RSR | 1 | d02 | Negative | Large | Yes | 0.001 |
| RSR | 2 | d02 | Negative | Medium | Yes | 0.004 |
| RSR | 3 | h32 | Positive | Large | Yes | 0.004 |
| RSR | 4 | h31 | Positive | Large | Yes | 0.006 |
| RSR | 5 | CVd02 | Positive | Large | Yes | 0.007 |
| ADR | 1 | STDd02 | Negative | Small | Yes | 0.020 |
| ADR | 2 | d31 | Positive | Small | No | 0.045 |
| ADR | 3 | CVh41 | Negative | Small | Yes | 0.082 |
| ADR | 4 | STDd01 | Negative | Small | Yes | 0.124 |
| ADR | 5 | CVh41 | Negative | Medium | Yes | 0.148 |
| RLL | 1 | d32 | Positive | Small | Yes | 0.011 |
| RLL | 2 | CVh42 | Negative | Small | No | 0.034 |
| RLL | 3 | d22 | Positive | Small | Yes | 0.064 |
| RLL | 4 | d31 | Positive | Small | Yes | 0.070 |
| RLL | 5 | d12 | Positive | Small | Yes | 0.082 |
| LL | 1 | STDd02 | Negative | Medium | No | 0.174 |
| LL | 2 | CVd01 | Negative | Small | No | 0.196 |
| LL | 3 | STDd01 | Negative | Small | Yes | 0.237 |
| LL | 4 | CVh42 | Positive | Medium | Yes | 0.306 |
| LL | 5 | CVd02 | Positive | Medium | No | 0.317 |

[a] Variable name convention: S.T.D.: standard deviation; CV: coefficient of variation; lower case d: density metrics; lower case h: height metrics; first number: fraction number across the range of laser heights (0: lower, 1and 2: middle, 3: upper, 4: full range); last number: first (1) or last (2) laser return echo.

[b] Area from which laser variables originate. Small: 25 m$^2$, medium: 100 m$^2$, large: 225 m$^2$.

[c] Indicates if the variable is correlated within clusters.

there was correlation within the cluster (effect of cluster). The most important result displayed in this table is the difference in significance level of the laser variables between response variables. All of the five highest ranking laser variables significantly explain some of the variation of the quantity response (RSR), whereas ADR and RLL have only one significant laser variable. No significant laser variable was found for LL (Table 6).

## 4. Discussion

### 4.1. Regeneration success rate

Establishment – measured as regeneration success rate – was positively correlated with large variation in density between quadrants, but negatively correlated with average density for the large and medium circles as a whole. This seems logical, as a lot of variation in density implies that there are dense spots, which may be created by tall trees which can be seed sources, mixed with open areas where a seedling would get enough light and lower competition for soil resources. In this natural forest, canopy openings are usually created by treefall gaps, which are good sites for regeneration (Kuuluvainen and Juntunen, 1998; Kuuluvainen and Kalmari, 2003). Lower density as an average for the whole plot also implies lower competition and an increased proportion of gaps. The best correspondence was found with density metrics and variation in density for the lower fraction, which value is affected by canopy variation from 3 m and up to the top of the canopy. Also some height variables gave significant *p*-values (Table 6), and the positive relationship between seedling number and height of the upper fraction (h31 and h32) suggests that tall trees are beneficial for regeneration. Tall trees indicate a possibility for high seed production, and also a mature stand where treefall gaps may occur.

Broadly speaking, the significant variables originated from the last return data. While the first return data describes the surface of the canopy, the last returns penetrate deeper into the canopy and thus account for more vertical canopy variation. Last return data are, therefore, a better representation of light conditions on the ground.

Establishment was linked to large-scale rather than small. Our ‘‘small-scale’’ ($r = 2.82$ m) seemed to be too small to give a meaningful description of forest structure in this context. For germination and early establishment of spruce seedlings, soil temperature and humidity are the most important factors (Mork, 1938; Bjor, 1971). Light levels affect both temperature and the distribution of bottom and field layer vegetation, which can be important for water availability. Even though the nearest neighbor trees may have a large influence on light conditions, light levels below the canopy will be affected by trees on a larger scale than 2.82 m radius in this mature forest. In fact, our ‘‘large-scale’’ with a radius of 8.46 m was not really very large, as light levels below the canopy are affected by trees or gaps up to at least twice the dominant stand height at northern latitudes (Flemming, 1962; Golser and Hasenauer, 1997). Our maximum radius was set to avoid overlap between adjacent plots, but in further studies the range of radii should be increased.

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

### 4.2. Seedling vitality

For seedling growth or vitality, measured as absolute or relative leader length as well as apical dominance, only a few laser data variables were significant. However, there were trends in the material. Growth was best correlated on a small to medium scale, density variables were higher ranked than height variables (as with establishment), and last return data were again better related than first return. After the germination and early establishment phase, growth is dependent on both light and soil resources. Competition for resources is affected by distance from and size of neighboring trees (i.e., Hegyi, 1974; Elliott and Vose, 1995). Near neighbors have greater effect than more distant ones. Small and medium scale variables will thus better account for competition than large-scale variables.

ADR is previously found to be a useful indicator of seedling vigor for Norway spruce and other shade tolerant species (Ruel et al., 2000; Duchesneau et al., 2001; Grassi and Giannini, 2005). Large ADR values indicate that seedlings experience good light conditions. With strong competition for light, the seedlings turn into an ‘‘umbrella-shape’’ with longer lateral branches and corresponding low apical dominance. ADR was negatively correlated to the canopy density variables as expected, but also negatively correlated to large density variation between quadrants at the smallest scale. Little is known about how the light values below the canopy are affected by such small-scale variation. Thus, it is difficult to have a strong prior expectation of this relationship.

Absolute leader length (LL) did not display any significant correlations with the laser variables. As LL is dependent on the size of the seedling or sapling, the explanation is probably that LL, like total height, is more dependent on seedling age than on stand structure as such. Thus, we expected RLL to be a better vitality indicator than LL. RLL was positively related to density in the upper canopy. This result is conflicting with our expectations, and must be attributed to the fact that our data contains a fairly narrow-ranged structure variation.

It was surprising that the relationship between forest structure represented by the laser variables and seedling quantity was stronger than between structure and seedling growth and vitality. Many studies have shown that spruce seedling growth is positively affected by decreased stand density (Nilson and Lundqvist, 2001; Nilsson et al., 2002; Hanssen et al., 2003) because this leads to increased light levels and reduced competition from the overstorey. Establishment, on the other hand, may be influenced by many stochastic factors, of which weather conditions are the most important, having a strong influence on seed production, germination, and seedling mortality. Also non-stochastic factors like soil conditions, ground vegetation or micro-topography may influence establishment, regardless of stand structure. Thus, we had expected to find better correlations between structure and seedling growth than with seedling quantity. However, our material concerning seedling vitality is rather small, as half of the plots did not have any seedlings or saplings at all. In addition, all plots are located in spruce forests with a fairly similar structure. For instance, basal area ranges from 22 to 46 $m^2\ ha^{-1}$, while the dominant height varies from 20 to 32 m, which indicates fully stocked and relatively mature forests in all plots. Thus, the plots may not contain enough variation in structure and density to find any significant effect on seedling growth. We need a larger material with more varied structure to draw conclusions on the relationship between structure and vitality.

The study was conducted in a multi-storeyed, natural spruce forest. In a managed spruce forest, the relationship between structure and regeneration may not be completely the same. One obvious difference may be the type and frequency of treefall gaps, which enhances regeneration by soil disturbance and woody debris and are important regeneration niches in a natural spruce forest (Kuuluvainen, 1994). Those elements are created mostly by the downfall of (over) mature trees and related to stand structure. In the managed forest trees are removed at an economic maturity age, and the presence of treefall gaps and downed logs are lower and not related to stand structure in a similar way as in natural forests. Thus, a separate study should be conducted for managed forests.

## 5. Conclusion

Our study is a screening which aims at identifying laser variables that explains regeneration success and seedling growth best, and at which scale the laser data should be retrieved. A full correlation between laser data variables derived from the canopy and regeneration will never be found, as factors not affected by canopy structure also strongly influence regeneration success. However, the study has shown that already existing data derived from laser scanning, for instance during a regular forest inventory, may give us surplus information on regeneration. Our data show that there is a relationship between canopy structure and seedling number, possibly strong enough for prediction of regeneration success in future prediction models. We may conclude from our results that the best condition for establishment of regeneration is a stand with a lot of variation in density, and with an overall sparse canopy. The laser data variables associated with the large-scale, last return, and lower fraction density data were the most significant ones. The correlations between regeneration growth/vitality and laser data variables were weaker than for seedling quantity, but on a modus group level the last return density variables on a small to medium scale were best correlated to seedling vitality. To draw conclusions here, we need a larger dataset from more varied forest than what we have been assessing.

## Acknowledgements

This research was founded by the Research Council of Norway (research project no. 153185/110). We wish to thank Blom Geomatics AS for collection and processing of the laser scanner data.

## References

Adams, D.M., Ek, A.R., 1974. Optimizing the management of uneven-aged forest stands. Can. J. For. Res. 4, 274–287.

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Auto. Control 19, 716–723.

Anon, 1999a. Pinnacle user's manual. Javad Positioning Systems, San Jose, CA.

Anon, 1999b. SAS OnlineDoc®, Version 8. SAS Institute Inc. Cary, NC.

Anon, 2004. TerraScan users guide. Terrasolid Ltd., Helsinki, Finland.

Bjor, K., 1971. Forest meteorological, soil climatological and germination investigations. Meddelelser fra Det norske Skogforsøksvesen 28, 429–526 (In Norwegian with English summary).

Bollandsås, O.M., Næsset, E., 2007. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. Scand. J. For. Res. 22, 33–47.

Duchesneau, R., Lesage, I., Messier, C., Morin, H., 2001. Effects of light and intraspecific competition on growth and crown morphology of two size classes of understorey balsam fir saplings. For. Ecol. Manage. 140, 215–225.

Eid, T., Tuhus, E., 2001. Models for individual tree mortality in Norway. For. Ecol. Manage. 154, 69–84.

Elliott, K.J., Vose, J.M., 1995. Evaluation of the competitive environment for White pine (*Pinus strobus* L) seedlings planted on prescribed burn sites in the Southern Appalachians. For. Sci. 41, 513–530.

Flemming, G., 1962. Strahlung und Wind an Bestandesrändern. Archiv für Forstwesen 11, 647–656 (in German).

Gobakken, T., Lexerød, N., Eid, T., 2005. A growth simulator based on models for individual trees. In: Hobbelstad, K. (Ed.), Proceedings of SNS meeting at Sjusjøen. Norway, September 6–8, 2004, NIJOS rapport 9/05, pp. 243–248.

Golser, M., Hasenauer, H., 1997. Predicting juvenile tree height growth in uneven-aged mixed species stands in Austria. For. Ecol. Manage. 97, 133–146.

Grassi, G., Giannini, R., 2005. Influence of light and competition on crown and shoot morphological parameters of Norway spruce and silver fir saplings. Ann. For. Sci. 62, 269–274.

Hanssen, K.H., Granhus, A., Brække, F.H., Haveraaen, O., 2003. Performance of sown and naturally regenerated *P. abies* seedlings under different scarification and harvesting regimes. Scand. J. For. Res. 18, 351–361.

Hegyi, F., 1974. A simulation model for managing jack-pine stands. In: Fries, J. (Ed.), Growth models for tree and stand simulation. International Union of Forestry Research Organizations working party S4.01-4: proceedings of meetings in 1973. Research Notes no. 30. Royal Coll. of For., Stockholm, Sweden.

Hodgson, M., Bresnahan, P., 2004. Accuracy of airborne lidar-derived elevation: empirical assessment and error budget. Photogramm. Eng. Remote Sens. 70, 331–339.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS J. Photogramm. Remote Sens. 53, 193–203.

Kuuluvainen, T., 1994. Gap disturbance, ground microtopography, and the regeneration dynamics of boreal coniferous forests in Finland—a review. Ann. Zool. Fennici 31, 35–51.

Kuuluvainen, T., Juntunen, P., 1998. Seedling establishment in relation to microhabitat variation in a windthrow gap in a boreal Pinus sylvestris forest. JVS 9, 551–562.

Kuuluvainen, T., Kalmari, R., 2003. Regeneration microsites of *Picea abies* seedlings in a windthrow area of a boreal old-growth forest in southern Finland. Ann. Bot. Fennici 40, 401–413.

Lexerød, N., 2005. Recruitment models for different tree species in Norway. For. Ecol. Manage. 206, 91–108.

Maltamo, M., Eerikäinen, K., Pitkäinen, J., Hyyppä, J., Vehmas, M., 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sens. Environ. 90, 319–331.

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., Pitkänen, J., 2005. Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. For. Ecol. Manage. 216, 41–50.

Mitchell, K.J., 1969. Simulation of the growth of evenaged stands of white spruce. School of Forestry Bulletin, No. 75. Yale University, New Haven.

Mitchell, K.J., 1975. Dynamics and simulated yield of Douglas-fir. For. Sci. Monogr. 17, 39.

Monserud, R.A., Sterba, H., 1999. Modeling individual tree mortality for Austrian forest species. For. Ecol. Manage. 113, 109–123.

Mork, E., 1938. Gran- og furufrøets spirning ved forskjellig temperatur og fuktighet. Meddelelser fra Det norske Skogforsøksvesen 20, 225–249 (in Norwegian).

Moser, J.W., 1972. Dynamics of uneven-aged forest stands. For. Sci. 18, 184–191.

Newham, R.M., 1964. The development of a stand model for Douglas-fir. Ph.D. Thesis. University of British Colombia, Vancouver.

Nilsson, M., 1996. Estimating tree heights and stand volume using an airborne lidar system. Remote Sens. Environ. 22, 1–7.

Nilson, K., Lundqvist, L., 2001. Effect of stand structure and density on development of natural regeneration in two *Picea abies* stands in Sweden. Scand. J. For. Res. 16, 253–259.

Nilsson, U., Gemmel, P., Johansson, U., Karlsson, M., Welander, T., 2002. Natural regeneration of Norway spruce, Scots pine and birch under Norway spruce shelterwoods of varying densities on a mesic-dry site in southern Sweden. For. Ecol. Manage. 161, 133–145.

Næsset, E., 2001. Effects of differential single- and dual-frequency GPS and GLONASS observations on point accuracy under forest canopies. Photogramm. Eng. Remote Sens. 67, 1021–1027.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88–99.

Næsset, E., 2004a. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. Remote Sens. Environ. 91, 243–255.

Næsset, E., 2004b. Practical large-scale forest stand inventory using small-footprint airborne scanning laser. Scand. J. For. Res. 19, 164–179.

Økland, B., 1994. Mycetophilidae (Diptera), an insect group vulnerable to forest practices? A comparison of clearcut, managed and semi-natural spruce forests in southeastern Norway. Biodivers. Conserv. 3, 68–85.

Parker, G.G., Russ, M.E., 2004. The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry. For. Ecol. Manage. 189, 307–315.

Pretzsch, H., Biber, P., Ďurský, J., 2002. The singletree-based stand simulator SILVA: construction, application and evaluation. For. Ecol. Manage. 162, 3–21.

Reutebuch, S.E., McGaughey, R.L., Andersen, H.-E., Carson, W.W., 2003. Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. Can. J. Remote Sens. 29, 527–535.

Ruel, J.C., Messier, C., Doucet, R., Claveau, Y., Comeau, P., 2000. Morphological indicators of growth response of coniferous advance regeneration to overstorey removal in the boreal forest. Forest. Chron. 76, 633–642.

Solberg, S., Næsset, E., Hanssen, K.H., Christiansen, E., 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. For. Ecol. Manage. 102, 364–376.

Tickle, P.K., Lee, A., Lucas, R.M., Austin, J., Witte, C., 2006. Quantifying Australian forest floristics and structure using small footprint LiDAR and large scale aerial photography. For. Ecol. Manage. 223, 379–394.

Vanclay, J.K., 1989. A growth model for north Queensland rainforests. For. Ecol. Manage. 98, 1–12.

Vanclay, J.K., 1992. Modelling regeneration and recruitment in tropical rainforest. Can. J. For. Res. 22, 1235–1248.

Please cite this article in press as: Bollandsås, O.M., et al., Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest, Forest Ecol. Manage. (2007), doi:10.1016/j.foreco.2007.10.017

# Paper III

# Height-diameter models for individual trees in Norway

Ole Martin Bollandsås*, Erik Næsset

Norwegian University of Life Science

Department of Ecology and Natural Resource Management

P.O. Box 5003, N-1432 Ås, Norway

* Phone: +47 64965753, e-mail: olebo@umb.no

## Abstract

Non-linear models for predicting tree height for most tree species in Norway were developed from national forest inventory data. Diameter, stand density, site productivity, altitude, and latitude explained 66% to 82% of the variation in tree height for the different tree species. No serious bias was detected. The random errors assessed by independent tests were 17.1 %, 19.0 %, 19.7 %, and 23.2 % of the observed mean tree height for Norway spruce, Scots pine, birch, and other broadleaves, respectively. A second set of models without stand density as an explanatory variable was also estimated because it can be inconvenient that tree heights are differently estimated just before and just after silvicultural operations that significantly affect density. The first set of models will therefore be best adapted to uneven-aged forestry with little variations in stand density. Diameter, site productivity, altitude, and latitude explained 63% to 79% of the variation in tree height for the different tree species for the second set of models. The random errors were approximately one percentage point larger than models that included stand density. The models provide an alternative to constructing height curves specific for stands or forests that might be resource demanding.

*Keywords:* Tree height; non-linear model; uneven-aged forest

## Introduction

Forest inventories in Norway have traditionally been stand oriented and computations of growth and yield based on mean characteristics such as mean height, mean diameter, stand basal area, stand age, number of stems, and site quality. This approach has been expedient since the majority of the economically important stands in Norway are even-aged and consequently well described by mean characteristics. However, uneven-aged, multi-species forests are recognized as important for preservation of biodiversity, and efforts are therefore made to change the forestry-practice towards selective cuttings. Consequently, the area covered by even-aged forest is most likely to decrease.

Mean characteristics are less relevant for uneven-aged stands, both for description and computations. Accurate computations of growth and yield of these forest types require information on tree- or diameter class level so that total stand volume can be calculated as the sum of individual tree volumes. Individual tree volumes are computed by means of equations (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967) where diameter and height are required as input variables. While diameter measurements can be obtained accurately at low cost, measurement of tree heights are both more resource demanding and less accurate. The advantage of being able to predict tree heights from diameter is then obvious. The most common practice of obtaining tree heights is to make stand-wise height-diameter equations from a sub-sample of height-measured trees. However, an accurate pre-estimated model valid for large areas would be preferable when the number of available sample trees is small, or conversely, if we want to reduce or omit height measurements and hence reduce cost.

Projections of future growth and yield by means of most single-tree based computer simulators require height-diameter models. The only height-diameter models developed on data from all of Norway until now were presented by Øyen and Andreassen (2002). However, these models are based on linear regressions that fail to account for the changing height-diameter relationship throughout the lifetime of a tree. Furthermore, the documentation of the validity of their models is only limited.

Construction of height-diameter models valid for large geographical areas is not straightforward. Within a stand, the relationship between diameter and height is sufficiently stable to allow diameter as the sole predictor. A height-diameter model for a large region or country will, however, require that conditions influencing the relationship between diameter and height be taken into account. This relationship depends on site index (Vanclay, 1994), local climate (Mäkinen, 1998), competition (Loetsch et al., 1973), and stand density (Fulton, 1999; Sharma and Zhang 2004).

Several authors have developed height-diameter models for large areas. Many of these models, however, include diameter as the only explanatory variable (Wykoff et al., 1982; Zhang, 1997; Lappi, 1997; Fang and Bailey, 1998; Peng, 1999), consequently assuming that the height-diameter relationship only depends on diameter. Others have included stand variables. Sharma and Zhang (2004) included stand basal area, number of stems, and site index. Larsen and Hann (1987) and Parresol (1992) included basal area. The model form varies, but mostly a non-linear function is fitted to represent the relationship between tree height and the explanatory variables.

Dorado et al. (2006) developed height-diameter models for northwestern Spain utilizing a mixed modeling approach to account for dependence between multiple measurements of the same plots and dependence between trees on the same plot. In the current study the plots are measured only once and the number of trees on each plot is generally low (average 2.3). The correlation between trees on the same plot is therefore minor and mixed modeling is hence not used.

The aim of this work was to develop species-specific predictive models of tree height valid for Norway by including stand variables in addition to diameter at breast height. Two sets of models were developed. One including stand basal area as an explanatory variable (Model set 1), and one without stand basal area (Model set 2).

Table 1. Summary of estimation data on stand level

| Variable | # of plots | Mean | SD | Range |
|---|---|---|---|---|
| Spruce | 2,675 | | | |
| *H40* (m) | | 12.8 | 4.3 | 6.0 - 26.0 |
| BA ($m^2ha^{-1}$) | | 23.6 | 12.1 | 0.3 - 91.6 |
| N ($ha^{-1}$) | | 1,310 | 722 | 40 - 5,080 |
| LAT[a] | | 61.2 | 1.9 | 58.0 - 69.5 |
| | | | | |
| Pine | 2,358 | | | |
| *H40* (m) | | 10.5 | 3.5 | 6.0 - 23.0 |
| BA ($m^2ha^{-1}$) | | 20.1 | 10.4 | 0.6 - 73.0 |
| N ($ha^{-1}$) | | 955 | 609 | 40 - 4,520 |
| LAT[a] | | 60.8 | 2.0 | 58.0 - 70.0 |
| | | | | |
| Birch | 1,824 | | | |
| *H40* (m) | | 10.7 | 3.7 | 6.0 - 23.0 |
| BA ($m^2ha^{-1}$) | | 18.3 | 10.0 | 0.6 - 73.0 |
| N ($ha^{-1}$) | | 1,362 | 766 | 80 - 6,080 |
| LAT[a] | | 62.8 | 3.5 | 58.0 - 70.5 |
| | | | | |
| Other | 677 | | | |
| *H40* (m) | | 13.1 | 4.2 | 6.0 - 26.0 |
| BA ($m^2ha^{-1}$) | | 24.0 | 11.9 | 0.2 - 70.3 |
| N ($ha^{-1}$) | | 1,589 | 892 | 40 - 6,520 |
| LAT[a] | | 61.5 | 3.2 | 58.0 - 69.9 |

[a] Degrees and minutes north according to Euref89. Minutes scaled by 100/60 (60˚ 30'N=60.5˚N).

## Material and methods

### *Estimation data*

Data used for model estimation was supplied by the Norwegian National Forest Inventory (NFI). All data originate from permanent 250 m$^2$ circular plots (radius=8.92 m), organized in a 3 × 3 km grid, yielding a representative sample of the Norwegian forest area. The NFI data in question was collected in the time period between 1999 and 2003. The data provide information on both stand- and tree characteristics, and summaries of essential variables are displayed in tables 1 and 2.

Table 2. Summary of estimation data on tree level

| Variable | # of trees | Mean | SD | Range |
|---|---|---|---|---|
| Spruce | 7,603 | | | |
| *DBH* (mm) | | 222 | 97.7 | 50 - 712 |
| *h* (dm) | | 154 | 54.3 | 26 - 329 |
| | | | | |
| Pine | 5,529 | | | |
| *DBH* (mm) | | 267 | 99.1 | 50 - 805 |
| *h* (dm) | | 148 | 43.9 | 33 - 312 |
| | | | | |
| Birch | 3,418 | | | |
| *DBH* (mm) | | 153 | 77.0 | 50 - 655 |
| *h* (dm) | | 106 | 38.1 | 32 - 288 |
| | | | | |
| Other | 1,515 | | | |
| *DBH* (mm) | | 181 | 99.1 | 50 - 999 |
| *h* (dm) | | 123 | 46.4 | 37 - 275 |

All stand variables are calculated from measurements inside the plot perimeter except site index (*H40*). *H40* was registered just outside the plot to avoid destructive sampling due to that a coring have to be performed to determine age. The *H40* variable is defined by average age at breast height and the average height of the 100 largest trees per hectare according to diameter at breast height. The specific *H40* value relate to the dominant height at an index age of 40 years (Tveite, 1977; Braastad, 1980).

Tree diameter (*DBH*) of 5 cm at breast-height (1.3 m) defined the lower limit of the sample size-range. Sample trees were selected with a probability proportional to stem basal area, and heights of these sample trees were measured by means of either a Suunto clinometer or a Vertex hypsometer. According to field instructions, tree height is defined from the stump-height to the top.

Species was registered for all callipered trees. In the present work, separate models were developed for Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), and birch (*Betula pubescens* and *Betula pendula*). For the remaining deciduous species one model independent of species was developed. These species were in alphabetical order: alder (*Alnus* spp.), ash (*Fraxinus excelsior*), aspen (*Populus tremula*), beech (*Fagus silvatica*), Bird cherry (*Prunus padus*), linden (*Tilia cordata*), marple (*Acer* spp.), oak (*Quercus* spp.), rowan (*Sorbus aucuparia*), sallow (*Salix caprea*), and Wych elm (*Ulmus glabra*). The main reason for this is that the number of observations for some of these species was scarce, so that species-specific models would have a poor basis. Furthermore, practical forest planning in Norway makes no distinction between these species because they occur relatively seldom. Both reasons also apply for the two birch species.

### *Data reduction*

The initial estimation data comprised near 8,600 plots. Plots intercepted by roads, water, agricultural land etc, were excluded to avoid edge effects. Furthermore, plots classified as unproductive, i.e. have a volume increment less than 1 $m^3ha^{-1}yr^{-1}$ were excluded because the *H40* site index system is not defined below this limit. Trees with reduced height as a result of breakage were excluded from further analysis to eliminate a source of systematic errors. Trees with dead top were excluded for the same reason. Trees where a new top was established after a break were kept. All conifer species not identified as Norway spruce or Scots pine, were excluded from further analysis. After data reduction 7,299 plots (16,960 trees) remained for further analysis.

Table 3. Summary of test data on stand level

| Variable | # of plots | Mean | SD | Range |
|---|---|---|---|---|
| Spruce | 877 | | | |
| *H40* (m) | | 12.3 | 4.2 | 6.0 - 23.0 |
| *BA* (m$^2$ha$^{-1}$) | | 18.8 | 12.9 | 0.1 - 64.3 |
| *N* (ha$^{-1}$) | | 1,111 | 721 | 40 - 5,080 |
| *LAT*[a] | | 61.2 | 2.0 | 58.0 - 69.6 |
| Pine | 650 | | | |
| *H40* (m) | | 10.5 | 3.5 | 6.0 - 20.0 |
| *BA* (m$^2$ha$^{-1}$) | | 17.5 | 11.0 | 0.1 - 60.1 |
| *N* (ha$^{-1}$) | | 924 | 660 | 40 - 4,440 |
| *LAT*[a] | | 61.1 | 2.2 | 58.0 - 69.9 |
| Birch | 862 | | | |
| *H40* (m) | | 10.4 | 3.6 | 6.0 - 23.0 |
| *BA* (m$^2$ha$^{-1}$) | | 14.7 | 9.9 | 0.1 - 61.4 |
| *N* (ha$^{-1}$) | | 1,059 | 695 | 40 - 5,080 |
| *LAT*[a] | | 62.7 | 3.4 | 58.0 - 70.5 |
| Other | 346 | | | |
| *H40* (m) | | 12.3 | 4.2 | 6.0 - 23.0 |
| *BA* (m$^2$ha$^{-1}$) | | 18.0 | 12.4 | 0.3 - 61.4 |
| *N* (ha$^{-1}$) | | 1,191 | 775 | 40 - 3,920 |
| *LAT*[a] | | 62.2 | 3.5 | 58.0 - 70.0 |

[a] Degrees and minutes north according to Euref89. Minutes scaled by 100/60 (60° 30'N=60.5°N).

### *Test data*

Like the estimation data, the test data originated from the national forest inventory in Norway. However, the majority of the test-trees were not included in the estimation data. All test-trees were measured for height during 2005 as a part of the first year of an extended height measurement campaign where, if possible, ten relascope selected trees per plot are measured for height. This practice of extended height measurements is conducted to improve volume and increment estimates from the inventory. Because of these additional height measurements, additional plots previously without height measurements, and hence not a part of the estimation data, were included in the test data. Summaries of the test data on plot and tree level are displayed in tables 3 and 4, respectively. Comparing them to tables 1 and 2, they show that the forest conditions do not range as wide as the estimation data (Tables 1 and 3), and the trees are on average smaller (Tables 2 and 4). The larger number of smaller trees in the test data is a consequence of the requirement of ten sample trees per plot which is obtained by adapting the relascope factor on each plot.

Table 4. Summary of test data on tree level

| Variable | # of trees | Mean | SD | Range |
|---|---|---|---|---|
| Spruce | 5,040 | | | |
| *DBH* (mm) | | 170 | 90.0 | 50 - 632 |
| *h* (dm) | | 120 | 56.7 | 23 - 328 |
| Pine | 3,117 | | | |
| *DBH* (mm) | | 209 | 98.6 | 50 - 612 |
| *h* (dm) | | 123 | 48.0 | 15 - 274 |
| Birch | 3,845 | | | |
| *DBH* (mm) | | 122 | 58.4 | 50 - 565 |
| *h* (dm) | | 88 | 32.6 | 26 - 273 |
| Other | 1,201 | | | |
| *DBH* (mm) | | 136 | 83.3 | 50 - 655 |
| *h* (dm) | | 100 | 44.6 | 22 - 297 |

### *Model form and statistical methods*

All models are fitted using the NLIN procedure of the statistical software SAS (Anon, 1999). The iterative method used was the Gauss-Newton. The method iteratively finds the parameter estimates that minimize the residual sum of squares (SSR). To ensure that the Gauss-Newton method converges at a global SSR-minimum, several starting points for the iterative process were set. Several preliminary model estimations were performed to find plausible values and ranges of these starting points.

The height growth of a tree typically follows an asymptotic pattern in relation to diameter (Ryan and Yoder, 1997). More specifically this means that the diameter growth is more durable than the height growth as height reaches its near maximum at an earlier age relative to diameter. In the present study it was assumed that external factors (stand variables) determine this maximum height. Many non-linear mathematical functional forms are suitable to describe such asymptotic relationships. Most frequently used are the Chapman-Richards function (Richards, 1959; Chapman, 1961), the Weibull (Yang et al., 1978), and the Korf/Lundqvist function (Stage, 1963; Zeide, 1989). In addition to these, different exponential and logistic functions are frequently used. An overview and tests of different model forms are presented by Peng et al. (2001). In the present work the model form used was a version of the Korf/Lundqvist function, given by:

$$h = 13 + \left(\beta 1 \times SI^{\beta 2} \times LAT^{\beta 3} \times ALT^{\beta 4}\right) \times \exp\left(\beta 5 \times DBH^{\beta 6} \times BA^{\beta 7}\right) \qquad (1)$$

where the βs are parameters to be estimated, *LAT* is latitude, *ALT* is altitude, and *BA* is stand basal area per hectare. In this model form, the first part of the expression represents the asymptotic value. 13 (dm) is added to force the tree height trough 13 dm when the diameter in breast height (1.3 m) approaches zero. The exponential part of the expression approaches 1 as *DBH* approaches infinity. A second set of models without stand density as explanatory variable was also estimated because the inclusion of *BA* may be inconvenient as the height of a tree will be differently estimated just before and just after silvicultural operations that significantly affect density. The model form for the second set of models equals the form presented in eq. (1), without *BA*.

During the estimation of the models, the residual standard error and the pseudo-coefficient of multiple determination (pseudo-$R^2$) have been considered to choose between candidate models. The pseudo-$R^2$ was computed as 1 minus the ratio between sum of residual squares (SSR) and the corrected total sum of squares (CSST) as presented in equation 2.

$$Pseudo - R^2 = 1 - \left( \frac{SSR}{CSST} \right) \tag{2}$$

The models were tested both by self-validation and height prediction on an independent data. The self-validation was performed at the same dataset as the model estimation, hence self-validation. Bias and random error estimates were calculated as the mean differences between predicted and observed values and corresponding standard deviations (SD), respectively. The main appraisal of the model adequacy is based on thresholds of prediction errors as discussed by Huang et al. (2003). Simply, these guidelines say that a mean prediction error of <±10% at 95% confidence level is acceptable if the prediction errors do not show adverse patterns. A prediction error between ±10% and ±20% indicate a level of uncertainty that calls for additional data and testing. Two-tailed *t*-tests were performed to assess the statistical significance of the difference between predicted and observed value. When these tests were performed simultaneously for *k* subgroups according to diameter, site quality, latitude, altitude, and basal area, Bonferroni *t*-tests (Miller, 1981) were used in order to control the type I error. Thus, the level of significance for each of the *k* test was $\alpha/2k$.

### *Variable selection*

The maximum or asymptotic height of a specific tree species depends on many factors. Most of these factors are related to stand conditions. The current study has considered *LAT*, *ALT*, and *H40* to account for the variation of maximum tree height based on the "hydraulic limitation hypothesis" (Friend, 1993; Ryan and Yoder, 1997; Koch et al., 2004). In brief, the hydraulic limitation hypothesis argues that trees reach a certain maximum height because the increased hydraulic resistance in tall trees causes the stomata (pores) of the leaves to close earlier in the day to prevent a collapse in the xylem function. Such a collapse occurs as a result of discontinuity in the xylem water-column because of high tension (Tyree and Sperry, 1988; Tyree and Zimmermann, 2002). The early closing of the stomata reduces evaporation and hence the ability to pull water from the soil. The hydraulic resistance is of course directly related to the tree height it self, but it is also related to the contents of nutrients and water in the soil. On nutrient-poor sites, trees form xylem with narrow tracheids with a consequential low permeability. The *H40* is therefore a logical explanatory variable, explaining that tall trees require increased water and nutrient availability to overcome gravity and resistance of the xylem. Both *LAT* and *ALT* represent climatic factors (temperature) limiting the maximum height of a tree, and are considered important to make the model applicable to the whole country. The increased viscosity of soil water on cold sites causes the hydraulic resistance to increase (Ryan and Yoder, 1997). Consequently, as the values of *LAT* and *ALT* increase, the maximum tree height is assumed to decrease, whereas increasing *H40* should allow for increasing maximum height. Furthermore, the height-diameter relationship and hence the shape of the height curve, is dependent on social status and stand density. This means that a tree with a given diameter will have a different height depending on the influence of neighboring trees. These effects can be accounted for by including stand basal area (*BA*) or a competition index. The competition index *BAL* (Wycoff, 1990) was considered in the present work. The value of *BAL* for a specific tree is the sum of tree basal areas of larger trees pre hectare. *BA* was also included in the models developed by Sharma and Zhang (2004), but contrary to the present work they include *BA* in the asymptotic part of their models.

## Results

### *The models*

Table 5 shows the estimated model parameters and the corresponding model adequacy diagnostics. All models include *H40* and *LAT* in the asymptotic part of the model. *ALT* was, however, only included in the models for Norway spruce and birch. In addition to *DBH*, *BA* was included in the shape part of Model set 1. *BAL* was not included because it created a strong trend in the prediction error. The pseudo-$R^2 \geq 0.63$ indicate a good fit for all models, and the random error was 23.2, 25.5, 19.0, and 26.7 dm for Norway spruce, Scots pine, birch, and other broadleaves, respectively for Model set 1. The random errors were approximately one percentage point higher for Model set 2.

Table 5. Parameter estimates[a] and model adequacy diagnostics for species specific diameter-height models. Model 1 includes stand basal area as explanatory variable, Model 2 does not.

| | | Species | | | |
|---|---|---|---|---|---|
| Parameter | Variable | Spruce | Pine | Birch | Other |
| | | *Model set 1. Including BA as explanatory variable* | | | |
| $\beta1$ | | 31,569 | 9,655.0 | 830.30 | 101,572 |
| $\beta2$ | $H40$ | 0.1543 | 0.3994 | 0.3854 | 0.3023 |
| $\beta3$ | $LAT$ | –1.0825 | –1.1371 | –0.4165 | –1.5432 |
| $\beta4$ | $ALT$ | –0.0221 | - | –0.0220 | - |
| $\beta5$ | | –33.332 | –105.50 | –14.821 | –15.455 |
| $\beta6$ | $DBH$ | –0.5436 | –0.8410 | –0.4312 | –0.4166 |
| $\beta7$ | $BA$ | –0.1233 | –0.2193 | –0.1230 | –0.1207 |
| Pseudo-$R^2$ | | 0.82 | 0.66 | 0.75 | 0.67 |
| SD (dm) | | 23.2 | 25.5 | 19.0 | 26.7 |
| BIAS (dm) | | –0.08ns | –0.12ns | –0.08ns | –0.10ns |
| Observations | | 7,601 | 5,519 | 3,381 | 1,486 |
| | | *Model set 2. Excluding BA as explanatory variable* | | | |
| $\beta1$ | | 17,142 | 17,125 | 1,605.5 | 131,104 |
| $\beta2$ | $H40$ | 0.2491 | 0.4604 | 0.4606 | 0.3723 |
| $\beta3$ | $LAT$ | –1.0428 | –1.3280 | –0.6737 | –1.6993 |
| $\beta4$ | $ALT$ | –0.0149 | - | –0.0224 | - |
| $\beta5$ | | –37.565 | –123.60 | –16.347 | –14.833 |
| $\beta6$ | $DBH$ | –0.6693 | –1.0175 | –0.5640 | –0.5207 |
| Pseudo-$R^2$ | | 0.79 | 0.63 | 0.72 | 0.65 |
| SD (dm) | | 24.9 | 26.6 | 20.3 | 27.7 |
| BIAS (dm) | | 0.03ns | 0.06 | 0.05 | 0.01 |
| Observations | | 7,601 | 5,519 | 3,381 | 1,486 |

[a] All parameter estimates are significant ($p<0.05$).

All models show different patterns in terms of shape of the height curve. Figure 1 displays all models of Model set 1 for average forest conditions i.e. $H40=17$, $LAT=63$, $ALT=50$, and $BA=25$. The figure shows that the models for the deciduous species predict higher trees for small diameters but that spruce ultimately will be superior in height. Figure 2 shows that the model for Norway spruce (Model set 1) is consistent with prior expectations considering the stand variables included in the models. The figure displays curves of predicted height values for different values of $BA$ (a), $LAT$ (b), $H40$ (c), and $ALT$ (d). The models for the other species show similar patterns.

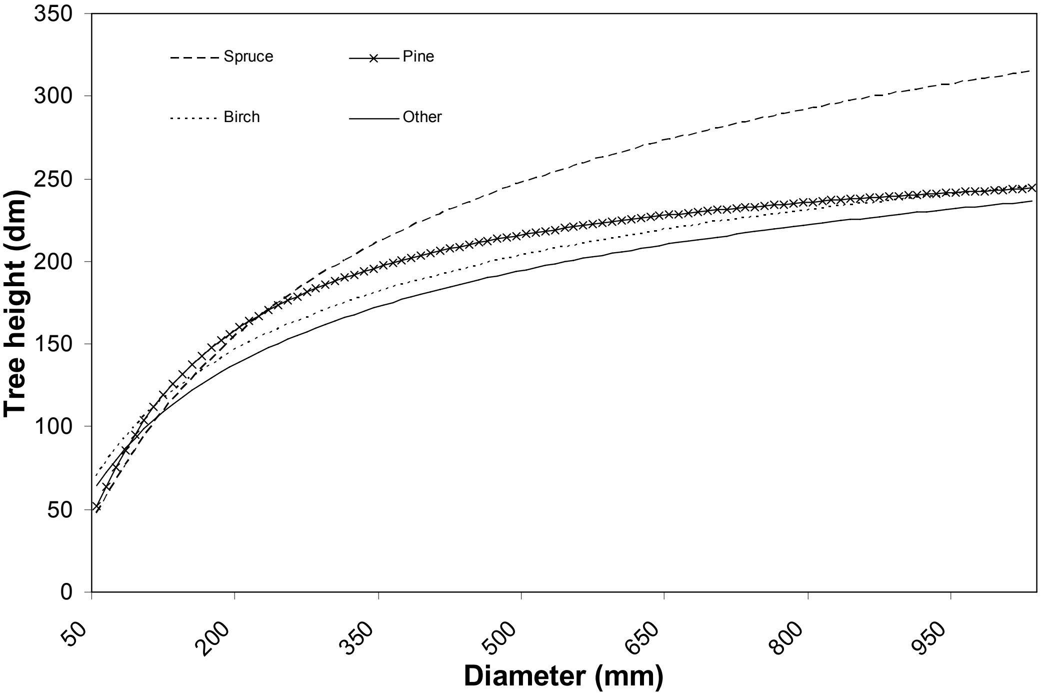

Figure 1. Tree height for different values of diameter (*DBH*) and for average forest conditions (*H40*=17, *LAT*=63, *ALT*=50, and *BA*=25) for all models.

*Self-validation*

Tables 6 and 7 display the self-validation results for the conifer and deciduous tree species, respectively. All values of bias were inside the range considered as acceptable by Huang et al. (2003). The self-validation also shows small random errors (%) for the conifer species and that they were smallest for Norway spruce. Some statistically significant biases were present for some of the subgroups, as displayed by tables 6 and 7. The main difference between the two set of models is observed from the prediction errors distributed on the subgroups of *BA* where it seemed that there was a trend for Model set 2.

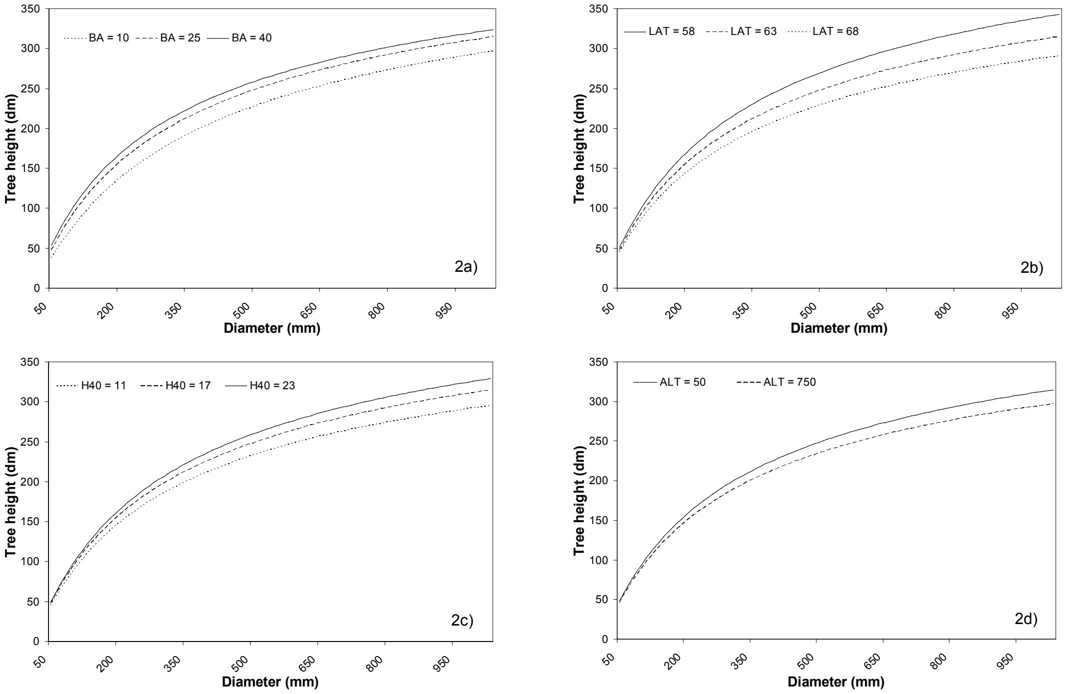

Figure 2. Tree height for different values of diameter and *BA* (2a), *LAT* (2b) *H40* (2c), and *ALT* (2d) utilizing the model for Norway spruce.

## *Independent model testing*

Tables 8 and 9 display results from the testing on the independent data for both model sets. For Model set 1, all differences between predicted and observed height values were smaller that 10%, except for the smallest diameter subgroup for spruce and pine (Table 8) and the second largest subgroup of diameter for other broadleaves (Table 9). Tested on the full range of the data, the prediction errors were –1.6%, –1.5%, –1.3%, and –3.8% for Norway spruce, Scots pine, birch, and other broadleaves, respectively. Distributed on the different subgroups of *DBH*, *H40*, *LAT*, *ALT* and *BA*, the model testing revealed no serious trends for the prediction error. The testing of Model set 2 produced similar results as for Model set 1. The main differences between the model sets were that Model set 2 seemed to produce larger height values on average. The trend for the *BA* subgroup for Model set 2 was similar to the self-validation.

Table 6. Results from model cross-validation for Norway spruce and Scots pine distributed on subgroups of diameter (*DBH*), site quality (*H40*), latitude (LAT), altitude (ALT), and stand basal area (BA)

| Dataset | Norway spruce | | | | Scots pine | | | |
|---|---|---|---|---|---|---|---|---|
| | # of obs | Obs. mean | Bias[a] 1 (%) | Bias[a] 2 (%) | # of obs | Obs. mean | Bias[a] 1 (%) | Bias[a] 2 (%) |
| All | 7,601 | 154 | **–0.1** *15.1* | **0.0** *16.2* | 5,519 | 148 | **–0.1** *17.2* | **0.0** *18.0* |
| 5< *DBH* ≤10 | 753 | 69 | **–1.9** *19.6* | 2.2 *20.3* | 211 | 69 | –8.0 *23.3* | 6.8 *27.2* |
| 10< *DBH* ≤15 | 1,200 | 107 | **0.2** *17.0* | **–0.3** *19.0* | 448 | 101 | **0.3** *19.6* | **–0.1** *21.9* |
| 15< *DBH* ≤20 | 1,504 | 138 | **0.5** *15.7* | **–0.7** *17.4* | 818 | 126 | 2.4 *17.8* | –2.8 *19.9* |
| 20< *DBH* ≤25 | 1,433 | 164 | **0.3** *14.6* | **–0.6** *16.1* | 1,024 | 147 | **–0.3** *16.8* | **0.3** *17.9* |
| 25< *DBH* ≤30 | 1,138 | 186 | **–1.0** *13.7* | 1.2 *14.5* | 1,030 | 158 | –1.4 *16.6* | 1.4 *17.4* |
| 30< *DBH* ≤35 | 780 | 203 | **–0.8** *13.4* | **0.8** *14.2* | 919 | 167 | **–1.0** *16.5* | **0.9** *16.7* |
| 35< *DBH* | 793 | 221 | **1.1** *13.5* | **–0.8** *13.5* | 1,069 | 177 | **1.2** *16.3* | **–1.0** *16.3* |
| *H40*≤8 | 1,550 | 132 | 2.0 *15.0* | –2.1 *15.0* | 2,342 | 129 | **–0.3** *17.6* | **–0.1** *17.8* |
| 8<*H40*≤14 | 3,302 | 151 | –1.8 *14.8* | 1.5 *16.2* | 2,586 | 159 | **–0.6** *16.8* | **0.7** *17.7* |
| 14<*H40*≤20 | 2,332 | 166 | **0.1** *15.1* | **0.1** *16.4* | 571 | 177 | 2.4 *16.5* | –2.2 *18.2* |
| 20<*H40* | 417 | 187 | 5.2 *14.3* | –4.3 *15.0* | 20 | 199 | **5.7** *14.6* | **–4.8** *14.3* |
| 58<*LAT*≤61 | 4,605 | 159 | **–0.4** *15.1* | **0.4** *16.1* | 3,596 | 155 | **–0.4** *16.3* | **0.4** *17.1* |
| 61<*LAT*≤64 | 2,063 | 148 | **0.7** *15.3* | **–0.2** *16.4* | 1,701 | 138 | **0.5** *19.1* | **–0.4** *19.8* |
| 64<*LAT*≤70 | 933 | 140 | **0.2** *14.4* | –1.4 *15.4* | 222 | 113 | **1.1** *18.9* | **–2.8** *19.3* |
| *ALT*≤150 | 1,324 | 160 | 1.7 *15.4* | –1.7 *16.2* | 1,144 | 145 | 4.7 *17.8* | –3.9 *18.8* |
| 150<*ALT*≤300 | 2,100 | 160 | –2.3 *15.1* | 2.2 *16.3* | 1,875 | 157 | –1.8 *16.6* | 2.0 *17.4* |
| 300<*ALT*≤450 | 1,566 | 153 | –0.9 *14.4* | **0.7** *15.8* | 1,125 | 150 | **–1.0** *16.3* | **0.7** *17.1* |
| 450<*ALT* | 2,611 | 146 | 1.5 *15.0* | –1.4 *15.9* | 1,375 | 138 | **–0.9** *17.5* | **–0.2** *18.2* |
| *BA* ≤ 16 | 1,261 | 107 | **–0.9** *16.9* | –11.1 *18.6* | 1,434 | 121 | **–0.6** *20.3* | –6.9 *20.7* |
| 16<*BA*≤24 | 1,646 | 137 | 1.3 *15.9* | –3.7 *16.4* | 1,678 | 144 | **–0.1** *16.9* | **–0.1** *17.3* |
| 24<*BA*≤32 | 1,827 | 156 | **0.2** *14.8* | **0.4** *15.1* | 1,290 | 156 | 1.7 *16.2* | **0.5** *16.5* |
| 32<*BA* | 2,867 | 182 | **–0.5** *14.2* | 4.3 *14.3* | 1,117 | 180 | –1.5 *15.6* | 5.7 *15.7* |

[a] Significance level: **bold** charcters=not significant ($p$>0.05). Bias 1: Bias using Model 1. Bias 2: Bias using Model 2. Small italic numbers are standard deviations of the respective biases.

## Discussion

Non-linear height-diameter models were estimated from NFI-data. For Model set 1, *LAT*, *ALT*, and *H40* determined an asymptotic height and *DBH* and *BA* the slope of the height curve. Model set 2 excluded *BA*. All signs were logic in relation to the effect on height-diameter relationship. The shapes of the height curves according to diameter using Model set 1 are displayed in Figure 1. As the figure shows, all species specific models produce different maximum heights and different slopes under otherwise equal conditions. The slopes as displayed in Figure 1 are in correspondence to prior expectations in relation to pioneer (pine, birch) and climax species (spruce). The shapes of the height curves are similar using Model set 2. Both pine and birch allocate more resources to height growth compared to spruce for small diameters, and the models reproduces this behavior. As a

Table 7. Results from model cross-validation for birch and other broadleaves distributed on subgroups of diameter (*DBH*), site quality (*H40*), latitude (LAT), and altitude (ALT)

| | Birch | | | | Other broadleaves | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | # of obs | Obs. mean | Bias[a] 1 (%) | Bias[a] 2 (%) | # of obs | Obs. mean | Bias[a] 1 (%) | Bias[a] 2 (%) |
| All | 3,381 | 106 | **–0.1** *17.8* | **0.0** *19.0* | 1,486 | 124 | **–0.1** *21.5* | **0.0** *22.3* |
| 5< *DBH* ≤10 | 894 | 74 | **–0.4** *18.9* | **0.3** *20.8* | 300 | 79 | **–2.2** *20.0* | **1.5** *20.8* |
| 10< *DBH* ≤15 | 1,067 | 98 | **0.4** *17.8* | **–0.3** *19.2* | 369 | 105 | **0.9** *20.6* | **–0.9** *22.0* |
| 15< *DBH* ≤20 | 682 | 118 | **–1.6** *16.5* | **1.0** *17.7* | 323 | 127 | **1.3** *21.5* | **–1.3** *22.7* |
| 20< *DBH* ≤25 | 397 | 135 | **0.5** *16.2* | **–0.4** *17.3* | 189 | 143 | **1.1** *20.1* | **–1.5** *21.2* |
| 25< *DBH* ≤30 | 166 | 151 | **2.1** *17.6* | **–1.8** *18.5* | 122 | 162 | **–1.0** *20.3* | **1.7** *20.4* |
| 30< *DBH* ≤35 | 100 | 164 | **–0.1** *16.2* | **0.5** *17.1* | 81 | 181 | –6.7 *16.4* | 6.7 *16.9* |
| 35< *DBH* | 75 | 187 | **0.1** *17.9* | **0.6** *17.1* | 102 | 187 | **1.9** *23.5* | **–2.0** *23.7* |
| *H40*≤8 | 1,455 | 89 | –1.3 *17.6* | **1.0** *18.7* | 255 | 97 | **0.3** *23.6* | **–0.8** *23.4* |
| 8<*H40*≤14 | 1,561 | 114 | **1.1** *17.6* | **–1.0** *18.4* | 740 | 119 | **0.4** *21.6* | **–0.6** *22.4* |
| 14<*H40*≤20 | 332 | 147 | **–1.5** *17.0* | **1.2** *19.6* | 417 | 143 | **–1.6** *19.7* | **1.5** *20.9* |
| 20<*H40* | 33 | 159 | **1.0** *17.9* | **1.1** *18.3* | 74 | 154 | **3.7** *23.6* | **–1.5** *24.8* |
| 58<*LAT*≤61 | 1,413 | 119 | **–1.1** *18.2* | 1.3 *19.6* | 880 | 134 | **–0.7** *20.9* | **0.9** *21.7* |
| 61<*LAT*≤64 | 844 | 100 | 3.8 *17.9* | –3.8 *18.5* | 392 | 117 | 3.8 *23.0* | –3.4 *23.8* |
| 64<*LAT*≤70 | 1,124 | 96 | –1.5 *16.2* | **1.1** *17.4* | 214 | 93 | –5.0 *19.5* | **2.7** *20.2* |
| *ALT*≤150 | 790 | 108 | **0.8** *17.7* | **–1.0** *19.1* | 749 | 124 | **–1.1** *20.0* | **1.5** *20.8* |
| 150<*ALT*≤300 | 994 | 115 | **–1.2** *18.1* | **1.6** *19.3* | 464 | 125 | **1.5** *23.4* | **–2.1** *23.7* |
| 300<*ALT*≤450 | 635 | 109 | –3.0 *17.0* | 2.8 *18.5* | 167 | 124 | **–0.8** *23.2* | **0.4** *24.6* |
| 450<*ALT* | 962 | 96 | 2.7 *17.4* | –3.0 *18.0* | 106 | 116 | **2.0** *20.4* | **–2.3** *22.5* |
| *BA* ≤ 16 | 1,490 | 89 | **–0.9** *17.5* | –5.1 *18.8* | 283 | 92 | **–0.4** *21.8* | –10.6 *22.8* |
| 16<*BA*≤24 | 1,030 | 107 | 1.8 *17.7* | **0.1** *18.1* | 352 | 117 | **–0.8** *22.1* | **–1.6** *22.4* |
| 24<*BA*≤32 | 512 | 127 | **–0.3** *17.9* | 4.5 *18.2* | 431 | 124 | **2.0** *20.9* | **–0.4** *21.0* |
| 32<*BA* | 349 | 149 | **–1.6** *17.0* | 7.5 *16.7* | 420 | 151 | **–1.2** *20.9* | 5.8 *21.0* |

[a] Significance level: **bold** charcters=not significant ($p$>0.05). Bias 1: Bias using Model 1. Bias 2: Bias using Model 2. Small italic numbers are standard deviations of the respective biases.

climax species, spruce attains greater maximum heights than the other tree species in Norway which also is reproduced by this specific model. The model for other broadleaves behaves almost similar to the model for birch. This model is, however, not as easily interpreted because it ranges over a large number of species. However, most of these species can be defined as pioneers, which support the model behavior. Species specific models were not estimated since all species included in the other broadleaves model are of little economic importance and that no distinctions are made between these

Table 8. Results from independent model test for Norway spruce and Scots pine distributed on subgroups of diameter (*DBH*), site quality (*H40*), latitude (LAT), and altitude (ALT)

| | Norway spruce | | | | Scots pine | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | # of obs | Obs. mean | Bias$^a$ 1 (%) | Bias$^a$ 2 (%) | # of obs | Obs. mean | Bias$^a$ 1 (%) | Bias$^a$ 2 (%) |
| All | 5,040 | 120 | –1.6 *17.1* | 3.1 *18.1* | 3,117 | 123 | –1.5 *19.0* | 1.2 *20.0* |
| 5< *DBH* ≤10 | 1,368 | 59 | –11.4 *20.2* | 4.5 *19.3* | 508 | 61 | –13.9 *21.8* | –5.2 *28.2* |
| 10< *DBH* ≤15 | 1,074 | 96 | **–0.1** *18.1* | 7.8 *19.7* | 462 | 90 | **0.1** *21.7* | 8.2 *24.8* |
| 15< *DBH* ≤20 | 945 | 133 | **–0.5** *16.5* | 2.8 *18.6* | 548 | 119 | **0.5** *19.1* | 4.1 *20.0* |
| 20< *DBH* ≤25 | 720 | 160 | **–0.5** *14.6* | **1.3** *15.6* | 580 | 141 | **–1.5** *17.5* | **–0.1** *18.0* |
| 25< *DBH* ≤30 | 484 | 179 | **0.1** *14.2* | **1.4** *15.3* | 455 | 155 | **–1.6** *16.8* | **–0.8** *17.3* |
| 30< *DBH* ≤35 | 242 | 200 | **0.6** *14.2* | **1.1** *15.2* | 285 | 164 | **–0.5** *17.5* | **–0.3** *17.4* |
| 35< *DBH* | 207 | 224 | **0.8** *13.6* | **0.6** *13.9* | 279 | 170 | **1.9** *16.8* | **1.9** *16.7* |
| *H40*≤8 | 1,280 | 104 | **0.5** *17.5* | 5.3 *16.9* | 1,479 | 110 | –3.7 *19.6* | **–0.1** *19.6* |
| 8<*H40*≤14 | 2,322 | 115 | –3.8 *16.8* | 1.8 *18.0* | 1,417 | 131 | **–0.1** *18.4* | 2.2 *19.9* |
| 14<*H40*≤20 | 1,321 | 136 | **0.3** *16.7* | 4.1 *18.4* | 221 | 163 | **1.6** *17.3* | **1.9** *19.8* |
| 20<*H40* | 117 | 200 | **–3.5** *15.4* | **–2.7** *16.4* | - | - | - | - |
| 58<*LAT*≤61 | 2,959 | 128 | –2.5 *16.2* | 1.4 *17.5* | 1,839 | 131 | –1.8 *17.7* | **–0.1** *19.2* |
| 61<*LAT*≤64 | 1,415 | 107 | **0.6** *19.3* | 7.0 *19.6* | 1,085 | 112 | **–0.1** *21.7* | 4.5 *20.8* |
| 64<*LAT*≤70 | 666 | 111 | –1.6 *16.7* | 3.7 *16.7* | 193 | 97 | –6.8 *19.6* | **–3.1** *20.7* |
| *ALT*≤150 | 863 | 131 | –2.5 *17.0* | **1.2** *18.3* | 550 | 122 | 3.8 *19.6* | 4.6 *20.6* |
| 150<*ALT*≤300 | 1,166 | 130 | –3.2 *16.1* | **0.0** *17.4* | 1,168 | 129 | –3.1 *18.4* | –1.7 *20.0* |
| 300<*ALT*≤450 | 1,192 | 123 | –2.6 *16.5* | 2.4 *17.5* | 559 | 127 | **–1.0** *16.7* | 2.7 *18.3* |
| 450<*ALT* | 1,819 | 106 | **0.9** *19.1* | 7.0 *18.4* | 840 | 114 | –2.9 *20.4* | 2.3 *20.0* |
| *BA* ≤ 16 | 2189 | 82 | –6.2 *18.7* | 11.6 *20.2* | 1,389 | 98 | –3.9 *21.4* | 7.2 *22.0* |
| 16<*BA*≤24 | 1093 | 126 | 1.4 *16.5* | 3.6 *17.1* | 907 | 127 | **0.7** *18.7* | **0.2** *19.6* |
| 24<*BA*≤32 | 873 | 150 | **0.4** *16.2* | **–0.3** *16.3* | 513 | 156 | **–0.5** *16.3* | –3.0 *16.6* |
| 32<*BA* | 885 | 178 | **–0.6** *14.2* | –4.3 *14.1* | 308 | 175 | **–1.2** *15.8* | –5.6 *15.9* |

[a] Significance level: **bold** charcters=not significant ($p$>0.05). Bias 1: Bias using Model 1. Bias 2: Bias using Model 2. Small italic numbers are standard deviations of the respective biases.

species in practical inventories. Species specific predictive models of height of such species should not be developed from NFI data, but from data collected specially for this purpose. A model for all these species was nevertheless developed here until sufficient data exist.

Table 9. Results from independent model test for birch and other broadleaves distributed on subgroups of diameter (*DBH*), site quality (*H40*), latitude (LAT), and altitude (ALT)

| Dataset | Birch | | | | Other broadleaves | | | |
|---|---|---|---|---|---|---|---|---|
| | # of obs | Obs. mean | Bias$^{a}$ 1 (%) | Bias$^{a}$ 2 (%) | # of obs | Obs. mean | Bias$^{a}$ 1 (%) | Bias$^{a}$ 2 (%) |
| All | 3,845 | 88 | –1.3 *19.7* | 3.8 *20.3* | 1,201 | 100 | –3.8 *23.2* | 2.5 *23.8* |
| 5< *DBH* ≤10 | 1,699 | 66 | –2.6 *20.9* | 6.9 *21.0* | 531 | 70 | –8.0 *23.6* | 5.7 *23.7* |
| 10< *DBH* ≤15 | 1,176 | 91 | **–0.4** *19.5* | 3.5 *19.9* | 273 | 97 | **–0.8** *23.6* | 5.5 *24.9* |
| 15< *DBH* ≤20 | 583 | 112 | **–1.3** *17.5* | **1.5** *18.6* | 190 | 126 | **–3.1** *22.0* | **0.3** *22.2* |
| 20< *DBH* ≤25 | 263 | 128 | **0.3** *17.0* | **2.1** *17.5* | 103 | 141 | **–0.5** *19.0* | **0.1** *19.6* |
| 25< *DBH* ≤30 | 70 | 148 | **–1.9** *18.8* | **–1.9** *20.0* | 51 | 158 | **–2.4** *20.6* | **–2.7** *21.3* |
| 30< *DBH* ≤35 | 36 | 162 | **0.7** *17.2* | **0.2** *18.5* | 16 | 195 | **–11.8** *16.3* | **–13.1** *17.4* |
| 35< *DBH* | 18 | 182 | **–1.7** *20.6* | **–2.6** *19.5* | 37 | 192 | **–0.6** *21.2* | **–0.9** *20.9* |
| *H40*≤8 | 2,168 | 79 | –3.1 *18.7* | 1.5 *19.8* | 253 | 83 | **–6.0** *24.1* | **0.9** *22.7* |
| 8<*H40*≤14 | 1,434 | 96 | **0.7** *19.9* | 6.0 *20.1* | 584 | 95 | –3.4 *24.4* | 3.7 *25.2* |
| 14<*H40*≤20 | 203 | 118 | **2.1** *21.4* | 8.7 *20.8* | 341 | 118 | –3.1 *20.5* | **2.4** *21.5* |
| 20<*H40* | 40 | 131 | –8.1 *12.0* | **0.0** *13.4* | 23 | 152 | **–5.3** *24.8* | **–7.2** *25.6* |
| 58<*LAT*≤61 | 1,272 | 98 | –2.8 *20.6* | 2.7 *21.0* | 645 | 111 | –4.9 *22.7* | **1.1** *23.5* |
| 61<*LAT*≤64 | 1,020 | 82 | 2.6 *20.6* | 7.9 *19.9* | 294 | 96 | **1.4** *23.0* | 6.9 *23.5* |
| 64<*LAT*≤70 | 1,553 | 83 | **–2.3** *18.6* | 2.2 *19.2* | 262 | 79 | –7.1 *23.5* | **1.2** *23.7* |
| *ALT*≤150 | 922 | 89 | **–1.3** *20.1* | 4.3 *20.9* | 521 | 99 | –5.3 *22.6* | **1.4** *23.1* |
| 150<*ALT*≤300 | 1,078 | 95 | –3.7 *19.1* | **0.9** *20.4* | 376 | 106 | –4.4 *22.5* | **0.6** *23.8* |
| 300<*ALT*≤450 | 679 | 91 | **–1.5** *19.3* | 2.7 *19.0* | 202 | 97 | **–0.6** *25.3* | 5.7 *24.9* |
| 450<*ALT* | 1,166 | 78 | 1.7 *19.9* | 7.4 *19.8* | 102 | 89 | **0.3** *23.8* | 9.5 *23.2* |
| *BA* ≤ 16 | 2,723 | 79 | –3.4 *19.2* | 6.6 *19.6* | 596 | 79 | –8.2 *24.2* | 9.1 *24.4* |
| 16<*BA*≤24 | 750 | 100 | 2.6 *18.7* | **0.3** *19.0* | 310 | 105 | **–0.8** *21.7* | **1.3** *21.7* |
| 24<*BA*≤32 | 257 | 117 | 4.8 *19.1* | **0.2** *19.4* | 139 | 126 | **2.0** *18.0* | **0.9** *18.0* |
| 32<*BA* | 115 | 145 | **–2.7** *21.5* | –9.4 *21.0* | 156 | 151 | **–3.4** *22.8* | –8.0 *22.9* |

[a] Significance level: **bold** charcters=not significant ($p$>0.05). Bias 1: Bias using Model 1. Bias 2: Bias using Model 2. Small italic numbers are standard deviations of the respective biases.

Table 5 displays the parameter estimates and fit. The pseudo-$R^2$ in the range of 0.63 to 0.82 is somewhat lower than for example Sharma and Zhang (2004). It is, however, difficult directly to compare the findings in the present work to other models because of differences in methodology, geographical scale, and tree species modeled.

The random error was somewhat smaller for spruce compared to all other species (tables 6, 7, 8, and 9). The reason for this is probably that spruce has a more regular conical crown shape with a clearly defined top. This yields less measurement errors compared to the other species where the top can be difficult to delimit exactly, especially when height measurements are executed from a steep angles.

The self-validation (tables 6 and 7) revealed no serious anomaly. However, Table 6 shows that the height of pine is more under-predicted for small diameters using Model set 1 than the other species. This may be related to the fact that pine is a pioneer species, allocating a large proportion of its available resources to height growth in early age, and that the model is not sufficiently flexible to account for this. However, comparing the tests of the two model sets it can also be observed that *BA* probably contributes significantly to the under-prediction of tree heights for small diameter trees. The flexibility of the model form is thus probably sufficient.

Similar to the self-validation, the independent test results (tables 8 and 9) indicate that *BA* yield under-predictions of the height of small trees. The reason may be related to collinearity between *DBH* and *BA*.

## Conclusions

The non-linear height-diameter models presented here are developed for Norway spruce, Scots pine, birch and other broadleaves in Norway. The test results distributed on different subgroups of *DBH* and stand variables show no serious trends in the prediction error. However, Model set 1 that includes *BA* as an explanatory variable under-predicts the height of small trees. The variables required as input for the models are available in almost all practical inventories and the models can be utilized on the whole geographical range of the forested part of Norway. The models therefore provide a good alternative to constructing height curves specifically for individual stands or forests. Stand specific height curves will still be superior with respect to accuracy and the tradeoff between less accurate models and less inventory costs must be considered in each case. The models are considered to be robust and suitable for implementation in a single-tree based computer simulator. The model set that do not include *BA* (Model set 2) is considered to be best suitable for inclusion in a simulator because *BA* can be changed by silvicultural operations. This will affect the height estimate of the residual trees. Model set 1 is most suitable for uneven-aged forestry where the changes in stand density due to human intervention are low.

## Acknowledgements

The research leading to this article was founded by the Research Council of Norway (research project no. 153185/110). I wish to thank Professor Tron Eid (Norwegian University of Life Sciences) for discussions during the research process. I also wish to thank Dr. Arne Sandnes, Professor Erik Næsset, and Associate Professor Terje Gobakken (Norwegian University of Life Sciences) for valuable comments on earlier versions of the manuscript, and Rune Eriksen (Norwegian Institute of Land Inventory) for supplying the data.

## References

Anon., 1999. SAS OnlineDoc®, Version 8. SAS Institute Inc. Cary, NC.

Braastad, H., 1966. *Volume tables for birch.* Communications of the Norwegian Forest Research Institute, 21, 265-365. (In Norwegian with English summary).

Braastad, H., 1980. *Tilvekstmodellprogram for furu. (Growth model computer program for Pinus sylvestris)*. Norwegian Forest Research Institute. Report 35(5), 272-359. (In Norwegian with English Summary).

Brantseg, A., 1967. *Volume functions and tables for Scots pine*. South Norway. Communications of the Norwegian Forest Research Institute, 22, 695-739. (In Norwegian with English summary).

Chapman, D.G., 1961. Statistical problems in dynamics of exploited fisheries populations. In: Neyman, J. (Ed.), Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability, vol 4, (pp. 153-168). Berkeley CA.

Dorado, F.C., Diéguez-Aranda, U., Anta, M.B., Rodríguez, M.S., Gadow, K.v., 2006. A generalized height-diameter model including random components for radiate pine plantations in northwestern Spain. *For. Ecol. Manage. 229*, 202-213.

Fang, Z., Bailey, R.L., 1998. Height-diameter models for tropical forests on Hainan Island in southern China. *For. Ecol. Manage 110*, 315-327.

Friend, A.D., 1993. The prediction and physiological significance of tree height. In: Solomon, A.M., Shugart, H.H. (Eds), Vegetation dynamics and global change (pp. 101-115). Chapman & Hall, New York.

Fulton, M.R., 1999. Patterns in height-diameter relationships for selected tree species and sites in eastern Texas. *Can. J. For. Res. 29*, 1445-1449.

Huang, S., Yang, Y., Wang, Y., 2003. A critical look at procedures for validating growth and yield models. In: Amaro, A., Reed, D., Soares, P. (Eds), Modelling forest systems (pp. 271-293). CABI, Wallingford.

Koch, G.W., Sillett, S.C., Jennings, G.M., Davis, S.D., 2004. The limits to tree height. *Nature 428*, 851-854.

Lappi, J., 1997. A longitudinal analysis of height/diameter curves. *For. Sci. 43*, 555-570.

Larsen, D.R., Hann, D.W., 1987. *Height-diameter equations for seventeen tree species in southwest Oregon*. Oregon State Univ. For. Res. Lab. Pap., 4 pp.

Loetsch, F., Zöhrer, F., Haller, K.E., 1973. Forest inventory, Volume II. BLV, München, Bern, Wien.

Mäkinen, H., 1998. The suitability of height and radial increment variation in Pinus Sylvestris (L.) for expressing environmental signals. *For. Ecol. Manage 112*, 191-198.

Miller, R.G., 1981. Simultaneous Statistical Inference. 2nd edition. Springer, New York.

Parresol, B.R., 1992. Baldcypress height-diameter equations and their prediction confidence intervals. *Can. J. For. Res. 22*, 1429-1434.

Peng, C.H., 1999. *Non-linear height-diameter models for nine tree species in Ontario boreal forests*. Ministry of Natural Resources, Ontario Forest Research Institute, OFRI-Rep. 155, 28 pp.

Peng, C.H., Zhang, L., Liu, J., 2001. Developing and validating nonlinear height-diameter models for major tree species of Ontario`s boreal forests. *North. J. Appl. For. 3*, 87-94.

Richards, F.J., 1959. A flexible growth function for empirical use. *J. Exp. Biol. 10*, 290-300.

Ryan, M.G., Yoder, B.J., 1997. Hydraulic limits to tree height and tree growth. *BioScience 47*, 235-242.

Sharma, M., Yin Zhang, S., 2004. Height-Diameter Models Using Stand Characteristics for *Pinus banksiana* and *Picea mariana*. *Scand. J. For. Res. 19*, 442-452.

Stage, A.R., 1963. A mathematical approach to polymorphic site index curves for grand fir. *For. Sci. 9*, 167-180.

Tveite, B., 1977. *Site index curves for Norway spruce (Picea abies* (L.) Karst.). Norwegian Forest Research Institute, Report 33, 1-84.

Tyree, M.T., Sperry, J.S., 1988. Do woody plants operate near the point of catastrophic xylem disfunction caused by dynamic water stress? *Plant Physiol. 88*, 574-580.

Tyree, M.T., Zimmermann, M.H., 2002. Xylem structure and the ascent of sap. 2nd ed. Springer-Verlag, Berlin.

Vanclay, J.K., 1994. Modelling forest growth and yield: Applications to mixed tropical forests. CABI, Wallingford.

Vestjordet, E., 1967. *Functions and tables for volume of standing trees. Norway spruce*. Communications of the Norwegian Forest Research Institute, 22, 543-574. (In Norwegian with English summary).

Wykoff, W.R., 1990. A basal area increment model for individual conifers in the northern Rocky Mountains. *For. Sci. 36*, 1077-1104.

Wykoff, W.R., Crookston, N.L., Stage, A.R., 1982. *User`s guide to the stand prognosis model*. USDA For. Serv. Gen. Tech. Rep. INT-133.

Yang, R. C., Kozak, A., Smith, J.H.G., 1978. The potential of Weibull-type functions as a flexible growth curve. *Can. J. For. Res. 8*, 424-431.

Zeide, B., 1989. Accuracy of equations describing diameter growth. *Can. J. For. Res. 19*, 1283-1286.

Zhang, L., 1997. Self-validation of non-linear growth functions for modeling tree height diameter relationships. *Ann. Bot. 79*, 251-257.

Øyen, B.H, Andreassen, K. (2002). *Modeller for å estimere trehøyde ved hjelp av diameter og bestandsdata*. Norwegian Forest Research Institute. Aktuelt 6, 13-14. (In Norwegian)

# Paper IV

# Predicting the growth of stands of trees of mixed species and size: A matrix model for Norway

Ole Martin Bollandsås[1*], Joseph Buongiorno[2], Terje Gobakken[1]

[1]Norwegian University of Life Science

Department of Ecology and Natural Resource Management

P.O. Box 5003, N-1432 Ås, Norway

* Phone: +47 64965753, e-mail: olebo@umb.no

[2]University of Wisconsin

Department of Forest and Wildlife Ecology

1630 Linden Drive, Madison, Wisconsin 53706, USA

## Abstract

The objective of this study was to predict the growth of forest stands of mixed tree species and size with natural recruitment. The stand state was defined by the number of spruce, pine, birch, and other broadleaf trees by ha in fifteen diameter classes from 50 to 750 mm. The change in stand state over 5 years was predicted with state-dependent matrices based on equations for recruitment, growth, and mortality. The data came from 7,241 plots of the national forest inventory of Norway, measured from 1994 to 2005. A short term validation was carried out by comparing predicted and actual growth over 10 years on 416 plots not used in model estimation. The model was also used to predict the long-term growth of stands with different initial species composition and diameter distribution. Irrespective of the initial condition the same steady state resulted, with characteristics similar to those observed in stands that had been undisturbed for 75 years.

*Keywords:* Biometrics, forest growth, matrix model, simulation.

## Introduction

Norway has a long north to south coast and large differences in elevation. Thus, the forest conditions vary along gradients of altitude, latitude, soil type, and humidity. Together with different management practices this has produced a wide range of forest types with different levels of recruitment, growth, and mortality.

Most existing Norwegian growth models are monospecific and predict only mean stand characteristics. This was expedient for commercial even-aged silviculture. However, because of certification systems and subsidy regulations that promote biological diversity, a larger part of the forest stands are likely to be of trees of various species and size in the future. Correspondingly, growth models to guide the management of such forests will need to predict the species composition of forest stands and their structure (diameter distribution).

Although stand level models are suitable to deal with even-aged forestry, uneven-aged management requires that the models distinguish individual trees or size and species classes to account for varying growth conditions due to stand structure and species mix within each stand. Models of tree growth (Andreassen and Tomter, 2003), recruitment (Lexerød, 2005), and mortality (Eid and Tuhus, 2001) have been estimated for Norwegian conditions previously. However, the model of Andreassen and Tomter (2003) includes a subjective field-assessed indicator variable to distinguish multi canopy layered stands from one layered stands that is not easily quantified and integrated in simulation models.

For long-term projections, the tree growth and related models must be implemented in computer tools. The tools that are available in Norway today, BESTPROG (Blingsmo and Veidahl, 1993), AVVIRK 2000 (Eid and Hobbelstad, 2000), and GAYA (Hoen and Eid, 1990) are for even-aged stands. Single tree simulators such as the SILVA simulator (Pretzsch et al., 2002) are able to handle uneven-aged forest management and a single tree simulator for Norwegian conditions is also under development (Gobakken et al. 2005). However, the complexity of single-tree simulators, the large amount of information they require, and the long processing time, still make them difficult to use for management which demands the study of many different policies over long periods of time. Simpler and more compact models dealing with species and size classes are more efficient and practical for that purpose.

The objective of this study was to develop a matrix model that would predict accurately the growth of forest stands of mixed tree species and size in Norway, over long periods of time. While dealing with species and size classes only, its parameters would rely on the same detailed data as an individual tree model. A main requirement was that the model be useful to deal with "uneven-aged", or "continuous-cover" forestry regimes, i.e. "silvicultural systems which involve continuous and uninterrupted maintenance of forests cover and which avoid clear cutting" (Pommering and Murphy, 2004).

The remainder of the paper presents the matrix model structure, its calibration with the extensive individual tree and stand data of the National Forest Inventory (NFI) of Norway, the short-term validation of the model with post-sample data, and its long term prediction of forest stand growth for various initial conditions.

## Materials and methods

While conceptually simple, matrix models are very general, and able to represent tree competition based on a wide variety of stand or individual tree data. Matrix models define the stand state by the numbers of trees in different size and species classes, and project this stand state by means of transition probabilities. Linear (i.e. state independent) and non-linear (state dependent) matrix models have been developed by several authors (e.g. Usher, 1966; Bosch, 1971; Buongiorno and Michie, 1980; Lu and Buongiorno, 1993; Liang et al., 2005).

The parameters of the matrix model proposed here account for a wide range of growth conditions by using national forest inventory (NFI) data, for applicability of the model to the whole productive forest area of Norway.

We defined a stand state by the number of trees per ha in four species groups and fifteen size classes. This species-size class definition is simply understood, and the measurements can be readily obtained from field measurements or remote sensing (Maltamo et al., 2000; Maltamo et al., 2005; Bollandsås and Næsset, 2007a), or with models that predict diameter distributions based on stand data (Vestjordet, 1972; Mønnes, 1982; Holte, 1993). At the same time, the species-size definition of stand states is sufficiently detailed to exploit the rich source of information in individual tree data to predict stand growth.

*Model structure*

Matrix models predict future stand states by means of transition matrices that operate on the previous stand state. The stand state is a vector of the number of trees per unit area in predefined species and diameter classes. The transition matrices used here are state dependent and also depend on site characteristics. The model general form is:

$$\boldsymbol{y}_{\mathrm{t+p}} = \boldsymbol{G}\ (\boldsymbol{y}_t - \boldsymbol{h}_t) + \boldsymbol{R} \qquad (1)$$

where $\boldsymbol{y}_t = [\boldsymbol{y}_{ijt}]$ is the vector state, in which each entry is the number of live trees per unit area of species group ($i = 1,\ldots,m$) and diameter class ($j = 1,\ldots,n$) at time $t$, $\boldsymbol{G}$ is a state-dependent transition probability matrix, $\boldsymbol{h}_t = [h_{ijt}]$, is the harvest vector in which each entry is the number of trees cut per unit area, and $p$ is the projection interval. $\boldsymbol{R}$ is the recruitment vector, also state dependent, in which each entry is the number of recruits per unit area by species and diameter class. The transition probability matrix $\boldsymbol{G}$ has one sub-matrix for each species group with the following structure:

$$\boldsymbol{G}_i = \begin{bmatrix} \boldsymbol{a}_{i1} & 0 & \cdots & 0 & 0 \\ \boldsymbol{b}_{i2} & \boldsymbol{a}_{i2} & \cdots & 0 & 0 \\ 0 & \boldsymbol{b}_{i3} & \ddots & 0 & 0 \\ \vdots & \cdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{b}_{in} & \boldsymbol{a}_{in} \end{bmatrix} \tag{2}$$

where $\boldsymbol{a}_{ij}$ ($i = 1, \ldots, m, j = 1, \ldots, n$) is the probability that a tree of species $i$ will remain in diameter class $j$ between $t$ and $t + p$, $\boldsymbol{b}_{ij}$ is the probability that a tree in diameter class $j$ grows into diameter class $j$+1 during the time interval $p$.

Similarly the recruitment vector consists of one sub-vector for each species group, indicating the number of recruits of a specific species $i$ that enter the smallest diameter class from $t$ to $t + p$:

$$\boldsymbol{R}_i = \begin{bmatrix} \boldsymbol{d}_i \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3}$$

where $\boldsymbol{d}_i$ is the number of trees of species $i$ that enter the smallest diameter class between $t$ and $t + p$.

The transition probability $b_{ij}$ was calculated here as the ratio of the rate of diameter growth to the width of a diameter class. The tree growth rate depended on the diameter class, on the stand state, and on site characteristics. The probability that a tree stayed in a diameter class, $\boldsymbol{a}_{ij}$, was calculated as $\boldsymbol{a}_{it} = 1 - \boldsymbol{b}_{ij} - \boldsymbol{m}_{ij}$, where $\boldsymbol{m}_{ij}$ was the probability that a tree of species group $i$

and diameter class *j* died during the interval *t* to *t+p*. Like the transition probability, the mortality rate was a function of the diameter class of the tree, and of the stand and site characteristics.

The recruitment rate, $\boldsymbol{d}_i$ was obtained as the product of the probability of positive recruitment and of the expected recruitment conditional on positive recruitment, which both depended on the stand state and the site characteristics.

### *Data*

The data used to estimate the model parameters came from 250 $m^2$ permanent field plots of the Norwegian national forest inventory, measured between 1994 and 2005. All plots are re-measured every five years, so that there were at least two measurements per plot and for 2/5 of the plots, there were three. The sample design is a 3 × 3 km grid which yields a representative sample of the Norwegian forest area. The northernmost county (Finnmark, above 70ºN latitude) was sampled differently because it has few forests, and was therefore excluded from the study.

Plots intercepted by roads, water, agricultural land etc, were also excluded to avoid edge effects. Plots classified as unproductive, i.e. with a volume increment of less than 1 $m^3ha^{-1}yr^{-1}$ were excluded because the site index (*SI*) is not registered below this level. The *SI* is defined by the height of the dominant trees at 40 years of age at breast height (Tveite, 1977; Braastad, 1980). Plots were also excluded when tree species could not be identified unambiguously, as was the case on some of the youngest stands. Furthermore, among the plots that had been measured three times, 416 plots (20%) were selected at random and put aside for validation. This left 7,241 plots with the characteristics summarized in Table 1. The youngest stands at the time of the first measurement were 20 years old, and the average age of all the stands was 87 years. Hence, the recruitment that occurred between measurements was assumed to be mostly due to natural regeneration.

Table 1. Summary of data for the 7,241 plots used for model calibration.

| Variable | Mean | Range | SD |
|---|---|---|---|
| Site index, *SI* (m) | 11.0 | 6.0 - 26.0 | 4.0 |
| Stand age (yr) | 86.7 | 20.0 - 344 | 35.6 |
| Basal area, *BA* ($m^2ha^{-1}$) | 18.4 | 0.1 - 79.5 | 10.3 |
| Growth seasons (yr) | 5.0 | 4.0 - 6.0 | 0.5 |
| Latitude, *LAT*[a] | 61.9 | 58.0 - 70.0 | 3.0 |
| Recruitment rate ($ha^{-1}$ $yr^{-5}$) | 62.6 | 0.0 - 1520 | 104 |
| Mortality rate ($ha^{-1}$ $yr^{-5}$) | 51.7 | 0 - 1560 | 98.5 |

[a] Degrees and minutes north according to Euref89. Minutes scaled by 100/60 (60° 30'N=60.5°N).

The species had been recorded for all callipered trees on the plots. For this study, the trees were sorted in four species groups: Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), birch (*Betula pubescens* and *Betula pendula*), and other broadleaved species which consisted of, in alphabetical order: alder (*Alnus spp.*), ash (*Fraxinus excelsior*), aspen (*Populus tremula*), beech (*Fagus silvatica*), Bird cherry (*Prunus padus*), linden (*Tilia cordata*), marple (*Acer spp.*), oak (*Quercus spp.*), rowan (*Sorbus aucuparia*), sallow (*Salix caprea*), and Wych elm (*Ulmus glabra*).

Table 2. Number of plots with recruitment, and number of recruit trees on the 7,241 calibration plots.

| Species | Plots with recruitment | | 5 yr recruitment ($ha^{-1}$) | |
|---|---|---|---|---|
| | Number | % of total plots | On plots with recruitment | On all plots |
| Spruce | 1841 | 25.4 | 86.1 | 21.9 |
| Pine | 383 | 5.3 | 57.0 | 3.0 |
| Birch | 1879 | 25.9 | 89.6 | 23.2 |
| Other broadleaves | 932 | 12.9 | 111.8 | 14.4 |

### *Recruitment model*

Recruits were trees that reached or exceeded a diameter of 50 mm between two inventories. Table 2 shows that the average recruitment rates were highest, and similar, for birch and spruce, while they were lowest for pines.

The logistic model (Agresti, 1996) appearing in equation (4) was used to predict the probability, $\pi_i$, of positive recruitment in $p$=5 years, for each species $i$, on a plot of given initial characteristics:

$$\pi_i = \left(1 + e^{-(\alpha_{i0} + \alpha_{i1} BA + \alpha_{i2} SI + \alpha_{i3} PBA)}\right)^{-1} \tag{4}$$

where *BA* is the stand basal area, *SI* is the site index, *PBA* is the percentage of basal area for the subject species. The parameters *α* were estimated with data from the 7,241 plots with the PROC LOGISTIC program of SAS (Anon, 1999). Model (4) is a simplified version of Lexerød (2005) with fewer explanatory variables.

The expected number of recruits conditional on positive recruitment, $CR_i$, was predicted with the following model:

$$CR_i = \beta_{i0} BA^{\beta_{i1}} SI^{\beta_{i2}} PBA^{\beta_{i3}} \tag{5}$$

The *β* parameters were estimated from the plots where recruitment did occur, by linear regression after logarithmic transformation of the variables. To predict the expected number of recruits, the intercept was adjusted for logarithmic bias (Flewelling and Pienaar, 1981; Miller, 1984).

The product of the probability of positive recruitment from (4) and the conditional expected recruitment (5) gave the expected number of recruits, $d_i$, given a stand state and site characteristics.

### *Tree growth model*

Diameters of trees ≥50 mm at breast-height were recorded to the nearest millimeter. The diameter increment was defined as the difference between two measurements at an interval of five years. However, as the actual number of growth seasons ranged from 4 to 6, all diameter growth measurements were adjusted to correspond to 5 growth seasons. Furthermore, increments smaller than –10 mm or larger than 70 mm, of which there were 809, were assumed to have large errors and were discarded. Table 3 summarizes the observations for a total of 178,472 trees. The pines were the largest trees on average, but they grew slightly slower than the spruces. The birches were the smallest trees and grew at half the rate of the spruces.

The diameter growth of a tree of species *i* over 5 years was represented by the equation:

$$I_{\mathrm{i}} = \theta_{\mathrm{i}0} + \theta_{\mathrm{i}1}DBH + \theta_{\mathrm{i}2}DBH^2 + \theta_{\mathrm{i}3}BAL + \theta_{\mathrm{i}4}SI + \theta_{\mathrm{i}5}BA + \theta_{\mathrm{i}6}LAT \tag{6}$$

where *DBH* is the tree diameter at breast height, *BAL* is the basal area per ha of the trees larger than the tree of interest in the stand in which the tree is growing, and *LAT* is the latitude were minutes are scaled by 100/60 to obtain a continuous value. The $\theta$ parameters were estimated by ordinary least squares from the individual tree and corresponding plot data.

Table 3. Diameter at the time of the first measurement (DBH), and five-year diameter increment ($I_{5yr}$) of trees that survived between measurements.

| Species | Number of trees | *DBH* (mm) Mean | Range | SD | $I_{5yr}$ (mm) Mean | Range | SD |
|---|---|---|---|---|---|---|---|
| Spruce | 68,584 | 136.4 | 50.0 - 700.0 | 74.0 | 10.6 | –10.0 - 69.8 | 9.7 |
| Pine | 29,447 | 183.0 | 50.0 - 800.0 | 89.3 | 9.6 | –9.97 - 68.6 | 7.7 |
| Birch | 60,229 | 99.9 | 50.0 - 586.0 | 46.5 | 5.3 | –10.0 - 69.0 | 6.3 |
| Other broadleaves | 20,212 | 105.9 | 50.0 - 647.0 | 57.9 | 8.5 | –10.0 - 69.4 | 8.5 |

*DBH* = diameter at breast height at time of first measurement. $I_{5yr}$ = five year diameter increment.

The probability that a tree grew from diameter class *j* to *j* + 1 between *t* and *t* + 5 years given a particular stand state and site characteristics was then estimated as:

$$b_{ij} = \frac{I_{ij}}{w} \tag{7}$$

where $I_{ij}$ is was the diameter increment predicted with equation (6) for the average tree in diameter class *j*, and *w* is the width (mm) of the diameter classes. Here, *w* was set at 50mm, to ensure that no tree could grow more than one diameter class in one period, while keeping the distribution of trees within a diameter class nearly uniform.

### Mortality model

A summary of the mortality data appears in Table 4. Spruces and pines had the lowest and nearly equal mortality rate of 3 percent, birches and other broadleaves had the highest mortality rates.

Table 4. Tree mortality data.

| Species | Total number of trees at first measurement[a] | Trees that died between measurements | |
|---|---|---|---|
| | | Trees | % |
| Spruce | 71,805 | 2,212 | 3.08 |
| Pine | 30,645 | 891 | 2.91 |
| Birch | 66,156 | 3,595 | 5.43 |
| Other broadleaves | 23,944 | 2,606 | 10.9 |

[a] The difference between total number of trees and trees that died between measurements, do not equal number of live trees used for modeling increment because 809 trees with abnormal increment were excluded from the increment data. Furthermore, there were missing data for one or more of the explanatory variables for 3965 observations.

The mortality models developed here are modifications of Eid and Tuhus (2001). The most important change is the inclusion of stand basal area as an explanatory variable. Otherwise, they use fewer explanatory variables: the probability of death of a tree in 5 years was a species-specific logistic function of tree size and stand basal area only:

$$m_{\mathrm{i}} = \left(1 + \mathrm{e}^{-\left(\delta_{\mathrm{i}0} + \delta_{\mathrm{i}1} DBH + \delta_{\mathrm{i}2} DBH^{2} + \delta_{\mathrm{i}3} BA\right)}\right)^{-1} \tag{8}$$

The $\delta$ parameters were estimated with the individual tree data by logistic regression. The equation (8) was then used to predict the expected mortality, $m_{ij}$, for trees in each size class $j$, based on the mean diameter of that size class, and the stand basal area.

### *Tree volume and tree height models*

The expected volume of trees by diameter class was estimated with species-specific volume equations for individual trees, based on *DBH* and height ($h$) (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967). The expected tree height needed in the volume equations is described in Bollandsås and Næsset (2007b).

### *Short-term validation*

The simulations for short-term and long-term validation were carried out with Visual Basic of Microsoft Excel. Short-term validation was done by comparing actual stand growth and model predictions on 416 randomly selected, independent, plots. On all validation plots growth was observed over 10 years (three measurements).

After estimating the parameters, the matrix model was applied to the plot data at the time of the first measurement, and two 5-year iterations of the model were performed. If harvests were recorded in the data, they were allocated in equal halves to the beginning and the end of the period during which they occurred, because their precise date was unknown.

Then, the predicted number of trees and basal area by species and diameter class after 10 years was compared to the observed, over all the post-sample plots (Lu and Buongiorno, 1993). Plots defined as “young forest”, which had not been used in developing the model, were not used in this validation.

### *Long-term validation*

The long term model validation consisted in simulating stand growth without harvest for 1,000 years. The hypothesis was that the predicted stand would reach a steady state corresponding to the ecological climax (Buongiorno et al. 1995). Furthermore, without climate change or other major disturbance, the predicted steady state should be independent of the initial stand state (Stenberg and Siriwardana, 2006).

Four simulations assumed a pure initial stand of either spruce, pine, birch, or other broadleaves. The stands had 100 trees of diameter class 175 and 275 mm, and 400 trees in diameter class 225 mm, for a total of 600 trees per ha. For each of these simulations we monitored the evolution of total basal area and basal area by species. Moreover, stand volume, number of stems, mean height, and mean diameter at the end of the 1,000 year period were evaluated.

Table 5. Model parameters for the probability of recruitment on a particular plot.

| Variable[b] | Parameter [a] (*Standard error*) | | | |
|---|---|---|---|---|
| | Spruce | Pine | Birch | Other broadleaves |
| Intercept | –2.291 (*0.087*) | –3.552(*0.163*) | –0.904 (*0.075*) | –3.438 (*0.124*) |
| *BA* ($m^2ha^{-1}$) | –0.018 (*0.003*) | –0.062 (*0.007*) | –0.037 (*0.003*) | –0.029 (*0.005*) |
| *SI* (m) | 0.066 (*0.008*) | | | 0.123 (*0.011*) |
| *PBA* (%) | 0.019 (*0.001*) | 0.031 (*0.002*) | 0.016 (*0.001*) | 0.048 (*0.002*) |
| Observations | 7,241 | 7,241 | 7,241 | 7,241 |
| Misclassified plots (%)[c] | 8.4 | 0.8 | 3.5 | 2.3 |

[a] $p<0.001$ for all parameters.
[b] *BA*=Stand basal area, *SI*=site index value, *PBA*=percentage of basal area of the subject species
[c] Misclassified plots according to the Hosmer and Lemeshow test.

Four other simulations were carried out for initial stands of mixed species on different sites (*SI*=6, 11, 17, and 23). The total number of trees and basal area were the same as for the single-species simulations, and the number of trees and basal area were uniformly distributed over the four species groups.

Table 6. Model parameters for number of recruits per ha conditional on positive recruitment.

| Variable[b] | Parameter [a] (*Standard error*)<br>Spruce | Pine | Birch | Other broadleaves |
|---|---|---|---|---|
| Intercept | 43.142 (*1.130*) | 67.152 (*1.094*) | 64.943 (*1.140*) | 31.438 (*1.202*) |
| $BA$ (m$^2$ha$^{-1}$) | –0.157 (*0.028*) | –0.076 (*0.035*) | –0.161 (*0.027*) | –0.1695 (*0.042*) |
| $SI$ (m) | 0.368 (*0.047*) |  | 0.143 (*0.050*) | 0.442 (*0.075*) |
| $PBA$[c] (%) | 0.051 (*0.013*) |  | 0.104 (*0.013*) | 0.193 (*0.016*) |
|  |  |  |  |  |
| Observations | 1,841 | 382 | 1,879 | 932 |
| $R^2$ | 0.05 | 0.01 | 0.07 | 0.18 |

[a] $p<0.05$ for all parameters.

[b] $BA$=stand basal area, $SI$=site index value, $PBA$=percentage of basal area of the subject species.

[c] 1 was added to the actual value of $PBA$ to enable log transformation when $PBA$ was 0.

We also did simulations for four mixed spruce/birch stands of very different initial basal area to verify that over time the stands converged to steady states of similar basal area.

As part of the long term validation, the simulated steady state for the pure spruce initial stand was compared with data from an old spruce-dominated forest (Solberg et al., 2006; Bollandsås and Næsset, 2007a). According to Økland (1994), no cutting had been carried out on this forest since 1940, and no even-aged forestry had been carried out in this forest before that time.

Table 7. Model parameters for individual tree diameter increment (mm/5yr).

| Variable[b] | Parameter[a] (*Standard error*)<br>Norway spruce | Scots pine | Birch | Other broadleaves |
|---|---|---|---|---|
| Intercept | 17.839 (*1.087*) | 25.543 (*1.310*) | 11.808 (*0.485*) | 2.204 (*0.289*) |
| $DBH$ (mm) | 0.0476 (*0.002*) | 0.0251 (*0.002*) |  | 0.063 (*0.003*) |
| $DBH^2$ (mm$^2$) (×10$^{-5}$) | –11.585 (*0.376*) | –5.660 (*0.363*) | 9.616 (*0.481*) | – 8.320 (*0.882*) |
| $DBH^3$ (mm$^3$) (×10$^{-8}$) |  |  | – 9.585 (*1.499*) |  |
| $BAL$ (m$^2$ha$^{-1}$) | –0.3412 (*0.007*) | –0.216 (*0.009*) |  |  |
| $SI$ (m) | 0.906 (*0.008*) | 0.698 (*0.013*) | 0.519 (*0.008*) | 0.359 (*0.015*) |
| $BA$ (m$^2$ha$^{-1}$) | –0.024 (*0.007*) | –0.123 (*0.008*) | – 0.152 (*0.003*) | –0.177 (*0.006*) |
| $LAT$ (°. m×100/60) | –0.268 (*0.018*) | –0.336 (*0.022*) | –0.161 (*0.007*) |  |
|  |  |  |  |  |
| Observations | 68,584 | 29,447 | 60,229 | 20,212 |
| $R^2$ | 0.28 | 0.19 | 0.10 | 0.09 |
| RMSE (mm) | 8.3 | 6.9 | 6.0 | 8.2 |

[a] $p<0.01$ for all parameters

[b] $DBH$=diameter at breast height, $BAL$=basal area of larger trees, $SI$=site index value, $BA$=stand basal area, $LAT$=latitude

## Results

### *Model parameters*

Table 5 shows the parameters of the models that predicted the probability of recruitment on a particular plot. For all species we obtained a significant negative sign for *BA*, so that the probability of recruitment would decrease with increasing stand density, as expected. The positive sign of *PBA* reflected greater probability of recruitment for a particular species when that species occupied an important part of the stand basal area. For the pioneer species (pine and birch) *SI* were not statistically significant. The model goodness of fit is indicated by the expected number of misclassified plots (Hosmer and Lemeshow, 2000), which ranged between 0.8 and 8.4 percent of the total number of plots.

The parameters of the models for the number of recruited trees per ha, conditional on positive recruitment are in Table 6. For all species, the number of recruits was negatively and significantly related to stand basal area, as expected. *SI* and *PBA* affected positively and significantly the number of recruits of spruce, birch, and other broadleaves. However, the low $R^2$'s, especially for pine, show that the models explained only a small part of the variation in number of recruits on plots that had recruitment.

The parameters of the diameter increment models for individual trees are in Table 7. For all species, as tree *DBH* increased, the corresponding diameter increment increased initially, reached a maximum, and then decreased. The total basal area per ha, *BA*, had a statistically significant, and as expected negative, effect diameter increment. However, the basal area of trees larger than the tree under consideration, *BAL*, had a significant and negative effect on the diameter growth of spruces and pines only.

Table 8. Parameters of individual tree mortality models (probability of dying in five years).

| Variable[b] | Parameter [a] (*Standard error*) | | | |
|---|---|---|---|---|
| | Spruce | Pine | Birch | Other broadleaves |
| Intercept | –2.492 (*0.083*) | –1.808 (*0.125*) | –2.188 (*0.074*) | –1.551 (*0.083*) |
| *DBH* | –0.020 (*0.001*) | –0.027 (*0.001*) | –0.016 (*0.001*) | –0.011 (*0.001*) |
| $DBH^2$ (mm$^2$) (×10$^{-5}$) | 3.200 (*0.241*) | 3.300 (*0.305*) | 2.700 (*0.350*) | 1.400 (*0.342*) |
| *BA* (m$^2$ha$^{-1}$) | 0.031 (*0.002*) | 0.055 (*0.004*) | 0.030 (*0.002*) | 0.016 (*0.002*) |
| | | | | |
| Observations | 71,805 | 30,645 | 66,156 | 23,944 |
| Misclassified trees (%) | <1 | <1 | <1 | <1 |

[a] $p<0.001$ for all parameters

[b] *DBH*=diameter at breast height, *BA*=stand basal area.

As expected, site index, *SI*, had a statistically significant and positive effect on the diameter increment of all species. The other site indicator, latitude (*LAT*), had a negative effect on individual tree growth, except for the other broadleaves. This was most likely due to the fact that most of the "other broadleaf" are in southern Norway.

Table 8 shows the estimates of the parameters of the mortality models. As expected, the probability that a tree would die was highest for the trees of smallest *DBH*, it decreased as *DBH* increased, reached a minimum, and then increased with *DBH*.

Also as expected, other things being equal, the mortality rate was significantly higher in stands of higher basal area, *BA*. Overall, the models misclassified less than one percent of the trees.

### *Short-term validation*

The comparisons of observed and predicted number of trees per ha, for the 416 plots not used in parameter estimation, are summarized in Figure 1. The model was applied to each plot, after subtracting the harvest, to predict the plot state after 10 years. Figure 1 shows the predicted average number of trees, over the 416 plots, after 10 years, by size class and species groups, and the 95 percent confidence interval around the mean observed number of trees.

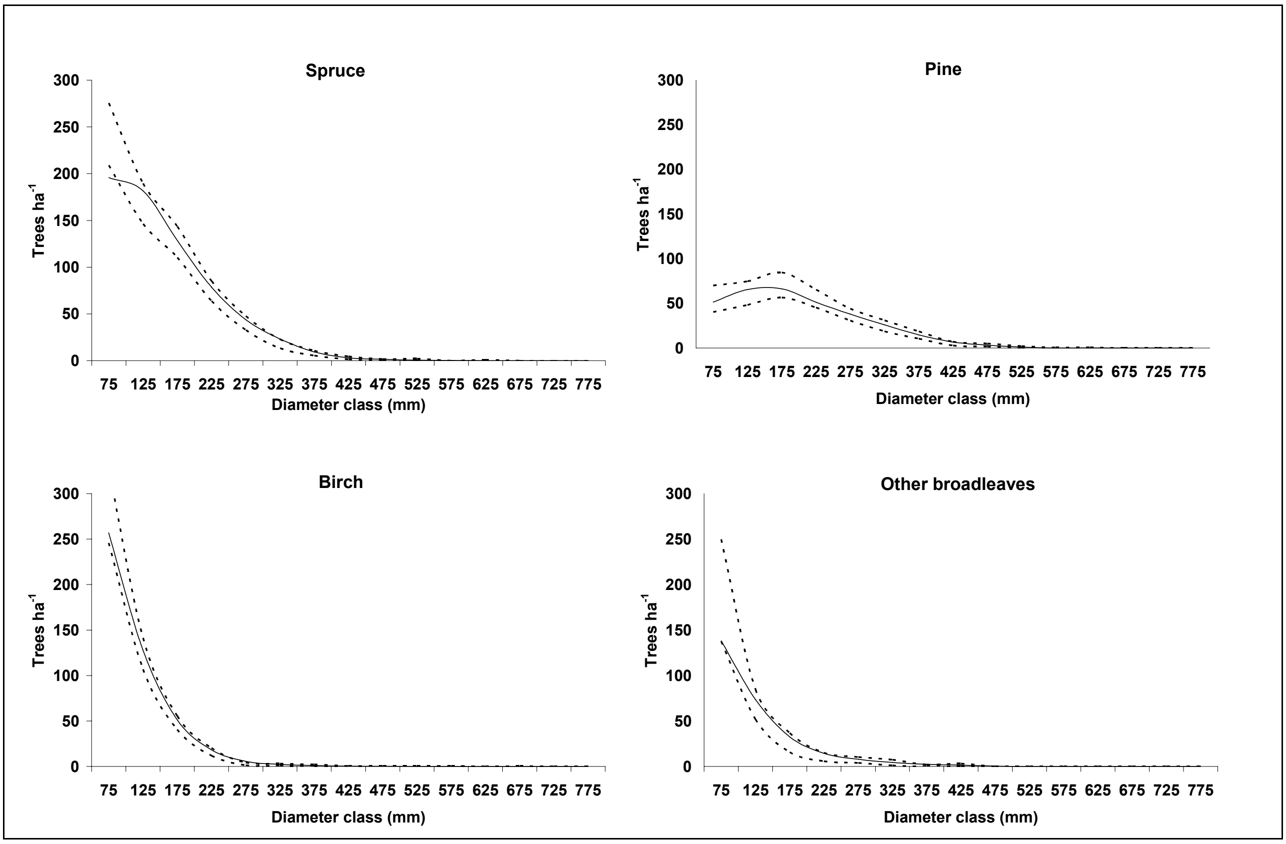

Figure 1. Average predicted number of trees (solid line) over 10 years on 416 post-sample plots. Dotted lines are the upper and lower bound of the 95% confidence interval of the mean number of observed number of trees.

### *Long-term validation*

The results of the long-term simulations of stand growth with initial pure stands of spruce, pine, birch, or other broadleaves are in Figure 2. The stands had the same initial number of trees by size class, and thus the same initial basal area. Also, the same site index, $SI$=14 m, was assumed.

Figure 3 illustrates the effect of site index on stand growth. The initial state was a mixed stand with 25 trees in diameter class 175 mm 100 trees in class 225 mm for all species. As the $SI$ increased from 6 m to 23 m the initial growth rate and the steady-state basal area increased, as expected. In all four simulations the proportion of spruce increased over time, most rapidly for high $SI$, to ultimately crowd out the other species. For the lowest value of $SI$, pine still made up an important part of the basal area at the end of the simulation period.

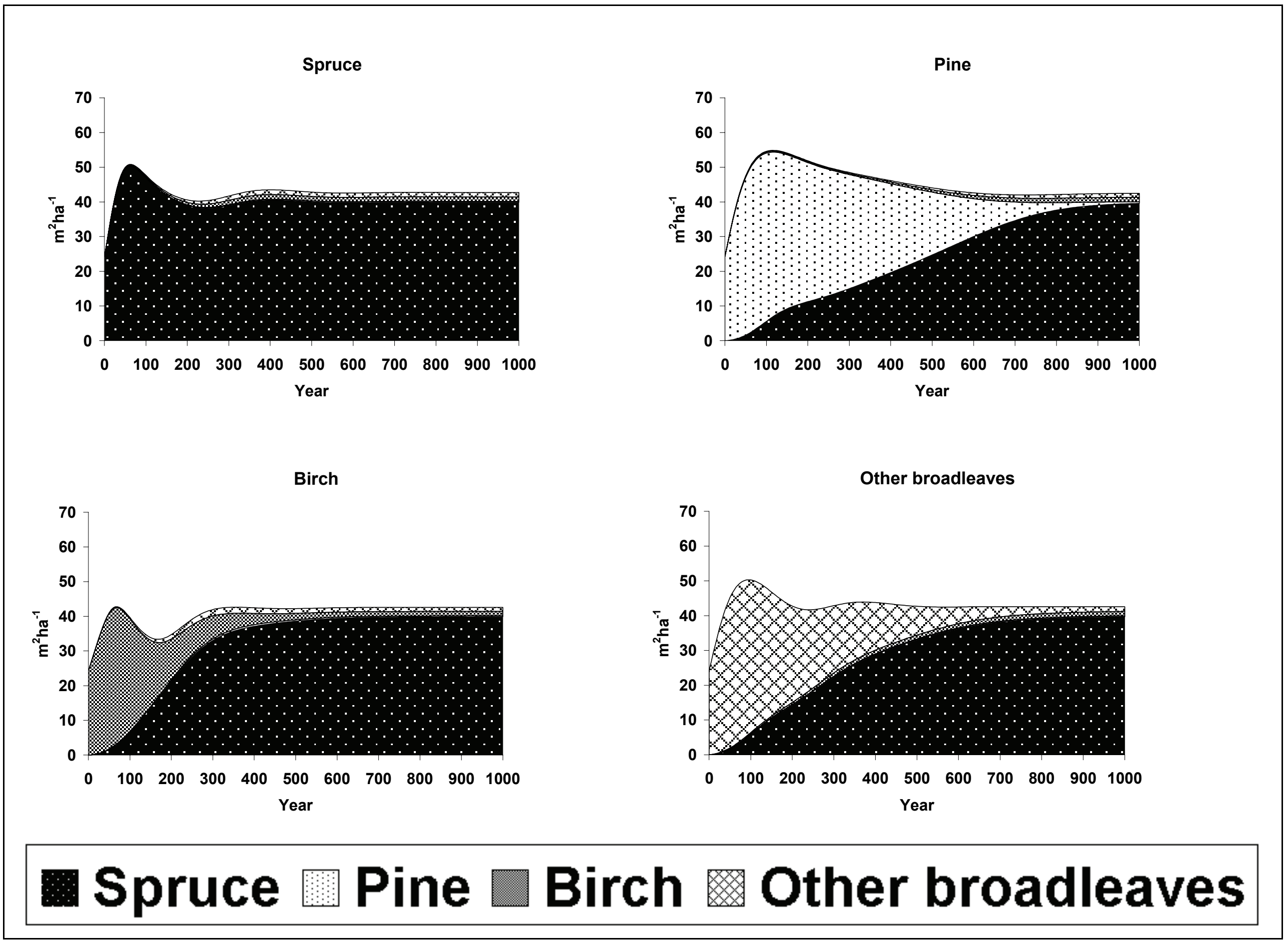

Figure 2. Predicted basal area of an initial stand of pure spruce, pine, birch, or other broadleaves.

Table 9 allows a comparison of the steady states obtained in the different simulations with observed old stands. The “average” refers to the average plot in a boreal nature forest that had not been disturbed for at least 75 years (Bollandsås and Næsset 2007a). The forest covers different sites and stands of varying composition. The “Best stand” refers to the plot with the highest basal area per ha and the highest site index in the same forest.

## Discussion

The model parameters in Tables 5, 6, 7, and 8 all had signs in accord with prior expectations. The Hosmer and Lemeshow (2000) tests show that the recruitment models (Table 5) misclassified only a few plots and the mortality models (Table 8) misclassified only a few trees, although this measure of fit is imperfect it does not suggest a flaw in the model (Greene 1993, p.

652). The model for spruce recruitment seems to be the least accurate. This is probably due to the fact that spruce is the species that is most often planted and thus will have least variability of recruitment explained by stand conditions. The $R^2$ of the models of the number of recruits conditional on positive recruitment (Table 6) ranged between 0.01 and 0.18. Thus, it appears that most of the variation in recruitment was dealt with the probability of recruitment models (Table 5). Nevertheless, because the effect of stand basal area on conditional recruitment was systematic across species and significant, the conditional models of the number of recruits were maintained.

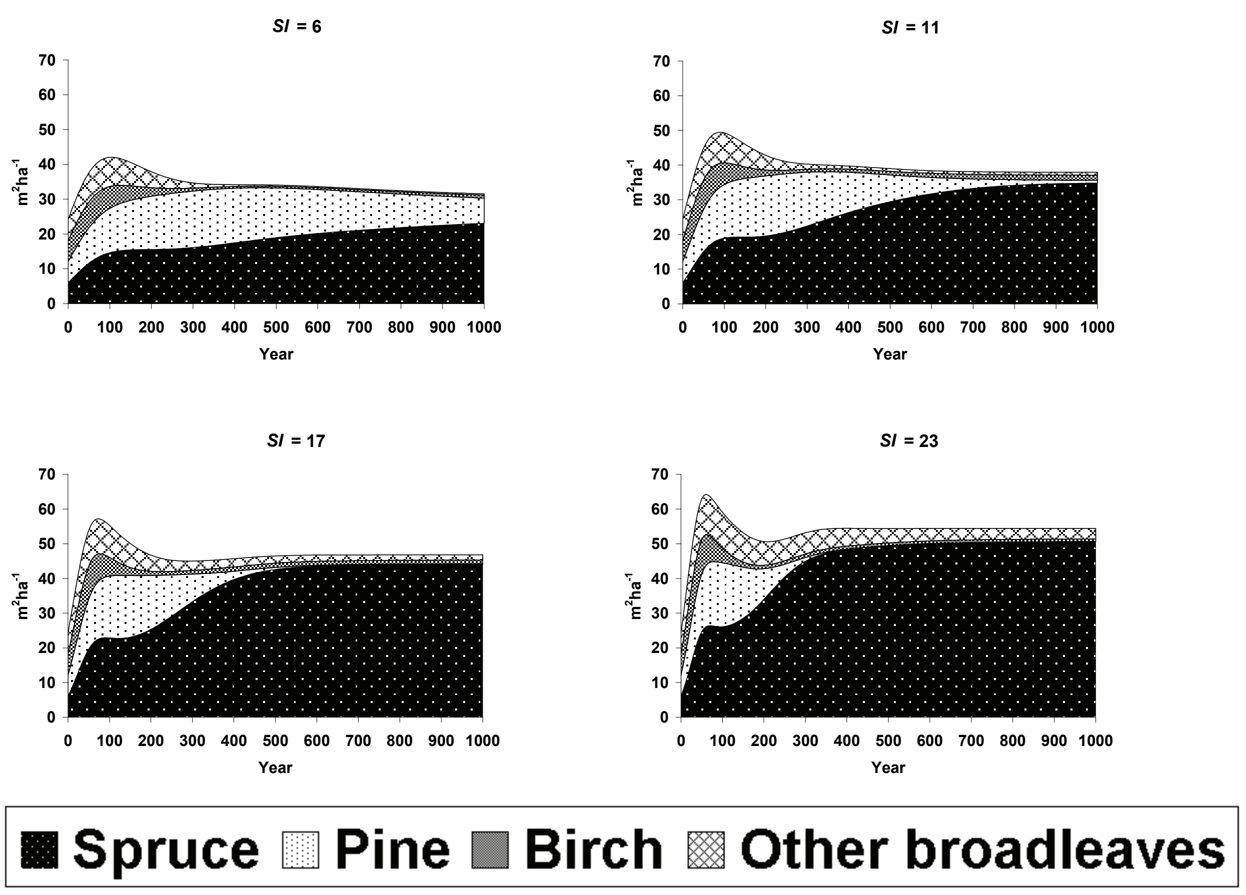

Figure 3. Predicted basal area of an initial mixed stand on four different sites.

The $R^2$ of the diameter increment models were low, as observed by other authors (e.g. Lin et al., 1996; Lin and Buongiorno, 1998; Ralston et al., 2003). This suggests that diameter increment could be modelled as a constant, at least for short-term projections, but for long forecasts the statistically significant effects of tree size and stand basal area on diameter growth could be biologically significant and were therefore kept in the model. This principle was applied to the other equations as well, where variables were maintained only if they were statistically significant and had plausible signs according to prior biological knowledge.

Growth and mortality data from trees on the same sample plot could be correlated. Hence, a mixed modelling approach with random plot effect would seem preferable. However, in this case, mixed models gave biased predictions of changes on independent plots. In contrast, the ordinary least squares method gave unbiased predictions, and was therefore preferred. This principle of choosing estimation methods according to the performance of the entire system of equations (1) was also applied in choosing the functional form and the variables of individual equations.

It can be noted that in their final version, the mortality models in Table 8 do not link mortality to the tree status in the stand canopy or with the site productivity as might be expected. Early versions of the model did include such variables, but they either had parameter signs inconsistent with prior knowledge, or were statistically insignificant.

Because from boundary plots were discarded, the models represent the growth of stands growing inside a forest. Boundary plots are affected by manure from agricultural fields, salt from roads, draught because of ditching, etc. Thus, their deletion should avoid bias due to human activities outside the forested area.

Figure 1 summarizes the short-term validation tests of the model. In predictions of the number of trees by species and size over all the post-sample plots, the average predicted number of trees stayed within the 95 percent confidence interval around the mean observed number of trees. The single exception was the low number of spruce trees in the smallest diameter class, . But this could well reflect the power of the test, which leaves a positive probability of rejecting the null hypothesis when it is true. A longer time interval would clearly have been better for such a test, but this was not feasible with the available data. Regardless, the wide variety of plots to which the model was applied served as a substitute, albeit imperfect, for the relatively short time span.

In all four cases of the long term validation appearing in Figure 2, the peak of basal area after one century could be attributed to the death of large trees after an aging phase of the stand (Stöcker, 2002). Afterwards, the stands converged to a steady state in the long run. As shown by Figure 2 and Table 9, this steady state was nearly the same regardless of the initial condition, in terms of total basal area, volume, number of trees, mean diameter and height, and species composition. In the long run, the stands were dominated by spruce, independently of the initial state. Figure 3 shows that the distribution of species depends on site index. This is in accord with prior expectation as low productivity sites can support mostly pioneer species like pine growing in low density stands with much light. As the site improved, spruce became the dominant species. The other broadleaves also became relatively more abundant, as expected.

Table 9 shows that the steady states predicted by the model were close to the values observed in a forest that has been little disturbed by silviculture. However, because of the relatively short time (75 years) since the last human intervention (Økland, 1994), the observed forest has most likely not reached a steady state, and the basal area will probably still increase, at least on the poorer sites where the evolution is slower. On the richer sites the basal area is more likely to be close to a steady state. The steady-states characteristics predicted by the simulations were in general agreement with those observed on the forest.  In particular, there was relatively more spruce in the observed forest on the better sites, in accord with the model predictions.

Table 9. Predicted stand characteristics after 1,000 years for different initial conditions, compared to the observed characteristics of old stands.

| | *BA* ($m^2ha^{-1}$) | *V* ($m^3ha^{-1}$) | *N* ($ha^{-1}$) | $d_g$ (mm) | $h_L$ (dm) | $BA_s$ (%) | $BA_p$ (%) | $BA_b$ (%) | $BA_o$ (%) |
|---|---|---|---|---|---|---|---|---|---|
| Observed old stands | | | | | | | | | |
| Average[a] | 34.2 | 361 | 1033 | 211 | 216 | 90.0 | 0.1 | 9.9[e] | |
| Best stand[b] | 45.5 | 635 | 630 | 303 | 289 | 100.0 | 0.0 | 0.0[e] | |
| Predicted stand state | | | | | | | | | |
| Spruce[c] | 42.5 | 381 | 826 | 257 | 216 | 93.7 | 0.7 | 3.0 | 2.6 |
| Pine[c] | 42.5 | 381 | 815 | 259 | 216 | 93.1 | 1.1 | 3.1 | 2.7 |
| Birch[c] | 42.6 | 381 | 825 | 257 | 216 | 93.7 | 0.7 | 3.0 | 2.6 |
| Other broadleaves[c] | 42.6 | 381 | 820 | 258 | 216 | 93.3 | 0.6 | 3.0 | 3.1 |
| Mixed[d] SI = 6 | 31.5 | 218 | 572 | 268 | 169 | 73.1 | 23.0 | 2.7 | 1.2 |
| Mixed[d] SI = 11 | 37.9 | 314 | 786 | 250 | 202 | 91.8 | 2.1 | 3.9 | 2.3 |
| Mixed[d] SI = 17 | 46.8 | 448 | 849 | 266 | 228 | 94.2 | 0.3 | 2.3 | 3.2 |
| Mixed[d] SI = 23 | 54.4 | 577 | 892 | 278 | 248 | 93.0 | 0.1 | 1.4 | 5.5 |

*BA*=stand basal area, *V*=total volume, *N*=number of stems, $d_g$=mean basal area diameter, $h_L$=basal area weighted mean height, $BA_s$, $BA_p$, $BA_b$, $BA_o$=basal area of spruce, pine, birch, and other broadleaves, *SI*= site index.

[a] mean of plots in boreal nature reserve (Bollandsås and Næsset, 2007a).

[b] "extreme" plot (highest observed basal area and site index) in boreal nature reserve.

[c] Initial pure stands of Spruce, Pine, Birch, and Other hardwoods, with 100 trees in diameter class 175 and 275 mm, and 400 trees in class 225 mm.

[d] Mixed initial stand with 25 trees in class 175 and 275 mm and 100 trees in class 225 mm for all species.

[e] Data for all deciduous trees.

## Summary and Conclusion

The description of forest stands by the number of trees per unit area in the main species groups and fifteen size classes is simple, but with enough detail for applications concerned with continuous-cover/uneven-aged forestry, where considerations of the species composition and forest stand structure are important.

The methods to estimate the parameters of the matrix growth model take advantage of the rich data set of the national forest inventory of Norway, both at stand and individual tree level. The data encompass a wide range of growth conditions. This cross-sectional variability compensates for the short time interval between observations, allowing the accurate estimation of parameters without the need of long-term experiments that are necessarily costly and limited in scope. Furthermore, the model predictions depend on the climate during the observation period. Thus, if the Norwegian climate were to change significantly, the recruitment, growth, and mortality equations would need to be updated when new data reflecting the climate change became available.

The results of the short-term validation showed that the model was sufficiently accurate for predictions of stand growth over at least a decade. However, the exclusion from the data of young forests, which was necessary to avoid bias in the recruitment rates due to artificial regeneration, means that the model is most suitable for well established stands that will continue to be managed with natural regeneration.

In applications of the model, accurate predictions are not to be expected on each and every stand, but only on average, on a large number of stands. As indicated by the statistics of the sub-models, given similar conditions the recruitment, mortality, and growth can vary considerably from stand to stand and tree to tree. Little can be done to reduce this uncertainty, although it can be taken into account in decision making (Lin and Buongiorno, 1998).

The prediction error can be expected to increase substantially with the length of the projection. Nevertheless, the long-term stand growth predicted deterministically with the present model was found to have desirable properties. In particular, it led to a unique steady state for each site level, independent of the initial tree distribution by size class and species group. The species composition of the predicted steady state was in agreement with ecological knowledge, such as the persistence of pines on the poorer sites and of spruces on the best sites. It was also in agreement with the available (though admittedly limited) observations on undisturbed forests, in terms of total basal area and volume, number of trees, average diameter, and species composition.

In this paper, the model was applied only to simulate natural stand growth without human disturbance, to check its validity as a predictor of expected stand growth. The same deterministic model form, or a stochastic extension (Liang et al. 2006), should in future applications prove useful to predict the economic and ecological effects of continuous-cover/uneven-aged forest management, by simulation, or optimization.

## Acknowledgements

The research leading to this paper was funded by the Research Council of Norway (research projects no. 153185/110 and 146675/110). We wish to thank Dr. Mo Zhou for her comments and suggestions and to three anonymous reviewers for their valuable comments. Thanks also to Rune Eriksen at the Norwegian Institute of Land Inventory for supplying the data.

**References**

Andreassen, K., Tomter, S.M. (2003). Basal area growth models for individual trees of Norway spruce, Scots pine, birch and other broadleaves in Norway. *Forest Ecology and Management 180*, 11-25.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: Wiley.

Anon., (1999). SAS OnlineDOC (R), Version 8. SAS Institute Inc. Cary, NC.

Blingsmo, K.R., Veidahl, A. (1993). *BESTPROG. Et hjelpemiddel for beslutninger om bestandsbehandling.* Internal report, Department of Forest Sciences, Agricultural University of Norway, Ås, 1-24. (In Norwegian)

Bollandsås, O.M. & Næsset, E. (2007a). Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. *Scandinavian Journal of Forest Research, 22*(1), 33-47.

Bollandsås, O.M. & Næsset, E. (2007b). Height-diameter models for individual trees in Norway. Submitted to Silva Fennica.

Bosch, C. A. (1971). Redwoods: a population model. *Science 172*, 345-349.

Braastad, H. (1966). *Volume tables for birch*. Communications of Norwegian Forest Research Institute 21, 265-365. (In Norwegian with English summary).

Braastad, H. (1980). *Growth model computer program for Pinus sylvestris.* Norwegian Forest Research Institute. Report 35(5), 272-359. (In Norwegian with English Summary).

Brantseg, A. (1967). *Volume functions and tables for Scots pine. South Norway.* Communications of Norwegian Forest Research Institute 22, 695-739. (In Norwegian with English summary).

Buongiorno, J. & Michie, B.R. (1980). A matrix model of uneven-aged forest management. *Forest Science, 26*(4), 609-625.

Buongiorno, J., Peyron, J.L., Houllier, F. & Bruciamacchie, M. (1995). Growth and management of mixed-species, uneven-aged forests in the French Jura: Implications for the economic returns and tree diversity. *Forest Science, 41*(3), 397-429.

Eid, T., Hobbelstad, K. (2000). AVVIRK 2000 - a large scale forestry scenario model for long-term investment, income and harvest analysis. *Scandinavian Journal of Forest Research 15,* 472-482.

Eid, T. & Tuhus, E. (2001). Models for individual tree mortality in Norway. *Forest Ecology and Management, 154*, 69-84.

Flewelling, J.W. & Pienaar, L.V. (1981). Multiplicative regression with lognormal errors. *Forest Science, 27*(2), 281-289.

Greene, W.H. (1993). *Econometric analysis*. New York: MacMillan.

Gobakken, T., Lexerød, N., Eid, T. (2005). A growth simulator based on models for individual trees. In: Hobbelstad, K. (Ed.). Forest inventory and planning in nordic countries. Proceedings of SNS meeting at Sjusjøen, Norway, September 6-8, 2004. NIJOS rapport 9/05, 243-248

Holte, A. (1993). *Diameter distribution functions for even-aged (Picea abies) stands*. Communications of Norwegian Forest Research Institute 46(1), 1-46.

Hoen, H.F., Eid, T. 1990. *En modell for analyse av behandlingsalternativer for skog ved bestandssimulering og lineær programmering.* Norwegian Research Institute, Ås. Report 9, 1-35. (In Norwegian).

Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression (2nd Edition).* New York: Wiley.

Liang, J., Buongiorno, J. & Monserud, R.A. (2005). Growth and yield of all-aged Douglas fir – western hemlock forest stands: a matrix model with stand diversity effects. *Canadian Journal of Forest Research 35*, 2368-2381.

Liang, J., Buongiorno, J. & Monserud, R.A. (2006). Bootstrap Simulation and Response Surface Optimization of Management Regimes for Douglas-Fir/Western Hemlock Stands. *Forest Science 52*, 579-594.

Lexerød, N. (2005). Recruitment models for different tree species in Norway. *Forest Ecology and Management 206*(1/3), 91-108.

Lin, C.R., Buongiorno, J., Vasievich, M. (1996). A multi-species density-dependent matrix growth model to predict tree diversity and income in northern hardwood stands. *Ecological Modelling 91,* 193-211.

Lin, C.R. & J. Buongiorno. (1998). Tree diversity, landscape diversity, and economics of maple-birch forests: Implications of a Markovian models. *Management Science 44*(10), 1351-1366.

Lu, H-C. & Buongiorno, J. (1993). Long- and short-term effects of alternative cutting regimes on economic returns and ecological diversity in mixed-species forests. *Forest Ecology and Management 58*(3/4), 173-192.

Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T. & Saramäki, J. (2000). Comparison of percentile based prediction methods and Weibull distribution in describing diameter distribution of heterogeneous Scots pine stands. *Forest Ecology and Management 133*(3), 263-274.

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J. & Pitkänen, J. (2005). Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management 216*(1/3), 41-50.

Miller, D.M. (1984). Reducing transformation bias in curve fitting. *The American Statistician 38* (2), 124–126.

Mønness, E. (1982). *Diameter distributions and height curves in even-aged stands of Pinus sylvestris L.* Communications of Norwegian Forest Research Institute 36 (15): 1-43.

Økland, B. (1994). Mycetophilidae (Diptera), an insect group vulnerable to forest practices? A comparison of clearcut, managed and semi-natural spruce forests in southeastern Norway. *Biodiversity and Conservation 3*, 68-85.

Pommering, A. & Murphy, S.T. (2004). A review of the history, definitions, and methods of continuous cover forestry with special attention to afforestation and restocking. *Forestry 77*(1), 37-44.

Pretzsch, H., Biber, P., Ďurský, J. (2002). The singletree-based stand simulator SILVA: construction, application and evaluation. *Forest Ecology and Management 162*, 3-21.

Ralston, R., Buongiorno, J., Schulte, B., Friend, J. (2003). Non-linear matrix modelling of forest growth with permanent plot data: The case of uneven-aged Douglas-fir stands. *International Transactions in Operational Research 10*, 461-482.

Solberg, S., Næsset, E. & Bollandsås, O.M. (2006). Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering and Remote Sensing 72*(12), 1369-1378

Stenberg, L.C. & Siriwardana, M. (2006). The steady-state treatment of forestry in CGE models. *International Journal of Agricultural Resources, Governance, and Ecology 5*, 1-17.

Stöcker, G. (2002). Growth dynamics of Norway Spruce (Picea abies (L.) Karst.) in natural spruce forest of the national park Hochharz – 2. Climax, ageing and decay phases. *European Journal of Forest Research 121*, 109-127.

Tveite, B. (1977). *Site index curves for Norway spruce (Picea abies (L.) Karst.).* Norwegian Forest Research Institute, Report 33, 1-84.

Usher, M.B. (1966). A matrix approach to the management of renewable resources, with special application to selection forests. *The Journal of Applied Ecology 3*, 355-367.

Vestjordet, E. (1967). *Functions and tables for volume of standing trees. Norway spruce.* Communications of Norwegian Forest Research Institute 22, 543-574. (In Norwegian with English summary).

Vestjordet, E. (1972). *Diameter distribution and height curves for even-aged stands of Norway spruce.* Communications of Norwegian Forest Research Institute 29(8), 469-557.

ISBN 978-82-575-0785-5
ISSN 1503-1667

Norwegian University of Life Sciences
NO–1432 Aas, Norway
Phone +47 64 96 50 00
www.umb.no, e-mail: postmottak@umb.no